А.Е. Кононюк

ДИСКРЕТНО-НЕПРЕРЫВНАЯ МАТЕМАТИКА

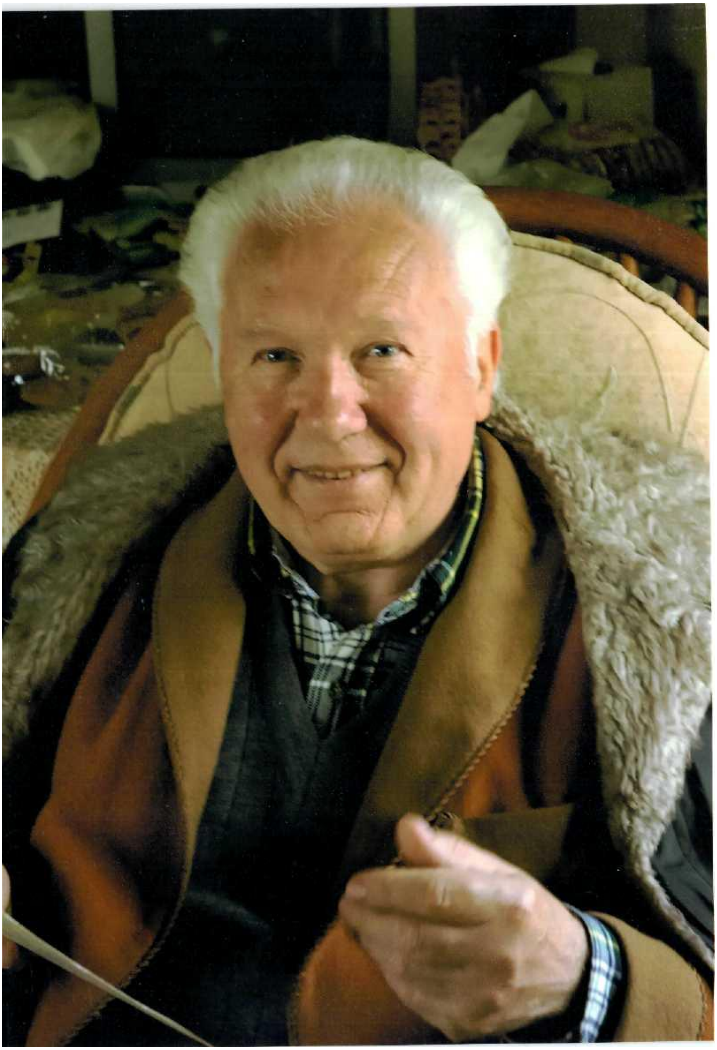
Книга 8

Пространства

Часть 1

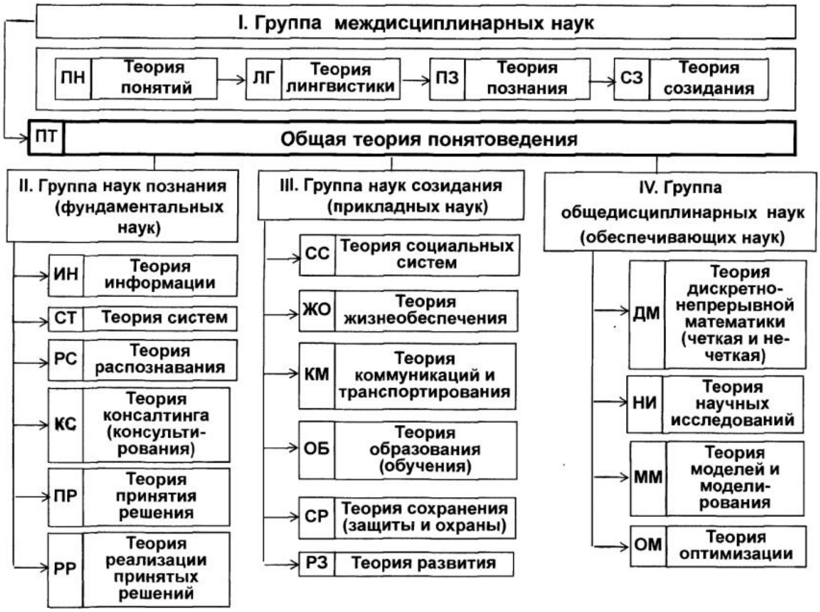
Введение в теорию пространств

Киев
«Освіта України»
2016



Кононюк Анатолий Ефимович

Структурная схема развивающейся панмедийной системы наук



УДК 51 (075.8)
ББК В161.я7
К65

Рецензенты:

В.В.Довгай - к-т физ.-мат. наук, доц. (Национальный технический университет „КПІ”);

В.В.Гавриленко - д-р физ.-мат. наук, проф., *О.П.Будя* - к-т техн. наук, проф. (Киевский университет экономики, туризма и права);

Н.К.Печурин - д-р техн. наук, проф. (Национальный авиационный университет).

Кононюк А. Е.
К213 Дискретно-непрерывная математика. (Пространства). — В 12-и кн. Кн.8. ч.1.— К.:Освіта України. 2016.—620 с.

ISBN 978-966-373-693-8 (многотомное издание)
ISBN 978-966-373-694-5 (книга 8. ч.1)

Многотомная работа содержит систематическое изложение математических дисциплин, исспользуемых при моделировании и исследованиях математических моделей систем.

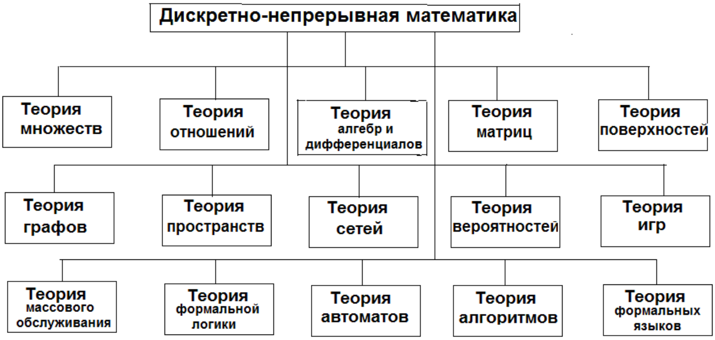
В работе излагаются основы теории множеств, отношений, поверхностей, пространств, алгебраических систем, матриц, графов, математической логики, теории формальных грамматик и автоматов, теории алгоритмов, которые в совокупности образуют единную методолгически взамосвязанную математическую систему «Дискретно-непрерывная математика».

Для бакалавров, специалистов, магистров, аспирантов, докторантов всех специальностей.

УДК 51 (075.8)
ББК В161.я7

ISBN 978-966-373-693-8 (многотомное издание)
ISBN 978-966-373-694-5 (книга 8. ч.1)

Структура
открытой развивающейся панмедийной системы математических наук (дисциплин)
"Дискретно-непрерывная математика"



Оглавление

Введение

Рассказывать о пространстве не как философской категории, а как о математическом понятии очень непросто. Наверное, нужно начать с того, что пространство — очень объемное понятие, которому в разных контекстах придаются самые различные смыслы: вспомним хотя бы космическое, воздушное, межклеточное или информационное пространство. Конечно, **разумно было бы начать с определения, однако пространство, как и любое другое основное понятие, является неопределяемым (аналогично число, множество, функция в алгебре, точка, прямая, плоскость в геометрии). Неопределяемые понятия мы можем только описывать, разъясняя, что имеется в виду, заменяя определяемое слово синонимами, обращаясь к интуиции и жизненному опыту**.

Итак, под **пространством понимают множество некоторых объектов, правила работы с ними и набор аксиом, которым эти правила должны подчиняться**. Иными словами, **это множество со введенной на нем структурой.** По-разному определяя множество и структуру, мы получаем разные пространства. Из изучаемых в школьном курсе можно отметить Евклидово пространство (одномерное, двумерное и трехмерное), линейное или векторное пространство и вероятностное пространство. Их элементами являются соответственно точки, векторы и элементарные события. Упоминается также четырехмерное пространство-время — пространство Минковского. (Заметим, что пространство Минковского не является четырехмерным Евклидовым пространством, поскольку имеет три пространственные и одну временную, но не пространственную, координату).

Одной из важнейших характеристик пространства является его размерность. С точки зрения аналитической геометрии, **размерность фигуры равна числу координат, нужных для определения положения лежащей на этой фигуре точки; например, положение точки на кривой определяется одной координатой, на поверхности — двумя координатами, в трёхмерном пространстве — тремя координатами**. (Необходимо отметить, что с середины 19 века геометрия стала изучать многомерные пространства, а в 20 веке — и пространства дробных размерностей).

Примером одномерного пространства является числовая прямая, положение каждой точки на ней можно характеризовать единственным числом. На прямой могут располагаться точки и отрезки (интервалы, полуинтервалы), они имеют одну пространственную характеристику — протяженность или длину. Примером двумерного пространства является плоскость, точки плоскости задаются двумя координатами. Плоские объекты характеризуются не только длиной, но и шириной. Примером трехмерного пространства является окружающий нас мир (Это утверждение верно до некоторых пределов. С точки зрения общей теории относительности, наш мир — искривляемое массивными телами четырехмерное пространство-время). Наряду с длиной и шириной появляется третье измерение, называемое высотой. Положение точек в трехмерном пространстве задается тремя пространственными координатами, в прямоугольной системе координат называемыми абсциссой, ординатой и аппликатой.

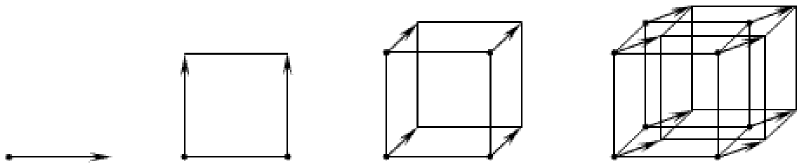
В рамках базовой программы изучается только прямоугольная система координат. Однако необходимо отметить, что существуют и непрямоугольные системы координат, например, полярная, сферическая и др.. В полярной системе координат на плоскости положение точки характеризуется двумя числами: расстоянием от точки до полюса и углом между полярной осью и отрезком, соединяющим точку с полюсом. Эта система координат удобна тем, что в ней уравнения некоторых линий принимают более простой вид, чем в прямоугольной системе координат. Аналогично, для задания положения точки на поверхности Земного шара понятия широты и долготы более удобны, чем какие-либо другие координаты. В сферической системе координат положение точки задают расстоянием от нее до начала координат, зенитным и азимутальным углами.

В физике используется «фазовая плоскость» — координатная плоскость, в которой по осям координат откладываются какие-либо две переменные (фазовые координаты), однозначно определяющие состояние системы. Каждому возможному состоянию системы соответствует точка фазового пространства. Изменение состояния системы отображается на фазовой плоскости движением этой точки. Фазовая плоскость является частным случаем фазового пространства, которое может иметь большую размерность. **Состояние сколь угодно сложной системы представляется в фазовом пространстве одной единственной точкой, а эволюция этой системы — перемещением этой точки.**

1. Виды и типы пространств

1.1. Пространство как математическое понятие

Наибольший интерес вызывает обсуждение многомерных гипертел. Приведем схему, позволяющую представить эволюцию геометрических объектов от точки до четырехмерного гиперкуба. Пусть у нас есть точка, которую мы сдвигаем вдоль прямой на некоторое расстояние. Если мы представим, что при движении точка оставляет след, то в результате сдвига мы получим отрезок. Будем теперь смещать отрезок в перпендикулярном ему направлении на расстояние, равное его длине. Результат построения — квадрат. Смещая теперь квадрат в перпендикулярном его плоскости направлению на расстояние, равное длине его стороны, получим куб. Рассуждая аналогично, мы можем вообразить четырехмерный гиперкуб. Он будет результатом смещения нашего трехмерного куба в направлении ему перпендикулярном на расстояние, равное длине любого из его ребер (см. рис.).



Больше того, наш алгоритм позволяет заметить геометрические характеристики гиперкуба! Проследим поэтапно: начальное и конечное положение движущейся точки обусловили наличие двух вершин отрезка. В свою очередь, четыре вершины квадрата образовались из двух вершин исходного и еще двух — смещенного отрезка. Число вершин куба представляет собой сумму числа вершин исходного и смещенного квадратов, поэтому их восемь, а число вершин гиперкуба будет равно шестнадцати — это вершины исходного и смещенного кубов. (Заметим, что число вершин 2, 4, 8, 16… является геометрической прогрессией.)

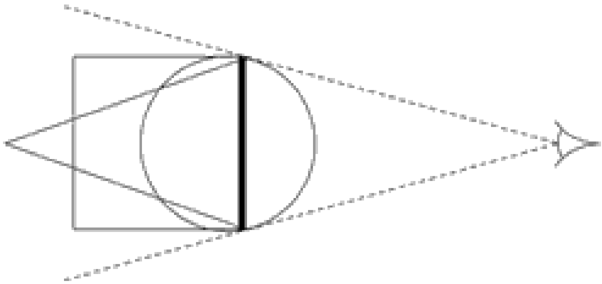
Продолжим наши наблюдения за объектом, который мы не в силах представить, но можем охарактеризовать. Единственная «сторона» отрезка — результат движения точки. Четыре стороны квадрата складываются из начального и конечного отрезков и еще двух отрезков, соединяющих исходные, и полученные смещением вершины этого отрезка. Поэтому же число ребер куба равно 12 (это по 4 стороны исходного и смещенного квадратов и еще 4 ребра, соединяющие соответствующие вершины исходного и смещенного). Аналогично число ребер гиперкуба — 32 (это по 12 ребер исходного и смещенного кубов и еще 8 ребер, образованных при движении 8 вершин исходного куба).

Далее, у нашего трехмерного куба 6 плоских граней — это исходный и смещенный квадраты, а также квадраты, полученные движением 4 сторон исходного квадрата. Аналогично число плоских граней гиперкуба — 24: по 6 граней у исходный и смещенного кубов, и еще 12, полученных движением 12 ребер исходного куба. Помимо 24 плоских граней, гиперкуб будет иметь еще 8 кубических граней: действительно, отрезок ограничен 2 точками, квадрат — 4 сторонами, куб — 6 гранями, гиперкуб — 8 кубами и т. д.; числа 2, 4, 6, 8… образуют арифметическую прогрессию. (Продолжая, мы можем рассматривать пятимерный гиперкуб. Он имеет 32 вершины, 80 ребер, 80 квадратных, 40 кубических и 10 гиперкубических граней. И так далее.)

При всей схожести построения объектов, пространства, в которых они существуют совсем разные.

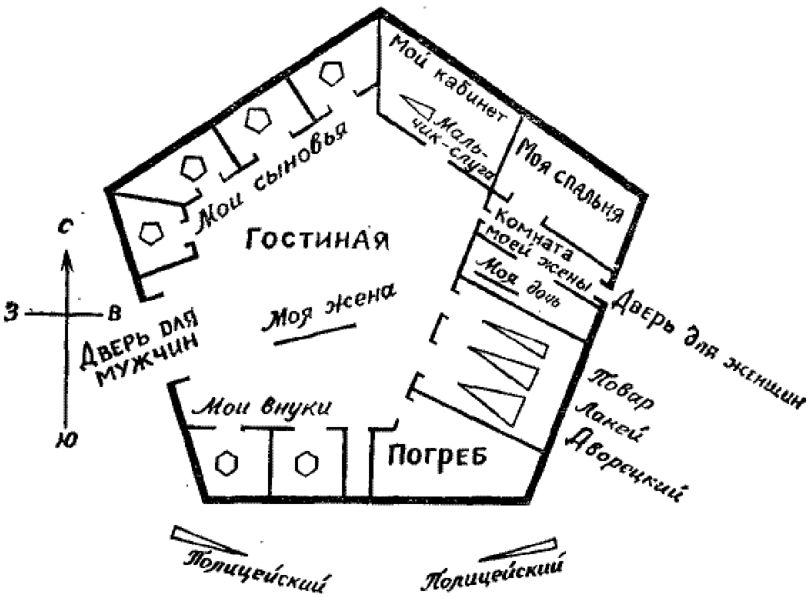
Рассмотрим мир жителей прямой. Вне прямой для них ничего не существует, просто ничего нет, даже пустоты. Внутри этого мира органами зрения можно видеть лишь точки. Это сами точки, или перпендикулярные проекции отрезков (представьте, что мы смотрим на иголку со стороны острия). Движение возможно только между двумя объектами на прямой, так как обойти точки, не сходя с этой нее, невозможно (см. рис.).

Двумерный мир представляет собой плоскость. Обитатели этого мира — точки, отрезки, плоские фигуры. Глядя друг на друга, они могут видеть лишь отрезки, которыми являются проекции плоских фигур на прямую (см. рис.).



Трехмерный мир окружает нас. Мы привыкли к нему, объекты, окружающие нас трехмерны, хоть в каждый момент времени мы видим только их плоские проекции — как на фотографии. Если бы мы были четырехмерными, в трехмерном мире перед нами открылись самые фантастические возможности.

Например, рассматривая плоский мир рисунка, мы видим внутренность всех комнат, равно как и их обитателей и всего дома (см. рис.).



Вид на дом Квадрата из Трехмерия.

Аналогично мы увидели бы улицы, горы и пещеры, извилистые лабиринты и спрятанные тайники, небеса и подземелья — все открывается взгляду и видно насквозь. Один из героев книги — Сфера — является перед жителем плоского мира Флатландии и пытается доказать ему существование неизвестного тому третьего измерения. Для убедительности со словами «… легкая боль, которую вы, быть может, ощутите, не идет ни в какое сравнение с выгодой, извлекаемой из этого прикосновения вашим умом», трехмерная Сфера даже касается желудка двумерного квадрата».

Итак, из следующего измерения мы получаем возможность видеть насквозь все, происходящее в пространстве меньшей размерности, возможность беспрепятственно перемещать предметы сквозь границы, видоизменяться в пространствах меньшей размерности, появляясь в них тем или иным своим многомерным сечением, исчезать в одной точке и появляться в другой точке беспричинно и непредсказуемо для обитателей меньших размерностей. В этом, несмотря на всю непохожесть, наши пространства имеют много общего.

Сила и красота математики, ее абстрактность и универсальность состоят как раз в том, что она находит универсальные единые методы описания пространств, что используемый язык позволяет строить плодотворные модели и аналогии для самых непохожих на первый взгляд объектов.

1.2. Пространство как философское понятие

Все мы знаем и имеем представление о том, что такое пространство, но, тем не менее, оно не вещь среди вещей и не предмет среди предметов. Вся материя содержится в пространстве и связана с ним, но само пространство находится как бы за всем этим материальным, **оно есть некий фон, пустота, сцена, на которой разворачивается вся драма материального мира**. И здесь мы вынуждены констатировать тот факт, что об окружающих вещах мы знаем куда больше, чем об этом фундаменте. Мы способны ощущать предметы окружающего мира, но как ощутить само пространство? ***Мы видим предметы, но как увидеть пустоту, в которой эти предметы существуют*?**

Сложность понимания пространства становится препятствием на пути его изучения. По мнению П. Девиса, «Такое представление пространства как отсутствие предметов делает для многих непонятным, почему ученые создают различные теории о нем. Ведь если пространство — ничто, то и говорить о нем нечего!».

Тем не менее, несмотря на всю сложность восприятия пространства, нам хорошо известны его свойства. Мы знаем о его трехмерности, изотропности, обратимости и т.д., но можем ли мы говорить о видовом пространственном делении? Кто-то может вспомнить физическое, психологическое и социальное время и пространство, но это не виды пространства, а всего лишь особенности восприятия пространства на индивидуальном и социальном уровне. Но **возможно ли видовое деление самого пространства?** При этом придется изначально отказаться от теорий о каких-либо параллельных пространственных измерениях, речь идет о видах одного-единого пространства, а не о слоеном пространственном «пироге».

Если следовать единству пространства, то вряд ли можно говорить о какой-либо видовой его классификации. Следуя этой логике, пространство изначально едино и мы не можем в нем выделить какого-либо еще пространства, помимо него самого. Но прежде чем дать окончательный ответ, попытаемся (хоть это не просто) помыслить и представить себе пространство. Что мы мыслим при попытке осознать пространство?

Возможно, первое, что мы осознаем, это некую фундаментальную пустоту, некое вместилище материи. Но, тем не менее, мы должны отказаться от понимания пространства как подобия глобальной пустоты. Пространство не тождественно небытию, оно имеет свои свойства и, возможно, свои виды. Обратимся в таком случае к свойствам этого пространства.

Пространство «Между»:

Есть свойства пространства, которые его описывают, но есть свойство, описывающее то, что это пространство делает. Так, трехмерность, изотропность, обратимость пространства говорит нам о том, какое наше пространство, но его протяженность говорит нам о том, что основной его функцией является способность соединять и объединять в своем единстве все вещи и предметы материального мира. **Именно это его свойство является главным и основным, поскольку оно показывает, как «действует» пространство, в то время как остальные свойства показывают, в какой форме реализуется это действие.**

Вполне возможно построить даже иерархию свойств пространства, где одно его свойство существует на основе другого, более универсального.

Так,

протяженность говорит о том, что пространство объединяет и связывает в себе все сущее;

трехмерность говорит о том, что это синтезирующее действо осуществляется по трем направлениям (X, Y, Z);

изотропность говорит о равенстве этих общих направлений (X=Y=Z);

однородность же говорит о равенстве сторон внутри каждого из этих направлений (X=-X, Y=-Y, Z=-Z).

Самое главное свойство пространства — **протяженность**. Само пространство, как пишет А.М. Мостепаненко, рассматривалось : «…как единая для всего мира эвклидова трехмерная протяженность, а время — как абсолютная универсальная длительность, которая всюду протекает равномерно».

Протяженность говорит о том, что оно объединяет в себе существование всех входящих в него вещей. Именно пространство соотносит существование вещей в плоскости единого сосуществования. Все предметы могут взаимодействовать друг с другом исключительно потому, что они входят в единое поле пространства. Как утверждает В.И. Жарков, «… протяженность есть прежде всего выражение устойчивости определенного типа связи сосуществующих объектов».

Такое единящее все и вся пространство можно назвать пространством **«между»**, поскольку оно, пролегая пустотой между предметами, объединяет их в единый предметный мир. В этом мире две материальные точки сосуществуют и взаимодействуют, поскольку между ними пролегает общее, единящее их пустое пространство. Но все ли пространство можно назвать этим пространством «между», или кроме него есть иное пространство, которое не описывается этим «между»?

***Пространство «В»*:**

Продолжим свой мысленный эксперимент и помыслим некую материальную точку в пространстве, которая, как мы уже говорили, сосуществует с другими материальными точками. Пространство их сосуществования, как мы уже определили, есть пространство «между», но таким ли будет пространство, которое можно охарактеризовать как внутреннее пространство этой точки?

Каждый предмет можно представить как находящийся в пространстве, которое соединяет этот предмет с другими предметами, а также имеющий пространство внутри себя. Одно из этих пространств будет пространство «между», а другое — пространство «в». Но есть ли принципиальная разница между ними? На первый взгляд, никакой. Например, если рассматривать дом — вне или внутри него мы обнаруживаем то же самое пространство. Но не все так просто даже с самим домом. Ведь пространство внутри дома и снаружи не есть одно и то же, пространство внутри делает этот дом именно домом, то есть тем строением, которое ограждает нас от внешнего мира. Но что касается самого пространства, то здесь дело обстоит принципиальнее.

Что касается нашей точки, то пространство «между» пролегает между ней и иными точками, *но внутри нет никакого «между»*, там нет никаких иных точек, кроме нее самой. **Все части этой точки соотносятся в ней в плане ее самотождественности, а не в плане общего нахождения в пространстве «между».**

Но можно ли выделить это *пространство «в»* в качестве альтернативного вида пространственного направления?

Прекрасной иллюстрацией взаимодействия пространства «между» и «в» являются черные дыры, где геометрические свойства приобретают физический смысл. По своему физическому действию **<u>черная дыра</u> — это бесконечное *падение* вовнутрь черного тела, которое происходит благодаря *сжимающей* силе гравитации**. По своему геометрическому воздействию она *преломляет* пространство, причем это преломление образуется со стороны всех направлений трехмерного пространства. Преломление пространства здесь происходит относительно *любой* точки внешнего пространства и относительно *всех* точек, вместе взятых.

Прямолинейное движение здесь лишь иллюзия и как бы верным ни казалось нам, что точка с ее «внутри» находится прямо перед нами, и что попасть внутрь нее можно, следуя прямому направлению по отношению к этой точке, — это всего лишь иллюзия, такая же иллюзия, как движение по прямой вблизи массивного тела. Мы уверены, что путь, совершаемый нами вблизи массивного тела, есть прямая, но, как показала теория относительности Эйнштейна, гравитация любого тела, имеющего массу, искривляет окружающее пространство. **Движение около такого тела не может быть прямолинейным.**

Если же мы посмотрим на наше пространство, то увидим, что нет никакой разницы, если мы будем двигаться прямо, вбок, вверх или вниз, то есть по всем трем измерениям — так или иначе мы совершаем одинаковое движение. Но движение вовнутрь принципиально отлично от движения по всем измерениям, это движение направлено в иное измерение. Можно было бы сказать, что *внутрь* — некое четвертое измерение, но это неверно, поскольку *относительно этого движения все остальные три измерения одинаковы*, поэтому мы можем говорить не об отношениях измерений, а об отношениях двух видов пространств.

Так, если обычное пространство пролегает по трем своим направлениям, то **плоскость пространства «в» альтернативна всем этим направлениям, причем одновременно.** Говоря языком геометрии, она образует угол ко всем трем направлениям пространства *одновременно*. Так, если взять классические обозначения пространственных направлений X, Y, Z, то каждое из этих направлений определяется как плоскость, лежащая под углом к двум другим. Каждое из этих направлений альтернативно по отношению к двум другим. Пространство же «в» способно находиться под углом ко всем этим трем направлениям, и оно способно быть альтернативным всему набору этих направлений. Если бы направление пространства «в» находилось под углом только к двум направлениям трехмерного пространства (допустим, X и Y), то оно совпало бы с третьим направлением (в нашем случае Z), но непокорное направление пространства «в» образует угол ко всем направлениям X, Y, Z.

В нашем пространстве положение точки характеризуется соотношением ее положения относительно всех трех направлений X, Yи Z, только сообща координаты данных направлений рисуют нам точку в пространстве. Так, если бы гравитационное искривление пространства происходило по двум направлениям, то положение точки укладывалось бы в рамки трех измерений и движение происходило бы по этому направлению. Но *в природе нет приоритета в пространственных направлениях, все измерения равны друг перед другом, и нет оснований считать какие-либо направления приоритетными при искривлении пространства*. **Искривление происходит по всем направлениям одновременно и равнозначно.**

О несовпадении пространственных координат пространства «в» может говорить число пи. Круг и круговое ускоренное центростремительное движение, по сути, является отражением взаимодействия двух видов пространства, где соотношение радиуса к окружности имеет бесконечное значение. Получается, что направление, идущее к центру круга, не может соотнестись с тем пространством, которое создается этим направлением. В то же время соотношение величин в трехмерном пространстве строго соотносятся между собой подобно тому, как в прямоугольном треугольнике квадрат гипотенузы соотносится с суммой квадратов катетов.

В результате соотношения координат такого искривленного пространства наша точка будет постоянно уменьшаться. **Уменьшение возникает в результате такого соотнесения пространственных координат, пытающихся как-то определить положение точки в искривленном пространстве**. Координаты точки оказываются смещенными относительно всех направлений трехмерного пространства, и при соотношении самих этих направлений происходит их уменьшение. Уменьшение также означает стремление пространства «в» выйти за пределы трехмерного пространства «между».

Более того, такое преломление пространства способно «нарисовать» в пространстве точку. Такой контраст пространственных направлений вырывает из нашего пространства некое иное бытие, которое противостоит пустоте окружающего пространства. **Наше трехмерное пространство, как уже было сказано, однородно, и эта однородность подобна пустоте, в которой ничего не может быть выделено. Пространство «в» противопоставляет этой однородной пустоте свое альтернативное <u>искривленное пространство</u>**. Она создает угол с направлениями пространства «между», который показывает, что данная точка отлична от всего остального пространства. **Именно этот угол делит пространство на внешнее и внутреннее, и если бы не существовало такового деления, не существовало бы отдельных предметов и все сливалось бы в единой плоскости существования**. **Не существовало бы границ между отдельными предметами или точками пространства**

Есть еще одно наиболее важное отличие движения по направлениям пространства «между» и «в», которое показывает принципиальную разницу этих видов пространств. Дело в том, что **<u>движение в трехмерном пространстве происходит по одному направлению</u> — слева направо, сверху вниз, от меня или на меня, <u>движение же внутрь происходит одинаково со всех сторон</u>. Сжимающаяся точка удаляется от меня со всех сторон одинаково, откуда бы я на нее ни смотрел.** Это не применимо к пространству «между», где всякое движение есть движение относительно двух точек пространства, где движение «от» одной точки есть движение «к» другой точке. **Движение же пространства «в» — это движение равноудаления со всех сторон относительно пространства «между».** Здесь нет никакого «вправо», «влево», «вверх», «вниз», есть только одно направление — «от».

Так где же находится это пространство «в», если оно не укладывается в рамки трехмерности нашего пространства? Создается впечатление, что оно оказывается за рамками самого пространства. Но, тем не менее, **пространство «в» выходит за рамки не самого пространства, а только за рамки его трехмерности**. Три измерения в целом образуют направления нашего пространства «между», которому противостоит направление искривленной плоскости пространства «в». При этом в качестве направлений мы не берем трехмерность, поскольку вся трехмерность укладывается в один вид движения пространства «между». **Трехмерность — это система однотипных направлений относительно одной точки пространства, где все эти «вправо», «влево», «вверх», «вниз», «от» и «к» показывают движение данной точки относительно иных точек пространства. Направление же пространства «в» противостоит всей этой трехмерности пространства «между» разом. Оно есть некое альтернативное направление, не укладывающееся в систему направлений пространства «между» и выходящее за рамки этой трехмерности направлений пространства «между». Это делает пространство «в» иным, отличным пространством относительно пространства «между».**

Но почему мы склонны говорить о разных видах пространства? Не проще ли было бы говорить о разных направлениях движения в пространстве? Эти виды пространств отличны, поскольку принципиально отличны направления этих пространств. **Направления пространства — это самый главный и определяющий признак пространства, если характер направлений пространств различен, различны и сами виды пространства**.

Движение.

Различными являются не только виды пространства «в» и «между», но и сами направления этих пространств. Рассмотрим их отличия.

Если взять направления пространства «между», то они уравновешивают друг друга как две чаши весов. **Направления этого вида пространства имеют зеркальную поляризацию, то есть если мы представим одно направление относительно какой-либо точки, то противоположное ему направление будет гармонично**

противопоставлено этому направлению как его зеркальная копия. ***Эти направления уравновешивают друг друга*.**

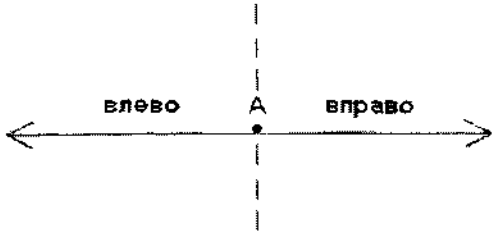


Рис. 1

Как показано на рис. 1, направления «вправо» и «влево» относительно точки А являются моментом стабильности и равновесия самой этой точки. **Изотропность пространства создает условие, по которым данная точка будет находиться в уравновешенном состоянии**. То есть **направления пространства «между» создают условия статичности и являются <u>основой состояния покоя этой точки</u>**. Но так ли это будет с направлениями пространства «в»?

Что касается направлений пространства «в», то здесь нет никакого равновесия. Направления пространства «в» могут располагаться двояким образом относительно материальной точки так, как это показано на рис. 2, но ни в одном из этих случаев нет равновесного состояния пространства, какое мы находим в направлениях пространства «между».

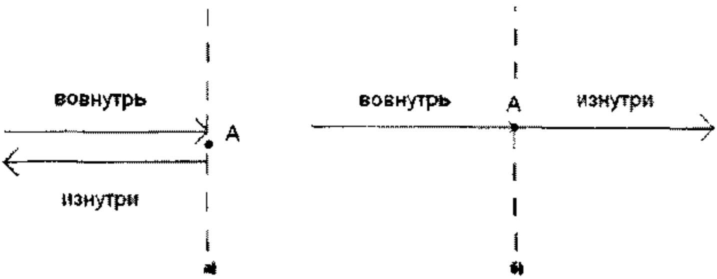


Рис. 2

Согласно рис. 2а, направления «вовнутрь» и «изнутри» направлены в противоположные стороны, но эти направления занимают только одну сторону относительно точки А. Что касается рис. 2б, в данной конфигурации направлений направления «вовнутрь» и «изнутри» имеют одно направление движения — они направлены только в одну сторону. Что следует из положения о неравновесности направлений пространства «в»? В данном случае мы можем видеть, что пространство «в» принципиально отлично от пространства «между» по своим свойствам. В отличие от пространства «между» пространство «в» *не изотропно*. Из этого следует, что основой статичности или покоя материальной точки является равновесие направлений пространства «между», **неравновесие же направлений пространства «в» является основой <u>движения материальной точки</u>, а напряжение между видами пространства реализуется в качестве <u>поля энергетического напряжения</u>. <u>Нарушение равновесия в физике непременно ведет к возникновению движения</u>**, тем более что в данном случае равновесие нарушает само пространство.

Это подобно тому, как если бы с одной стороны на точку давило бы пространство бытия, а с противоположной стороны ее бы засасывало в своеобразный пространственный вакуум. **Лишенное изотропности направлений пространство будет двигать эту точку в определенном направлении или же двигаться само относительно этой материальной точки.**

Но правомочно ли рассматривать пространственные направления в качестве основы существования движения?

Момент покоя и движения, как мы знаем из теории относительности, — явление относительное и определяется ни чем иным, как инерциальной системой отсчета, то есть пространственными координатами, связанными с материальной точкой в этом пространстве. **«Убрать» движение, так же как и «породить» его, возможно путем смены пространственных систем отсчета**. Как показала теория относительности, движение непосредственным образом связано с пространством, но только снова возникает вопрос о приоритете пространства и движения. И если теория относительности говорит о зависимости пространства от движения, то вполне возможна и обратная зависимость. Тем более что **<u>движение</u>, как фундаментальная категория, может появиться только в результате нарушения равновесия более фундаментальной категории, которой может быть только пространство. <u>Энергия же в этом случае является состоянием напряжения между «сломанными» плоскостями пространства.</u>**

1.3. Виды и типы пространств

Простра́нство на уровне повседневного восприятия интуитивно понимается как место, в котором возможно движение, различные положения и взаимные расположения объектов, отношения близости-дальности, понятие направления, как арена событий и действий, универсально содержащая все места и вмещающая объекты и структуры; иногда — как специфическое место, в значительной мере определяющее сущность происходящих в нём событий.

С геометрической точки зрения, термин «пространство» без дополнительных уточнений обычно обозначает трёхмерное евклидово пространство.

1. Абстрактное пространство. Сначала приведем общее определение пространства, а затем попробуем уяснить его содержание. В современной математике **пространство определяется как множество однородных объектов (предметов, явлений, состояний, переменных и т.п.), между которыми имеются пространственно подобные отношения.** Часто слова «однородные» и «пространственно подобные» опускают и определяют пространство как кортеж $(M, A_1, A_2, ..., A_n)$, где M — некоторое множество, а $A_1, A_2, ..., A_n$ - отношения между его элементами. **Иногда о пространстве говорят просто как о множестве *M*, между элементами которого подразумеваются некоторые отношения.**

Столь широкое понятие пространства сформировалось в результате абстрагирования и обобщения трехмерной евклидовой геометрии (геометрия Лобачевського и другие неевклидовы геометрии, различные геометрические преобразования - проективное, аффинное, конформное, топологическое и т.п.), развития понятия числа (комплексные числа, кватернионы и гиперкомплексные числа), а также стремление использовать геометрический язык и пространственные представления для соотношений с любым количеством переменных (подобно аналитической геометрии и векторной алгебре).

2. От трехмерного к многомерному пространству. Положение точки обычного трехмерного пространства в некоторой системе координат определяется тройкой чисел (x, y, z), называемых *ее координатами* (рис. 3).

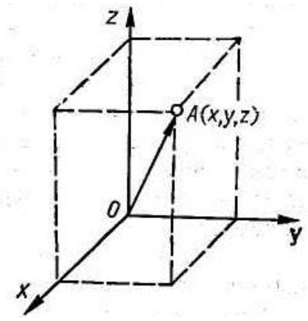


Рис. 3. Вектор в трехмерном пространстве.

Каждой точке соотносится пространственный *вектор,* который выходит из начала координат и заканчивается в этой точке. Числа x, y, z есть *проекциями* вектора на оси координат и называются *компонентами (составляющими)* вектора. Как и точка, вектор полностью определяется этой тройкой чисел, если строго придерживается порядок их следования, т.е. $\boldsymbol{a}=(x, y, z)$. Итак, **между точками и векторами пространства устанавливается взаимо-однозначное соответствие.** Поэтому в зависимости от удобства **можно говорить о пространстве как о множестве точек или векторов.** Так, многие физических величин (силы и скорости, напряженности электрического и магнитного полей и др.) представляются векторами, а различные фигуры удобно рассматривать как геометрические места точек, которые удовлетворяют соответствующим соотношением.

Действия над векторами сводятся к операциям над тройками чисел. Так, если

$$\boldsymbol{a}=(x, y, z) \text{ и } \boldsymbol{b}=(x', y', z'), \text{ то } \boldsymbol{a}+\boldsymbol{b}=(x+x', y+y', z+z')$$

и

$$\alpha\boldsymbol{a}=(\alpha x, \alpha x, \alpha z)$$

где α - некоторое число (скаляр).

Длина вектора

$$|a|=\sqrt{x^2+y^2+z^2}\ .$$

Расстояние между двумя точками пространства, соответствующими векторам $\boldsymbol{a}$ и $\boldsymbol{b}$, есть длина вектора $\boldsymbol{a}$-$\boldsymbol{b}$ и, следовательно,

$$|a-b|=[(x-x')^2+(y+y')^2+(z+z')^2]^{1/2}$$

Скалярное произведение двух векторов определяется соотношением

$$\boldsymbol{a}\text{-}\boldsymbol{b}=|a|\cdot|b|\cos\gamma=xx'+yy'+zz'),$$

где γ — угол между векторами $\boldsymbol{a}$ и $\boldsymbol{b}$. Отсюда

$$\cos\gamma = \boldsymbol{a}\cdot \boldsymbol{b}\,\frac{1}{|a||b|}\ .$$

Единичные векторы *(орты),* совпадающие по направлению с координатными осями, выражаются соответственно как

$$\boldsymbol{i}=(1, 0, 0), \boldsymbol{j}=(0, 1, 0), \boldsymbol{k}=(0, 0, 1).$$

Каждый вектор однозначно представляется через орты, которые образуют единичный базис в прямоугольной системе координат:

$$\boldsymbol{a}=(x, y, z)=x(1, 0, 0)+y(0, 1, 0)+z(0, 0, 1)= x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}.$$

Формальное обобщение трехмерного пространства состоит в том, что в качестве вектора принимается любая упорядоченная последовательность n чисел $a=(x_1, x_2, .., x_n)$, называемая n-мерным вектором или точкой многомерного пространства. Числа $x_1, x_2, ..., x_n$ называют *компонентами (составляющими, координатами)* n-мерного вектора, а множество таких векторов — *числовым* или *точечным векторным пространством.* Определяя соответствующим образом операции над векторами и задавая на множестве векторов отношения, подобный длине, расстоянию, углу и т.п. в обычном пространстве, получают специальные типы пространств.

До сих пор предполагалось, что количество составляющих вектора конечно и равно n. Ничто не мешает сделать следующий важный шаг на пути расширения понятия пространства: не ограничивать количество составляющих векторов и считать векторами любые (конечные или бесконечные) числовые последовательности $x=(x_1, x_2, ..., x_n, ...)$. Такими последовательностями выражаются, например, множества многочленов произвольной степени, ряды, различные разложения функций. Пространство, точки которого определяются бесконечными последовательностями, называется *бесконечномерным пространством.*

Более того, в качестве элементов пространства можно принять множество непрерывных функций на данном отрезке, совокупность всех решений дифференциального уравнения определенного типа и т.п. В подобных случаях функция $f(x)$ представляется как точка в пространстве, а ее координатами служит бесконечное множество (мощности континуума) значений функции при всевозможных значениях аргумента x. Пространства, элементами которых являются функции, называются *функциональными пространствами.*

Представление функций как объектов пространства есть одним из исходных положений важного раздела современной математики - функционального анализа.

3. От числовых к абстрактным пространствам. Другая линия обобщения пространства связана с содержанием понятия вектора.

Уже отмечалось, что в трехмерном пространстве с этим понятием связываются различные физические величины, которые характеризуются числовым значением и направлением. Наряду этим элементы пространства (векторы или точки) могут отождествляться с объектами любой физической природы. Например, в **практике широко используется трехмерное *цветовое пространство,* векторы которого соответствуют цветовым ощущениям и определяются тремя компонентами - <u>интенсивностями красного, зеленого и синего цветов</u>.**

Состояние физической системы описывается некоторой совокупностью переменных (тока и напряжения электрической цепи, температуры и концентрации веществ в химическом реакторе и т.п.). Каждое такое состояние можно представить вектором n-мерного пространства, называемого *пространством переменных состояний.*

В приведенных примерах объекты пространства характеризуются совокупностью чисел, и эти числа рассматриваются как составляющие соответствующих этим объектам векторов. Можно говорить об отображении множества объектов на множество векторов, но в конечном счете отношения между объектами пространства сводятся к отношениям на множестве векторов в числовых пространствах.

Но и в случаях, **когда объекты характеризуются свойствами, которые не являются числами (форма, цвет, материал), совокупность таких объектов можно также рассматривать как векторное пространство**. При этом свойства кодируются с помощью чисел или каких-нибудь символов, которые можно истолковывать как составляющие векторов (объектов) пространства. Подобные коды используют для передачи сообщений, обработки различной информации с помощью вычислительных машин и т.п. В простейших случаях кодирование свойств объектов сводится к простой нумерации, и каждый объект рассматривается как совокупность номеров присущих ему свойств. Подобный способ был использован нами, например, при рассмотрении отношения толерантности.

Наконец, можно говорить об объектах пространства как его «точках», вовсе не связывая эти объекты с обычным представлением о векторах, как последовательностях чисел, кодов или символов. **Пространство можно рассматривать как множество объектов в «чистом» виде (слова, понятия, люди, животные, детали механизма, компоненты электронной цепи). Операции на множестве таких объектов выполняются по специально устанавливаемым правилам, а отношения между ними выражаются в форме некоторого описания словесного или символического.** Так, множество объектов M с определенным на нем отношением толерантности τ можно рассматривать как *пространство толерантности (М,* τ), структура которого определяет сходство между объектами.

Что же остается при такой степени обобщения от первоначального понятия обычного трехмерного пространства? Если возвратить к определению, приведенному в (1), то найдем там слова «пространственно подобные отношения». Как уже указывалось, здесь имеются в виду обобщения таких отношений, как длина, расстояние, угол, фигура и т.п. Не следует искать слишком прямолинейного истолкования абстрактного пространства в категориях реального трехмерного мира. Это понятие введено математиками для того, чтобы использовать геометрические образы и терминологию для описания и изучение таких отношений, которые не допускают интерпретации в обычном трехмерном пространстве.

4. Метрические пространства. Рассмотрение множеств как совокупностей некоторых объектов имеет ограниченное применение, так как в природе все материальные объекты находятся во взаимной связи и взаимодействии. Поэтому понятие множества необходимо увязать с установлением тех или других отношений между его элементами.

Говорят, что множество имеет *структуру,* если между элементами множества установленные определенные отношения или над ними определены некоторые операции. Множество, которое наделено структура, еще называют *пространством.*

Изучение пространств мы начнем с простейшего вида пространств, называемых **метрическими пространствами**, для определения которых необходимо ввести понятие расстояния между элементами множества .

С понятием расстояния человек сталкивается повседневно, связывая это понятие с пространственным размещением предметов и понимая под расстоянием *меру удаленности предметов друг от друга*. Обычное расстояние *d(M, N)* между точками *M* и *N* измеряется длиной отрезка, который соединяет эти точки (рис. 4.).

Рис. 4. Иллюстрация аксиомы треугольника.

Однако такое определение расстояния часто оказывается недостаточным. Так, даже в быту расстояние между двумя городами определяется не однозначно (расстояние по железной дороге, расстояние по водному пути и т.п.). В городе, который разделен на кварталы, как показано на рис. 5, измерять расстояние отрезком прямой, которая соединяет точки *M* и *N,* не имеет смысла, так как двигаться можно только по улицам.

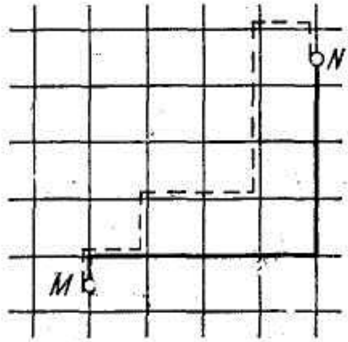


Рис. 5. Определение расстояния в городе, разделенном на кварталы.

С другой стороны, слово удаленность мы часто не связываем с пространством в обычном понимании. Так, близость двух клеточек на шахматной доске можно оценить числом ходов, которые нужно сделать, чтобы перевести шахматную фигуру с одной клеточки на другую. При этом две соседние клеточки близки для короля, но далеки для коня и являются бесконечно далекими, если из одной в другую нужно перевести слона.

В математике часто стит задача замены одной функции *y=f(x),* которая почему-либо неудобна для рассмотрения, другой функцией *y=g(x).* Такая замена возможна, если функция *g(x)* близка к функции *f(x)* (рис. 6,*а*).

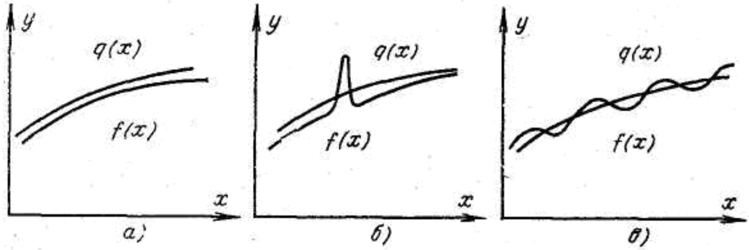


Рис. 6. Различные случаи близости двух функций.

При этом следует договориться, как понимать близость друг к другу двух функций. Если близость оценивать по максимальному отклонению значений *y* в этих функциях, то в случае на рис. 6, *б* функции уже нельзя считать близкими, а если учитывать еще и характер функции, например, кроме близости значений самых функций потребовать близости производных, то функции не будут близкими и в случае на рис. 6, *в*. Если же за условие близости принять ограничиваемые функциями площади, то во всех трех случаях функции близки друг к другу.

Из приведенных примеров видно, что должно существовать некоторое общее **определение расстояния как меры удаленности объектов**, а следовательно, и просторанства, в котором эти объекты существуют, причем в различных конкретных ситуациях эти понятия могут иметь разное содержание. Поскольку совокупности различных объектов представляют собой множества, то понятие пространства и расстояния должны быть связаны с понятием множества.

Определение метрического пространства

Пусть *X* — произвольное множество. Понятие расстояния между элементами из *X* получается путем обобщения фундаментальных свойств, которые можно интуитивно ожидать от понятия расстояния и которые легко могут быть понятны из рассмотрения рис. 2.

Свяжем с каждой парой элементов из *X* некоторое действительное неотрицательное число $d \geq 0$.

Это число называется *расстоянием* или *метрикой* в *X*, если для любых $y,x,z \in X$ оно удовлетворяет следующим трем условиям:

1) $d(x, y)=0$, причем тогда и только тогда, когда $x=y$ (аксиома идентичности);

2) $d(x, y)=d(y, x)$ (аксиома симметрии);

3) для любой тройки $y,x,z \in X$ имеет место

$$d(x, y) \leq d(x, z)+ d(z, y)$$

(аксиома треугольника или неравенство треугольника).

Метрическим пространством называется пара *(X, d)*, т.е. *множество X с определенной на нем метрикой* d. Элементы множества *X* называют точками метрического пространства *(X, d)*.

Из данного определения следует, что множество *X* только тогда превращается в метрическое пространство, когда в него *введена соответствующая метрика d(x, y).* Если в одном и том же множестве *X* ввести разлияные метрики, то получаться и различные пространства. Так, пространства, которые изображены на рис. 4 и 5, имеют в качестве элементов множества точек плоскости, но обладают различными метриками.

Примеры метрических пространств

Пусть x, y — любые элементы из множества R вещественных чисел. Множество R можно превратить в метрическое пространство, если определить расстояние между x и y по формуле

$$d(x, y)=|x- y|. \qquad (1)$$

Именно по этой формуле находят расстояние между точками действительной оси, которая является простейшим примером метрического пространства.

Как мы знаем, упорядоченные n-элементные множества удобно рассматривать как точки воображаемого n-мерного пространства R^n. Можно расширить представление о таком пространстве, введя понятие расстояния между отдельными точками, т.е. рассматривая это пространство как метрическое. Обозначим через $x=(x_1,..., x_n)$, $y=(y_1,...,y_n)$ отдельные точки пространства R^n. Пространство R^n можно превратить в метрическое несколькими различными способами.

Наиболее часто расстояние между точками x и y определяют по формуле

$$d_2(x,y)=\left\{\sum_{i=1}^{n}|x_i-y_i|^2\right\}^{1/2}. \qquad (2)$$

В случае n=2,3 это определение совпадает с обычным геометрическим понятием расстояния. Свойства 1, 2 и 3 для этого расстояния очевидны из рассмотрения рис. 4.

Метрика $d_2(x, y)$ называется евклидовой, а пространство R^n с такой метрикой называется евклидовым и обозначается E_n. (Более детально о евклидовом пространстве речь будет идти ниже).

Для множества R^n расстояние может быть определено и другими способами, например:

$$d_1(x,y)=\sum_{i=1}^{n}|x_i-y_i| \qquad (3)$$

или

$$d_\infty(x, y)=\max(|x_1- y_1|,...,|x_n- y_n|). \qquad (4)$$

Легко видеть, что метрики d_2, d_1, d_∞ являются частными случаями метрики

$$d_p(x,y)=\left\{\sum_{i=1}^{n}|x_i-y_i|^p\right\}^{1/p}$$

и получаются соответственно при p=2, p=1 и $p=\infty$.

Свойства 1 и 2 для метрик (3) и (4) очевидны.

Для доказательства свойства 3 введем в рассмотрение еще одну точку

$z=(z_1,\dots,z_n)\in R^n$...

Для расстояния $d_1(x, y)$ имеем:

$$d_1(x,y)=\sum_{i=1}^{n}|x_i-z_i+z_i-y_i|\leq\sum_{i=1}^{n}(|x_i-z_i|+|z_i-y_i|)=$$

$$=d(x,z)+d(z,y).$$

Для расстояния $d_\infty(x, y)$ свойство 3 проверяется следующим образом. Предположим, что $|x_k-y_k|$ - самая большая из соответствующих разностей точек x и y.

Тогда

$$d_\infty(x, y)=|x_k-y_k|=|x_k-z_k+z_k-y_k|\leq|x_k-z_k|+|z_k-y_k|.$$

Очевидно, что

$$|x_k-z_k|\leq\max(|x_1-z_1|,\dots,|x_1-z_1|)=d_\infty(x, z);$$
$$|z_k-y_k|\leq\max(|z_1-y_1|,\dots,|z_1-y_1|)=d_\infty(z, y).$$

Следовательно

$$d_\infty(x, y)\leq d_\infty(x, z)+d_\infty(z, y).$$

Рассмотрим множество всевозможных функций времени, непрерывных на интервале $a\leq t\leq b$. Пусть $x(t)$ и $y(t)$ — две такие функции. Расстояние между ними можно определить из соотношения

$$d(x, y)=\max_{a\leq t\leq b}|x(t)-y(t), \tag{5}$$

которое, как легко проверить, удовлетворяет всем свойствам метрики. Пространство с такой метрикой обозначается $C_{[a,b]}$.

Впервые строгое математическое построение теории непрерывных функций было проведено немецким математиком К. Вейерштрассом. Ему принадлежат две наиболее известные теоремы о непрерывных функциях, которые так и называют первой и второй теоремой Вейерштрасса. Обе эти теоремы можно объединить и кратко сформулировать так: каждая функция, непрерывная на отрезке, принимает в некоторой точке этого отрезка наибольшее значение. Иначе говоря, если функция $f(t)$ непрерывна на отрезке $[a, b]$, то существует точка $t_0\in[a, b]$ такая, что

$$f(t)\leqslant f(t_0)$$

для всех $t\in[a, b]$.

Значение $f(t_0)$ называют *максимумом* функции f на отрезке $[a, b]$ и обозначают так:

$$\max_{t\in[a,b]}f(t)=f(t_0).$$

Например,

$$\max_{t\in[0,\pi]}\sin t=\sin(\pi/2)=1.$$

Понятие максимума играет важнейшую роль в математике, особенно в теории оптимизации. Опираясь на свойства максимума, введем расстояние в пространстве $C[a, b]$. Функции, принадлежащие $C[a, b]$, т. е. непрерывные на отрезке $[a, b]$, будем обозначать не только символами $f(t)$, $g(t)$, но и символами $x(t)$, $y(t)$, $z(t)$.

Итак, пусть $x \in C[a, b]$ и $y \in C[a, b]$, т. е. функции $x(t)$ и $y(t)$ непрерывны на $[a, b]$. Тогда разность $x(t)—y(t)$ также непрерывна, и, следовательно, непрерывна функция

$$|x(t)-y(t)|.$$

Последнее вытекает из того, что абсолютная величина непрерывной функции есть непрерывная функция. По теоремам Вейерштрасса существует максимум этой абсолютной величины, который мы и назовем расстоянием от x до y. Таким образом,

$$d(x, y)=\max_{t \in [a, b]} |x(t)-y(t)|. \tag{6}$$

Графически расстояние в $C[a, b]$ иллюстрируется на рис. 7.

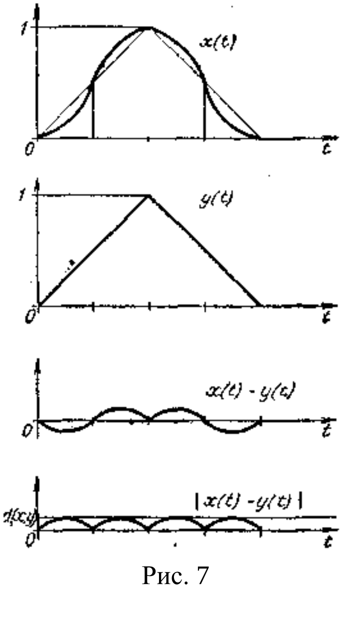


Рис. 7

Для того чтобы установить корректность этого определения, необходимо проверить выполнение условий 1°—3°:

1° $d(x, y)>0$, если $x \neq y$, и $d(x, y)=0$, если $x=y$;
2° $d(x, y)=d(y, x)$;

3° для любых трех элементов *x, y, z* из *M* справедливо неравенство треугольника

$$d(x, y) \leqslant d(x, z)+d(z, y).$$

То, что условия 1° и 2° выполняются, проверяется очень легко, а свойство 3°следует из неравенства

$$|x(t)-y(t)| \leqslant |x(t)-z(t)| + |z(t)-y(t)|.$$

Таким образом, формула (6) действительно определяет расстояние в пространстве $C[a, b]$, которое тем самым становится метрическим.

На этом мы закончим рассмотрение примеров и вернемся к изучению произвольных метрических пространств.

Пусть M — метрическое пространство; рассмотрим последовательность элементов $x_n \in M$. Понятие расстояния позволяет определить предел последовательности x_n.

Элемент $\overline{x} \in M$ называется *пределом* последовательности x_n, если расстояние от x_n до $\overline{x}$ стремится к нулю:

$$d(x_n, \overline{x}) \to 0 \text{ при } n \to \infty.$$

При этом говорят, что последовательность x_n *сходится* к $\overline{x}$ и пишут $x_n \to \overline{x}$ при $n \to \infty$. Другими словами, $x_n \to \overline{x}$, если числовая последовательность $d(x_n, \overline{x})$ является бесконечно малой. Иногда, подчеркивая, что сходимость понимается в смысле расстояния в метрическом пространстве, говорят, что последовательность сходится в *метрике M.*

Например, если в качестве метрического пространства взять числовую прямую R, то сходимость в метрике R совпадает с обычной сходимостью числовой последовательности. Действительно, в этом случае

$$d(x_n, \overline{x})=|x_n-\overline{x}|,$$

и если $d(x_n, \overline{x}) \to 0$, то $|x_n-\overline{x}| \to 0$.

Посмотрим, что означает сходимость в метрике пространства $C[a, b]$. Пусть $x_n \in C[a, b]$, $\overline{x} \in C[a, b]$ и $d(x_n, \overline{x}) \to 0$. Тогда для всех $t \in C[a, b]$

$$|x_n(t)-\overline{x}(t)| \leqslant d(x_n, \overline{x}),$$

и, следовательно, разность функций $-x_n(t)$ и предельной функции $\overline{x}(t)$ может быть сделана по абсолютной величине меньше любого положительного числа *сразу для всех точек t из отрезка* $[a, b]$. Такая сходимость последовательности функций была введена Вейерштрассом и названа равномерной сходимостью.

Итак, сходимость в пространстве $C[a, b]$ есть равномерная сходимость последовательности непрерывных функций на отрезке $[a, b]$.

Рассмотрим теперь задачу о приближенном решении уравнений в метрических пространствах.

Пусть даны два метрических пространства M и L. Предположим, что задано также **отображение T метрического пространства M в L**. Это значит, что **указано правило, по которому каждому элементу $x \in M$ сопоставляется определенный элемент Tx из L.** Фиксируем произвольный элемент $h \in L$ и рассмотрим уравнение

$$Tx=h. \qquad (7)$$

В этом уравнении T — заданное отображение, называемое также **оператором**, h — известная правая часть, x — неизвестный элемент, который требуется найти.

Для того чтобы показать, насколько широк класс уравнений вида (7), рассмотрим следующий пример. Пусть метрические пространства M в L совпадают с пространством функций, непрерывных на отрезке [0, 1], т. е. $M=L=C[0, 1]$. Определим отображение T равенством

$$Tx(t) = x(t) + \lambda \int_0^t (t-s) \sin x(s)\, ds, \qquad (8)$$

где λ — действительное число (параметр).

Интегральный оператор T каждой непрерывной функции $x(t)$ ставит в соответствие непрерывную функцию $Tx(t)_x$ т. е. действует из $C[0, 1]$в $C[0, 1]$. Уравнение (7) преобразуется к виду

$$x(t) + \lambda \int_0^t (t-s) \sin x(s)\, ds = h(t), \qquad (9)$$

где $h(t)$ — заданная функция, непрерывная на отрезке [0, 1]. Как видим, рассматриваемый нами класс уравнений (7) охватывает нелинейные интегральные уравнения (нелинейные, потому что подынтегральное выражение в (8) нелинейно зависит от неизвестной функции x).

Обычный путь получения уравнений — математическая формализация практических задач, т. е. построение математической модели. Не составляет исключения и интегральное уравнение (9).

С математической точки зрения правильная формализация должна быть такой, чтобы для уравнения (7) были справедливы теоремы о существовании и единственности решения при любой допустимой правой части.

Если нет теоремы единственности, то может оказаться, что уравнение (7) для некоторой правой части h имеет несколько решений. Как правило, это означает, что при составлении математической модели не были учтены какие-то важные факторы, характеризующие данное явление.

Приведем простейший пример подобного рода. Пусть требуется найти сторону квадратной комнаты площадью 18 м^2. Построение математической модели начинается с того, что реальный объект — пол комнаты заменяется абстрактной фигурой — квадратом. Если обозначить искомую длину стороны буквой x, то для ее определения получаем уравнение

$$x^2=18.$$

Правильно ли это уравнение приближает исходную задачу? Если его рассматривать на множестве всех действительных чисел R, то у него будет два решения; $x_1=\sqrt{18}$, $x_2=-\sqrt{18}$, т. е. единственности нет. Для выбора единственного решения вновь обратимся к исходной задаче. Из условия ясно, что ответ должен быть положительным. Следовательно, надо сузить множество, в котором ищется решение, и вместо всей числовой прямой R взять положительную полупрямую R_+ (т. е. множество всех положительных чисел). Теперь уравнение $x^2=18$ имеет ровно один корень, который и дает требуемое значение длины стороны.

Если нет теоремы существования, т. е. уравнение (7) не имеет решения, то обычно это означает, что мы предъявили слишком много требований к решению. Часто бывет достаточным расширить метрическое пространство, чтобы решение появилось.

Например, в той же задаче о нахождении длины квадрата по заданной площади мы могли бы вначале предположить, что искомая длина выражается рациональным числом. Тогда решение уравнения $x^2=18$ следовало бы искать в множестве положительных рациональных чисел. Положительный корень должен равняться $3\sqrt{2}$, но это число иррациональное. Следовательно, в множестве рациональных чисел рассматриваемое уравнение корней не имеет.

Расширим множество, в котором ищутся корни, и вместо рациональных чисел возьмем положительные действительные числа. Теперь уравнение $x^2=18$ имеет решение, и притом только одно.

Рассмотренная простейшая ситуация довольно точно характеризует трудности, возникающие в общей постановке задачи. Предположим, что уравнение (7), рассматриваемое в метрическом пространстве M, имеет единственное решение x_0. Удалим этот элемент x_0 из метрического пространства M. Оставшееся множество по-прежнему является метрическим пространством, которое мы обозначим символом M_0. Если теперь рассмотреть уравнение (7) в пространстве M_0, то оно уже решений иметь не будет. С абстрактной точки зрения пространства M и M_0 не отличаются, однако в первом уравнение имеет решение, а во втором — нет. Мы должны каким-то образом научиться различать эти пространства и исключать из рассмотрения те, которые получаются из «хороших» пространств удалением каких-то элементов.

Для решений этой задачи нам необходимо привлечь фундаментальные последовательности.

Последовательность элементов x_n метрического пространства называется фундаментальной, если

$$d(x_n,\ x_m)\to 0 \text{ при } n\to\infty \text{ и } m\to\infty.$$

Если в качестве метрического пространства M взята числовая прямая R, то это определение фундаментальной последовательности совпадает с тем, которое было дано ранее. Как и в случае числовой прямой, всякая сходящаяся последовательность элементов из M является фундаментальной. Это следует из неравенства треугольника:

$$d(x_n,\ x_m)\leqslant d(x_n,\ \bar{x})+d(\bar{x},\ x_m).$$

Если $x_n\to\bar{x}$, то оба слагаемых в правой части стремятся к нулю, поэтому стремится к нулю и левая часть т. е. последовательность x_n фундаментальна.

Как известно, в случае, когда метрическое пространство M есть числовая прямая, верно и обратное: всякая фундаментальная последовательность является сходящейся. Это утверждение составляет основное содержание критерия Коши.

Однако в некоторых метрических пространствах обратное утверждение может оказаться и неверным. Рассмотрим, например, множество всех рациональных чисел, которое обычно обозначают буквой Q. Введем в нем естественным образом расстояние от x до y:

$$d(x, y)=|x—y|.$$

Рассмотрим последовательность рациональных приближений x_n числа $\sqrt{2}$. Последовательность x_n сходится к $\sqrt{2}$ в пространстве действительных чисел R и поэтому является фундаментальной в R. Очевидно, она фундаментальная и в пространстве Q, однако в Q она уже не является сходящейся, так как число $\sqrt{2}$ множеству Q не принадлежит.

Вообще, удалим из числовой прямой R какую-нибудь точку a и обозначим полученное пространство M_a. Тогда любая последовательность отличных от a действительных чисел, сходящаяся к a, дает пример фундаментальной, но не сходящейся последовательности в M_a. Такая ситуация оказалась возможной вследствие того, что в метрическом пространстве M_a недостает элемента a, т. е. пространство неполно. Эти наводящие соображения приводят к следующему определению полного метрического пространства.

Метрическое пространство называется *полным,* если всякая фундаментальная последовательность его элементов имеет предел.

Классический пример полного метрического пространства — числовая прямая. Другой, более сложный пример — пространство $C[a, b]$ функций, непрерывных на отрезке $[a, b]$. Сходимость в метрике $C[a, b]$ совпадает с равномерной сходимостью, поэтому полнота пространства $C[a, b]$ следует из теоремы, доказываемой в курсе математического анализа: предел равномерно сходящейся последовательности непрерывных функций есть непрерывная функция.
Действительно, пусть x_n — фундаментальная последовательность функккций из $C[a, b]$. Тогда для любого фиксированного t из $[a, b]$ числовая последовательность $x_n(t)$ является фундаментальной в R и, согласно критерию Коши сходится. Ее предел, зависящий от t, обозначим через $\overline{x}(t)$:

$$\bar{x}(t) = \lim_{n \to \infty} x_n(t).$$

Так как равенство выполняется в каждой точке, то говорят, что последовательность $x_n(t)$ сходится к $\overline{x}(t)$ *по-точечно.* Докажем теперь, что она сходится равномерно.

Воспользуемся неравенством

$$|x_n(t) - x_m(t)| \leqslant d(x_n, x_m).$$

Последовательность x_n фундаментальна в $C[a, b]$, поэтому для любого положительного числа ε, сколь бы малым оно ни было, при достаточно больших номерах n и m будет выполняться неравенство

$$d(x_n, x_m) \leqslant \varepsilon$$

Поэтому при достаточно больших n и m неравенство

$$|x_n(t) - \bar{x}_m(t)| \leqslant \varepsilon$$

справедливо для всех $t \in [a, b]$.
Отсюда, устремляя m к бесконечности, получаем:

$$|x_n(t) - \bar{x}(t)| \leqslant \varepsilon$$

для всех $t \in [a, b]$. Ноэто и означает, что x_n сходится к $\overline{x}$ равномерно на отрезке $[a, b]$.

Теперь можно применить уже упомянутую выше теорему о непрерывности предела равномерно сходящейся последовательности непрерывных функций и заключить, что функция $\overline{x}$ непрерывна, т. е. $\tilde{x} \in C[a, b]$.

Обратим внимание на один существенный момент доказательства полноты $C[a, b]$. Первый шаг заключался в том, что, взяв произвольную фундаментальную последовательность $x_n \in C[[a, b]$, мы нашли функцию $\overline{x}(t)$, к которой функции $x_n(t)$ сходятся поточечно. Это утверждение еще далеко от требуемого: необходимо доказать, что функция $\overline{x}(t)$ непрерывна и что x_n сходятся к $\overline{x}$ равномерно. Однако именно этот первый шаг доказательства часто бывает одним из самых трудных — требовалось найти хоть какую-нибудь функцию (непрерывность была установлена позже), к которой данная последовательность сходится в каком-либо смысле (а затем было доказано, что сходнамость равномерная).

Следует обратить также внимание на то, что для доказательства существования функции $x(t)$ мы воспользовались критерием Коши сходимости числовых последовательностей. Таким образом, полнота пространства $C[a, b]$ является следствием полноты множества действительных чисел.

2. Расстояния в пространстве

На основе понятия расстояния можно изучать задачи о кратчайших путях на поверхностях, геометрические свойства многомерных пространств, методы помехоустойчивого кодирования сообщений, методы «сглаживания» результатов измерений и др.

На примере «расстояния» видно, какую роль в математике играет создание общих понятий, находящих порой самые неожиданные применения и связи. Кроме «расстояния», можно было бы еще указать на понятия «функции», «предела», «пространства», «преобразования» и менее известные в широкой аудитории понятия «изоморфизма», «группы», «кольца» и др. Среди этих понятий «расстояние» — одно из наиболее важных в математике.

Начальные разделы раскрывают, как происходит переход от обычного геометрического определения к общему понятию «расстояния» и что это «понятие» означает в различных конкретных случаях.

Далее описывается так называемое пространство сообщений, играющее важную роль в теории информации и общей теории связи.

Следующий параграф посвящен описанию методов кодирования сообщений, при которых сообщение оказывается устойчивым к ошибкам, возникающим в процессах передачи. Так как во всех реальных средствах связи время от времени возможны ошибки, то эти методы кодирования чрезвычайно существенны для систем связи и управления. Так, при передаче с борта космической ракеты на Землю фотографии обратной стороны Луны использовались помехоустойчивые методы кодирования этого сообщения. Необходимо отметить, что основная идея этих методов состоит в надлежащем использовании свойств расстояния в пространстве сообщений.

Рассматриваются возможные обобщения понятия расстояния. Показано, что далеко не всякое обобщение является содержательным, т. е. обладающим интересными свойствами. Сформулировать хорошее обобщение какого-либо математического понятия на самом деле не просто. В основе содержательного обобщения всегда лежат какие-то существенные свойства реального мира. В частности, важность понятия расстояния состоит в том, что целый ряд свойств многих реальных объектов связан с их взаимным расположением, которое часто можно охарактеризовать надлежащим образом определенным расстоянием. Так, например, хотя электроны в оболочке атомов нельзя представлять себе в виде материальных точек, все же в квантовой механике можно особым образом определить «расстояние» между различными состояниями электронов в атоме.

2.1. К определению понятия расстояние

Что такое расстояние, знает, казалось бы, каждый старшеклассник. Да и человек, основательно забывший школьную геометрию и не умеющий четко сформулировать определение расстояния, скорее всего будет утверждать, что ему хорошо понятно, что это такое.

Вместе с тем дело обстоит гораздо сложнее.

Слово «расстояние» может иметь разный смысл в зависимости от того, о каком пространстве идет речь. Сейчас мы увидим, что так обстоит дело даже в привычных для нас ситуациях.

На плоскости и в обычном трехмерном (так называемом евклидовом) пространстве расстояние между двумя точками M и N определяется как длина отрезка MN, соединяющего эти точки. Однако когда мы имеем дело с расстояниями между географическими объектами, на поверхности земли, то мы обычно имеем в виду длину дуги большого круга, соединяющего эти объекты. Разница между этими двумя видами расстояний особенно наглядна, если вычислить расстояния между северным N и южным S полюсами (см. рис. 1).

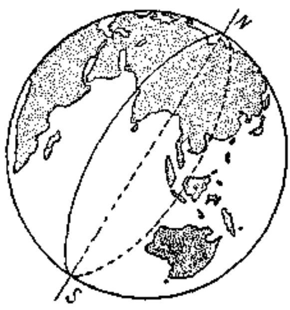


Рис. 1

Обычное (евклидово) расстояние между полюсами равно диаметру земли, т. е. около 12 800 *км.* Расстояние между полюсами по поверхности земли больше первого в $\frac{\pi}{2}$ раз, т. е. равно примерно 20 100 *км.* К этому примеру можно добавить, что в житейском обиходе при оценке расстояния между городами учитывается еще и способ передвижения. Так, расстояние по железной и шоссейной дороге между двумя пунктами может быть различным.

Новый пример расстояния мы получим, если будем рассматривать пункты на пересеченной местности, определяя расстояние между двумя пунктами как время, необходимое пешеходу для перехода из одного пункта в другой.

Ясно, что это расстояние не имеет ничего общего с длиной отрезка, соединяющего соответствующие пункты, так как путь по прямой, вообще говоря, не является наилучшим.

Вместе с тем любой путник исчисляет расстояние между пунктами своего маршрута по времени перехода.

Однако, несмотря на различие всех этих примеров, видно, что слово «расстояние» всюду употреблено в сходном смысле. **Это слово везде означает меру удаленности каких-то объектов.**

Итак, можно предположить, что есть некое общее определение расстояния, которое в разных конкретных ситуациях может иметь разное содержание. Такое общее определение будет сформулировано в позже.

А сейчас мы обсудим, какие требования надо предъявлять к определению математического понятия.

Современная математика — это язык естествознания. В основе наиболее важных математических образов лежат какие-то пространственно-временные формы мира, в котором мы живем.

Однако отношение между этими формами и математическими образами является очень сложным.

В любом разделе математики есть какие-то основные первоначальные понятия, которые в нашем сознании связаны с теми или иными физическими образами. Некоторые основные свойства этих понятий формулируются в виде **аксиом (и постулатов), т. е. <u>истин</u>, которые не доказываются, а принимаются в качестве исходных**. **Все остальные положения данного раздела математики логически выводятся из этих аксиом без дополнительных ссылок на свойства физического мира. Сами формулировки аксиом отражают какие-то интуитивно очевидные свойства этих понятий, соответствующие очевидным свойствам их физических прообразов.**

В геометрии такими основными понятиями являются: точка, прямая, плоскость, пространство и т. п. В систематических курсах геометрии можно найти перечень основных свойств этих объектов, выраженных в виде так называемых аксиом и постулатов, на основе которых строится все здание геометрии (Впервые так построил курс геометрии древнегреческий математик Езклид (IV—III век до н. э.).)

В алгебре такими основными понятиями являются <u>множество чисел и действия над числами</u>. В школе последовательно изучаются действия над целыми, дробными (или рациональными) относительными, вещественными (рациональными и иррациональными) числами и, наконец, комплексными числами.

При этом в каждом из перечисленных пяти случаев проверяется выполнимость основных законов (свойств действий над числами). Это — переместительный (коммутативный) закон для сложения $(a+b=b+a)$, сочетательный (ассоциативный) закон для сложения $((a+b)+c=a+(b+c))$, переместительный (коммутативный) закон для умножения $(ab=ba)$, сочетательный (ассоциативный) закон для умножения $((ab)c=a(bc))$, распределительный (дистрибутивный) закон $(a+b)c=ac+bc$ и правила: $a-a=0$; $a\cdot\frac{1}{a}=1$, характеризующие связь между основными (сложение и умножение) и обратными (вычитание, деление) операциями. Все эти законы выполняются во всех перечисленных числовых системах. Однако в некоторых из них деление выполнимо всегда, а в некоторых нет. **Если числовая система состоит только из положительных чисел, то в ней <u>не всегда возможно и вычитание</u>.** Оказывается, что правила алгебраических преобразований различных выражений основаны только на перечисленных выше свойствах. Все правила решения уравнений первой степени и систем таких уравнений также основаны на этих законах и возможности выполнения деления.

Оказывается, что свойства различных числовых систем можно изучать, исходя из общего понятия системы величин, над которыми определены операции (называемые сложением и умножением), удовлетворяющие перечисленным выше свойствам. **Такие системы называются в алгебре *кольцами* или *полями* (в зависимости от того, можно ли всегда в этой системе выполнять деление).**

Можно рассматривать правила преобразования выражений и правила решения уравнений в произвольном поле или кольце и как частный случай получить отсюда всю школьную алгебру.

В алгебре рассматриваются кольца и поля, являющиеся обобщением изучаемых в школе числовых систем. При этих обобщениях в основу кладутся основные свойства действий, проверенные для целых или дробных чисел, и изучаются факты, вытекающие логически только из этих свойств.

При таком подходе математика интересует не только открытие новых свойств физического мира и установление связей между его свойствами, но и выяснение свойств «воображаемых» миров, вытекающих из каких-то положенных в основу представлений.

Эта черта математики не менее важна, чем возможность описания физического мира. Замечательный русский ученый Н. И. Лобачевский, изменив один из постулатов Евклида, создал так называемую **«воображаемую» геометрию**. **Эта геометрия, спустя очень долгое время, легла в основу новых физических представлений о мире, возникших в результате открытия А. Эйнштейном теории относительности.**

Здесь мы рассматриваем одно из важных математических понятий — понятие «расстояния». Сначала мы проанализируем, какие основные законы, связанные с расстоянием, имеют место в элементарной геометрии. Положив в основу эти законы, мы получим определение так называемого **«метрического пространства»** и изучим различные примеры таких пространств. Мы увидим, что такой специфичный для математики подход к изучению различных объектов с точки зрения какого-то общего понятия открывает много интересных фактов.

Этот подход — создание новых общих понятий и попытка описать разнообразные объекты с помощью этих понятий — наиболее характерен для математики и ее приложений. С этой точки зрения «расстояние» является удобным примером для того, чтобы убедиться в плодотворности такого подхода.

2.2. Понятие расстояния и его свойства

К общему определению расстояния мы придем, обобщая свойства обычного расстояния в трехмерном пространстве. Поэтому мы сначала сформулируем основные свойства обычного расстояния.

Расстояние между двумя точками M и N в трехмерном пространстве— длину отрезка MN — принято обозначать $\rho\,(M, N)$.

Это обозначение подчеркивает, что расстояние есть вещественное число, определяемое точками M и N. Иначе говоря, расстояние есть числовая функция от пары точек. Если каждую из точек характеризовать тройкой координат $M(x, y, z)$ и $N(x, y, z)$, то расстояние в пространстве есть функция шести переменных

$$\rho\,(M, N)= F(x, y, z, x_1, y_1, z_1).$$

С помощью рис. 2 можно получить явное алгебраическое выражение для этой функции.

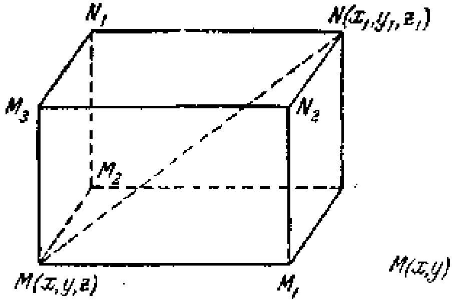


Рис. 2.

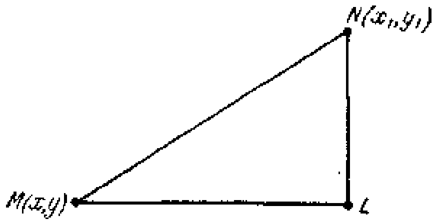


Рис. 3.

На этом рисунке изображен параллелепипед со сторонами, параллельными осям координат. Известно, что квадрат диагонали параллелепипеда равен сумме квадратов диагоналей его сторон.

Следовательно,

$$MN^2 = MM_1^2 + MM_2^2 + MM_3^2 = (x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2.$$

Отсюда

$$\varrho\,(M,\ N) = \sqrt{(x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2}\,. \qquad (1)$$

Еще проще вычисляется расстояние между точками $M(x, y)$ и $N(x_1, y_1)$ на плоскости (см. рис. 3). Для этого надо заметить, что длина отрезка ML равна $|x - x_1|$, а длина отрезка LN равна $|y-y_1|$. По теореме Пифагора

$$MN^2 = ML^2 + LN^2,$$

откуда

$$\varrho(M,\ N) = \sqrt{(x - x_1)^2 + (y - y_1)^2}. \qquad (2)$$

Несмотря на важность соотношений (1) и (2), нужные нам свойства расстояния могут быть получены без использования координатной системы.

Эти свойства можно записать следующим образом:

1. $\varrho(M,\ N) = \varrho(N,\ M)$ (симметричность),

2. $\varrho(M,\ N) \geqslant 0$ (неотрицательность).

3. $\varrho(M,\ N) = 0$ в том и только в том случае, когда точки M и N совпадают (невырожденность).

4. $\varrho(M,\ N) \leqslant \varrho(M,\ L) + \varrho(L,\ N)$ — для любой тройки точек M, L, N (аксиома неравенства треугольника).

Свойства 1, 2 и 3 очевидны. Они означают просто, что длина отрезка MN не зависит от порядка, в котором берутся точки M и N, всегда неотрицательна и равна нулю, только если начало и конец отрезка MN совпадают.

Свойство 4 становится очевидным, если через точки M, N и L провести плоскость и рассмотреть в этой плоскости треугольник MLN (см. рис. 4).

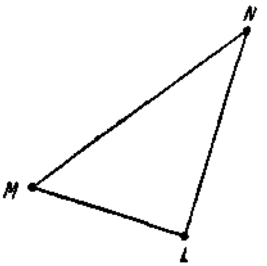


Рис. 4.

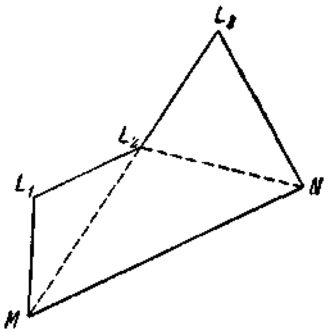


Рис. 5.

Тогда свойство 4 означает просто, что длина стороны MN не превосходит суммы длин остальных сторон треугольника MLN, откуда и происходит название свойства 4. Это свойство означает, что прямолинейный отрезок MN является кратчайшим путем, соединяющим точки M и N.

В самом деле, неравенство треугольника буквально означает, что длина отрезка MN меньше длины двухзвенной ломаной MLN, если точка L не лежит внутри этого отрезка. Отсюда можно вывести, что длина отрезка MN меньше длины ломаной с любым числом звеньев, соединяющей точки M и N. Для этого (см. рис. 5) будем последовательно уменьшать число звеньев ломаной на единицу, спрямляя по два звена. Каждый раз при этом длина ломаной будет уменьшаться, пока мы не дойдем до отрезка MN. Так, на рис. 5 мы переходим от ломаной $ML_1L_2L_3N$ к ломаной ML_1L_2N, затем к ломаной ML_2N, а затем к отрезку MN. Каждый раз длина ломаной уменьшается, а, значит, длина исходной ломаной больше длины отрезка MN.

Отметим, что в этом рассуждении мы пользовались только неравенством треугольника. Действительно, составим такую таблицу:

Ломаная	Ее длина	Используемое неравенство треугольника
$ML_1L_2L_3N$	$\varrho(M, L_1)+\varrho(L_1, L_2)+$ $+\varrho(L_2, L_3)+\varrho(L_3, N)$	$\varrho(L_2, L_3)+\varrho(L_3, N)>$ $>\varrho(L_2, N)$
ML_1L_2N	$\varrho(M, L_1)+\varrho(L_1, L_2)+$ $+\varrho(L_2, N)$	$\varrho(M, L_1)+\varrho(L_1, L_2)>$ $>\varrho(M, L_2)$
ML_2N	$\varrho(M, L_2)+\varrho(L_2, N)$	$\varrho(M, L_2)+\varrho(L_2, N)>$ $>\varrho(M, N)$
MN	$\varrho(M, N)$	

Из этой таблицы видно, как, подставляя в суммы, стоящие во втором столбце, неравенства из третьего столбца, мы приходим, наконец, к выводу, что

$$\varrho(M, L_1)+\varrho(L_1, L_2)+\varrho(L_2, L_3)+\varrho(L_3, N)>\varrho(M, N).$$

Если воспользоваться дополнительно тем фактом, что длина кривой есть предел длин вписанных в эту кривую ломаных, то можно доказать следующее утверждение.

Среди всех линий, соединяющих точки M и N, отрезок MN имеет наименьшую длину.

Из неравенства треугольника следует, что

$$\varrho(L, N)\geqslant\varrho(M, N)-\varrho(M, L). \qquad (3)$$

Неравенство треугольника показывает, что если точка M близка к точке L, а точка L близка к точке N, то точки M и N тоже близки. Подчеркнем, что в свойстве 4 неравенство превращается в равенство в том и только том случае, когда точки M, L и N лежат на одной прямой, а точка L находится между M и N (принадлежит отрезку MN).

Рассмотрим теперь расстояние на поверхности сферы S радиуса R. Расстояние между двумя точками M и N на поверхности сферы определим как длину меньшей дуги большого круга, проходящей через точки M и N. Напомним, что большим кругом на сфере называется окружность, лежащая на поверхности сферы с центром в центре сферы.

Большой круг лежит в плоскости, проходящей через точки *M, N* и *O* (центр сферы). Отсюда следует, что для несовпадающих точек *M* и *N* большой круг определяется единственным образом, поскольку для него заданы центр и две точки на окружности. Определенное таким образом расстояние $\varrho_S(M,\ N)$, очевидно, удовлетворяет свойствам 1, 2 и 3. Нетрудно также видеть, что для любых точек на сфере

$$\varrho_S(M,\ N)\leq\pi R, \tag{4}$$

причем равенство достигается только для точек *M* и *N*, лежащих на концах общего диаметра сферы (например, северный и южный полюсы).

Для того чтобы проверить свойство 4, надо рассмотреть сферический треугольник *MLN* (см. рис. 6).

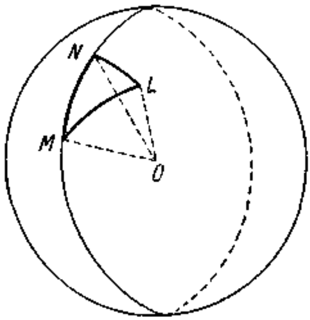


Рис. 6.

На этом рисунке точка *O* есть центр сферы. Ясно, что

$$\varrho_S(M,\ N)=R\alpha,\ \varrho_S(L,\ N)=R\beta,\ \varrho_S(L,\ M)=R\gamma,$$

где

$$\alpha=\angle MON;\ \beta=\angle LON,\ \gamma=\angle LOM.$$

Известно, что в трехгранном угле каждый плоский угол не превосходит суммы двух других плоских углов, т. е.

$$\alpha\leqslant\beta+\gamma. \tag{5}$$

Умножая обе части этого неравенства на радиус *R*, получим

$$R\alpha\leqslant R\beta+R\gamma.$$

Это и означает, что

$$\varrho_S(M,\ N)\leqslant\varrho_S(M,\ L)+\varrho_S(L,\ N), \tag{6}$$

что и требозалось доказать.

Таким образом, для сферического расстояния $\varrho_S(M,\ N)$ выполнены все основные свойства обычного расстояния.

Нетрудно проверить, что неравенство (6) обращается в равенство, только когда точка L расположена на том же большом круге, что и точки M и N. При этом точка L должна находиться между точками M и N, т. е. на меньшей дуге большого круга, соединяющей точки M и N.

Это следует из того, что неравенство (6) обращается в равенство только тогда, когда в равенство обращается (5). Последнее же может произойти, только когда трехгранный угол вырождается в плоский, т. е. когда точки M, L и N лежат в одной плоскости с центром сферы O и луч OL находится между лучами OM и ON. Но это и значит, что точка L принадлежит меньшей дуге большого круга, соединяющего точки M и N.

Отсюда видно, что меньшая дуга большого круга, соединяющая точки M и N, обладает свойствами, аналогичными свойствам отрезка в обычной (не сферической) геометрии. А именно:

1) через любые не совпадающие точки проходит ровно одна такая дуга (исключением является случай, когда точки M и N лежат на концах некоторого диаметра сферы — являются антиподами; в этом случае обе дуги любого большого круга, соединяющего M и N, равноправны, а таких кругов бесконечно много);

2) две такие дуги могут пересекаться только в одной точке;

3) для любой точки L, лежащей на такой дуге, соединяющей точки M и N, имеет место равенство $\varrho_S(M, L) + \varrho_S(L, N) = \varrho_S(M, N)$.

Заметим теперь следующее важное обстоятельство. Для обычного расстояния мы показали, что длина ломаной, соединяющей две точки M и N, больше, чем расстояние между точками M и N, т. е. чем длина отрезка MN. При этом рассуждение основывалось только на неравенстве треугольника и том факте, что оно обращается в равенство, лишь когда точки M, L и N лежат на одном отрезке. Так как неравенство треугольника верно и для расстояния на сфере, а обычным отрезкам здесь соответствуют меньшие дуги больших кругов, то ясно, что аналогичное утверждение верно и на сфере. Именно, если точки M и N соединены цепочкой из дуг больших кругов (см. рис. 7), последовательно соединенных друг с другом, то суммарная длина такой «сферической ломаной» больше (разумеется, если эта цепочка не лежит целиком на меньшей дуге большого круга, соединяющей точки M и N.), чем расстояние $\varrho_S(M, N)$.

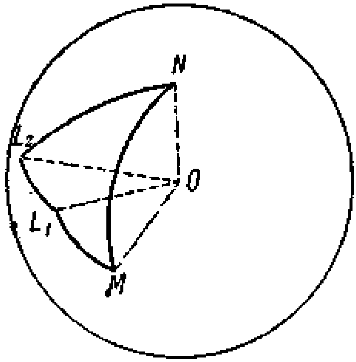


Рис. 7

Можно предложить читателю для упражнения восстановить полное доказательство этого утверждения по аналогии с приведенным выше доказательством для обычной ломаной. Это утверждение можно легко обобщить в следующей форме.

Длина меньшей дуги большого круга, соединяющей точки M и N, меньше, чем длина любой другой линии на сфере, соединяющей эти точки.

Итак, мы рассмотрели два вида «расстояний» и убедились, что их основные свойства совпадают. Дополнительные свойства типа (4) играют гораздо меньшую роль.

Поэтому следующий наш шаг будет состоять в том, что мы основные свойства расстояния (1, 2, 3, 4) возьмем в качестве исходных и будем изучать различные «пространства», в которых определено «расстояние», обладающее этими исходными свойствами. В этом параграфе мы рассмотрели два простейших примера таких пространств: обычное пространство и поверхность сферы.

2.3. Определение метрического пространства и расстояния

Мы начнем с напоминания того, что такое множество. Так же как «точка» в геометрии, понятие «множества» является первоначальным и определению не подлежит. Под словом «множество» в математике понимается совокупность каких-то объектов, называемых элементами множества.

Понятие «множества» состоит в сущности в том, что каким-то объектам приписывается свойство общности. Иначе, эти объекты по каким-то признакам попадают в некий класс, т. е. образуют множество. Задать множество — это и значит задать классификацию, т. е. для любого мыслимого объекта иметь возможность определить, попадает он в этот класс или нет. Принято говорить, что множество *содержит* каждый из своих элементов, а каждый из элементов данного множества в нем содержится. Множество считается заданным, если для любого объекта можно как-то проверить, содержится ли он в данном множестве.

Можно рассматривать, например, множество всех целых чисел. Солнце не содержится в этом множестве, так как оно не есть число, а объект совсем другой природы. Число π также не содержится в этом множестве, ибо оно не целое. С другой стороны, корни уравнения $x^2—3x+2 = 0$ входят в это множество. Можно рассматривать множество всех планет солнечной системы, понимая под планетами тела, движущиеся вокруг Солнца по замкнутым орбитам и имеющие массу не меньше чем 1 тонна. Солнце не содержится в этом множестве, так как является звездой, а не планетой. Земля содержится в этом множестве. Советская ракета, запущенная с Земли 2 января 1959 года, содержится в этом множестве, так как является искусственной планетой.

Пусть *E*—некоторое множество, а *N*—его элемент. Этот факт символически записывается как $N \in E$. Читается: «*N* является элементом *E*». Применяется также символическая запись типа $E\{L,\ M,\ N, \ldots\}$, где в скобках перечислены все элементы этого множества. Так, множество E_c, состоящее из всех столиц республик, можно было бы записать символически так: E_c{Москва, Киев, Минск, Тбилиси, Ереван, Баку, Рига, Таллин, Вильнюс, Ташкент, Алма-Ата, Фрунзе, Ашхабад, Дюшамбе, Кишинев}.

Если любой элемент множества *E* является в то же время элементом множества E_1, то множество *E* называется *подмножеством* множества E_1. Это записывается так: $E \subset E_1$ («*E* содержится в E_1»).

Так, множество целых чисел является подмножеством множества вещественных чисел.

Множество *E* называется *конечным,* если его элементы можно перенумеровать с помощью некоторого отрезка натурального ряда — множества $E_n\{1,\ 2;\ 3, \ldots, n\}$. Например, множество столиц республик E_c конечно, так как его можно перенумеровать элементами множества E_{ls}, что видно из следующей таблицы соответствий:

Москва	1	Ереван	5	Вильнюс	9	Ашхабад	13
Киев	2	Баку	6	Ташкент	10	Дюшамбе	14
Минск	3	Рига	7	Алма-Ата	11	Кишинев	15
Тбилиси	4	Таллин	8	Фрунзе	12		

Теперь мы можем дать определение метрического пространства. *Метрическим пространством* называется такое множество *E,* что для любой пары его элементов *M* и *N* определено вещественное число $\varrho(M, N)$, обладающее следующими свойствами:

1. $\varrho(M, N)=\varrho(N, M)$ (симметричность),
2. $\varrho(M, N)\geqslant 0$ (неотрицательность).
3. $\varrho(M, N)=0$ в том и только в том случае, когда *M* и *N* один и тот же элемент (невырожденность).
4. $\varrho(M, N)\leqslant\varrho(M, L)+\varrho(L, N)$ для любой тройки *M, L, N* элементов множества *E* (аксиома неравенства треугольника).

Элементы множества *E* принято называть *точками* соответствующего *пространства.* Метрическое пространство определяется, таким образом, выбором множества *E* и функции $\varrho(M, N)$ — *расстояния в пространстве.* Для простоты будем обозначать получаемое пространство той же буквой, что и соответствующее множество, хотя на самом деле пространство и множество его элементов суть разные объекты. Действительно, в одном и том же множестве *E* можно ввести различные расстояния и получить разные метрические пространства. Ниже мы построим различные определения расстояния на плоскости.

Вместо указанных четырех свойств расстояния можно было бы ввести только два (предполагая заранее, что $\varrho(M, N)$ — вещественное число):

1'. $\varrho(M, N)=0$ тогда и только тогда, когда точки *M* и *N* совпадают.

2'. $\varrho(M, N)\leqslant\varrho(M, L)+\varrho(N, L)$. Действительно, эти свойства вытекают из свойств 1, 2, 3 и 4, так как свойство 1' совпадает со свойством 3, а свойство 2' вытекает из неравенства треугольника 4 и условия 1.

С другой стороны, из свойств 1' и 2' можно вывести все условия 1, 2, 3 и 4.

Покажем это. Положим сначала в свойстве 2' *L=M,* тогда получим:

$$\varrho(M, N)\leqslant\varrho(M, M)+\varrho(N, M).$$

В силу 1' $\varrho(M, M)=0$. Отсюда следует $\varrho(M, N)\leqslant\varrho(N, M)$. Меняя в 2' местами точки *M* и *N* и проводя аналогичное рассуждение, убеждаемся, что $\varrho(N, M)\leqslant\varrho(M, N)$. Из последних двух неравенств следует свойство симметрии (1):

$$\varrho(M,\ N)=\varrho(N,\ M).$$

Подставляя в 2' *M* вместо *N* и *L*, получаем

$$\varrho(M,\ M)\leqslant\varrho(M,\ M)+\varrho(N,\ M)$$

или

$$0\leqslant\varrho(N,\ M),$$

что доказывает условие неотрицательности расстояния (2). Наконец, используя доказанное условие симметрии, можно в 2' поменять во втором слагаемом правой части *N* и *L* и получить неравенство треугольника 4. Таким образом, система свойств 1' и 2' равносильна системе свойств 1, 2, 3 и 4. Удобнее пользоваться более полной системой, которая в явном виде содержит больше фактических свойств расстояния. Однако интересно знать, что все эти свойства могут быть выведены всего из двух.

С точки зрения введенного определения содержание предыдущего параграфа можно было бы охарактеризовать как доказательство, что множество точек трехмерного пространства с расстоянием, определенным как длина отрезка или по формуле (1), является метрическим пространством. Аналогичный факт доказан в конце того же параграфа про множество точек сферы *S* с расстоянием $\varrho_S(M,\ N)$.

Еще один пример метрического пространства получится, если взять множество точек некоторой поверхности П в трехмерном пространстве, а в качестве расстояния $\varrho_\Pi(M,\ N)$ взять минимум длин путей, проходящих по поверхности П и соединяющих точки *M* и *N* (Для простоты мы предполагаем, что такой наикратчайший путь существует для любых точек *M* и *N* на поверхности П. При некоторых предположениях о свойствах поверхности П это можно доказать.). Первые три свойства расстояния очевидны.

Аксиома треугольника проверяется так. Соединим точки М и *L* наикратчайшим путем, и точки *L* и *N* также соединим наикратчайшим путем. Тогда путь, состоящий из двух наикратчайших линий *ML* и *LN*, соединяет точки *M* и *N*. Ясно, что длина этого пути не может быть меньше длины наикратчайшего пути из *M* в *N*. Так как длина этого пути есть $\varrho_\Pi(M,\ L)+\varrho_\Pi(L,\ N)$, а длина наикратчайшего пути из *M* в *N* есть $\varrho_\Pi(M,\ N)$, то отсюда следует искомое соотношение:

$$\varrho_\Pi(M,\ N)\leqslant\varrho_\Pi(M,\ L)+\varrho_\Pi(L,\ N). \qquad (7)$$

Заметим, что на поверхности сферы наикратчайший путь между точками есть как раз наименьшая дуга большого круга, соединяющая эти точки. Доказательство этого факта было намечено в конце предыдущего параграфа. Это доказательство как раз и основано на том, что независимыми рассуждениями было получено неравенство треугольника и показано, что для наименьших дуг большого круга оно обращается в равенство.

Введем понятие отрезка в произвольном метрическом пространстве. Именно, мы назовем *отрезком, соединяющим точки M и N в метрическом пространстве E,* такое множество точек $E_{M,N}$ этого пространства, что для любой точки L множества $E_{M,N}$ выполнено равенство

$$\varrho(M, N) = \varrho(M, L) + \varrho(L, N). \tag{8}$$

Нетрудно видеть, что для обычного расстояния на плоскости или в трехмерном пространстве множество $E_{M,N}$ совпадает с отрезком MN в обычном смысле слова. На сфере S при расстоянии $\varrho_S(N, M)$, введенном ранее, отрезком $E_{M,N}$ является наименьшая дуга большого круга, соединяющая точки M и N (если они не лежат на одном диаметре) и весь большой круг, если точки M и N — антиподы на сфере.

Предоставляем читателю убедиться, что для расстояния

$$\varrho_{\Pi}(M, N),$$

введенного выше, отрезок $E_{M,N}$ есть кратчайший путь (или так называемая геодезическая линия), соединяющий точки M и N.

В любом метрическом пространстве E можно также ввести понятие сферы $S_{M,r}$ с центром в точке M и радиусом r, обозначая этимтермином множество таких точек N, для которых расстояние $\varrho(M, N) = r$.

На плоскости это понятие соответствует окружности; в трехмерном пространстве — обычной сфере; для расстояния $\varrho_S(M, N)$ — окружностям на сфере S. В качестве еще одного (тривиального) примера метрического пространства возьмем произвольное множество E и определим расстояние между двумя любыми не совпадающими точками M и N равным единице, а между совпадающими точками — нулю. Легко видеть, что все необходимые условия при таком определении выполнены.

Разнообразные примеры метрических пространств будут рассмотрены ниже.

В метрическом пространстве можно всегда определить операцию предельного перехода.

Последовательность точек метрического пространства E: $L_1, L_2, \ldots, L_k$ называется *сходящейся* к точке $L \in E$, если для всякого положительного вещественного числа ε можно подобрать такой номер $n(\varepsilon)$, что из условия $k > n(\varepsilon)$ следует

$$\varrho(L, L_k) < \varepsilon.$$

Иными словами, последовательность сходится к точке *L*, если, начиная с некоторого номера, расстояние между членами последовательности и точкой *L* (пределом) становится меньше любого наперед заданного положительного числа. По аналогии с обычной записью можно ввести обозначение

$$L = \lim_{k \to \infty} L_k.$$

Нетрудно убедиться, что для метрического пространства, состоящего из вещественных чисел *x*, с расстоянием, определяемым по правилу

$$\varrho(x, y) = |x - y|,$$

это определение предела совпадает с обычным.

Для метрического пространства E_3, состоящего из точек трехмерного пространства с обычным расстоянием, данное понятие предела позволяет хорошо определить, что такое предел последовательности точек в трехмерном пространстве.

Заметим, что в этом случае совокупность точек *M*, для ко торых $\varrho(L, M) < \varepsilon$, образует сферу с центром в точке *L* и радиусом ε. Последовательность геометрических точек $L_1, L_2, \ldots, L_k, \ldots$ оказывается *сходящейся* к точке *L*, если, начиная с некоторого номера *n*(ε), все члены этой последовательности с *k*>*n*(ε) лежат в сфере с центром в *L* и радиусом ε, причем такой номер *n*(ε) должен существовать для всякого положительного ε.

Теорема. *Если последовательность элементов метрического пространства* E: $L_1, L_2, \ldots, L_k, \ldots$ *сходится к пределу L, то для всякого положительного числа* ε *существует такой номер m* (ε), *что при* $k > m(\varepsilon)$ *и* $k' > m(\varepsilon)$ *выполняется условие* $\varrho(L_k, L_{k'}) < \varepsilon$.

Доказательство. На основании определения предела можно выбрать такой номер $n\left(\frac{\varepsilon}{2}\right)$, что при $k > n\left(\frac{\varepsilon}{2}\right)$ и $k' > n\left(\frac{\varepsilon}{2}\right)$ будут справедливы неравенства

$$\varrho(L_k, L) < \frac{\varepsilon}{2}; \quad \varrho(L_{k'}, L) < \frac{\varepsilon}{2}.$$

Но по аксиоме треугольника и свойству симметричности

$$\varrho(L_k, L_{k'}) \leqslant \varrho(L_k, L) + \varrho(L, L_{k'}) =$$

$$= \varrho(L_k, L) + \varrho(L_{k'}, L) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Иными словами, если положить $m(\varepsilon) = n\left(\frac{\varepsilon}{2}\right)$, то при *k*>*m* (ε) и $k' > m(\varepsilon)$ будет иметь место неравенство

$$\varrho(L_k,\ L_{k'})<\varepsilon,$$

что и доказывает теорему.

Эта теорема позволяет охарактеризовать внутренние свойства сходящихся последовательностей.

Если в пространстве *E* справедливо утверждение, обратное этой теореме, то пространство *E* называется *полным* метрическим пространством.

Определение полного метрического пространства удобно дать в следующей развернутой форме. Последовательность точек метрического пространства $E: L_1, L_2, \ldots, L_k, \ldots$ называется *фундаментальной,* если для всякого числа ε >0 найдется номер *m*(ε), такой, что при *k*>*m* (ε) и *k*' (ε) выполнено условие

$$\varrho(L_k,\ L_{k'})<\varepsilon.$$

Метрическое пространство *E* называется *полным,* если в нем всякая фундаментальная последовательность имеет предел.

Прямая, плоскость и трехмерное пространство с обычным расстоянием являются полными метрическими пространствами.

Вопрос о полноте метрических пространств очень существен при использовании этих пространств в математическом анализе (для математического анализа принципиальное значение имеет полнота ряда метрических пространств, точками которых являются функции.)

Два метрических пространства называются *изометрическими,* если между множествами их элементов можно установить взаимно однозначное соответствие, так что расстояния между парами сопоставленных друг другу элементов равны. С точки зрения теории метрических пространств изометрические пространства являются одинаковыми.

Рассмотрим пример. Пусть пространство *B* есть плоскость с обычным расстоянием, а пространство E_K есть множество комплексных чисел *z* с расстоянием, определяемым по формуле

$$\varrho_K(z,\ z_1)=|z-z_1|.$$

Обычный способ изображения комплексных чисел точками плоскости осуществляет взаимно однозначное соответствие между этими пространствами. Нетрудно проверить, что это соответствие является изометричным, так как величина

$$|z-z_1|=\sqrt{(x-x_1)^2+(y-y_1)^2}$$

равна геометрическому расстоянию между соответствующими точками.

Приведенное нами определение метрического пространства с расстоянием не является наиболее общим, встречающимся в математике. Существуют различные обобщения этого понятия. Так, можно было бы, сохраняя все свойства расстояния, разрешить ему принимать для некоторых точек бесконечные значения. Это обобщение не очень содержательно. В ряде математических задач приходится иметь дело с расстоянием, у которого отсутствует свойство симметрии. Свойства такого расстояния мы изучим ниже. Наконец, в теории относительности приходится рассматривать расстояние, которое может принимать и мнимые значения.

2.4. Примеры метрических пространств

Рассмотрим ряд примеров метрических пространств с необычно определяемым расстоянием.

Ряд интересных примеров получается, если пространство есть множество точек обычной плоскости, а расстояние вводится различными способами. Точки плоскости мы будем при этом изображать с помощью раз и навсегда выбранной координатной системы, так что каждая точка плоскости задается двумя координатами *x* и *y*. Поэтому точку плоскости удобно обозначать как *M (x, y).*

Метрическое пространство *l* возникает, если расстояние между точками *M(x, y)* и *N(x,y)* определить по формуле

$$\varrho_1(M, N) = |x - x_1| + |y - y_1|. \qquad (9)$$

Из рис. 3 можно усмотреть, что $\varrho_1(M, N)$ равно сумме длин катетов треугольника *MLN,* в котором *MN* — гипотенуза, а катеты *ML* и *LN* параллельны осям координат. Так как длина гипотенузы не превосходит суммы длин катетов, то всегда

$$\varrho(M, N) \leqslant \varrho_1(M, N), \qquad (10)$$

где $\varrho(M, N)$ есть обычное геометрическое расстояние. Неравенство (10) превращается в равенство в том и только в том случае, когда отрезок *MN* или вертикален, или горизонтален, т. е. параллелен одной из осей координат.

Если в неравенство (10) подставить алгебраические выражения для соответствующих расстояний (9) и (2), то мы получим неравенство

$$\sqrt{(x-x_1)^2+(y-y_1)^2} \leqslant |x-x_1| + |y-y_1|.$$

При x_1—y_1=0 отсюда следует простое неравенство

$$\sqrt{x^2+y^2} \leqslant |x| + |y|. \qquad (11)$$

Свойства 1, 2 и 3 для расстояния $\varrho_1(M, N)$ очевидные.

Докажем, что выполнено и свойство 4. Для этого рассмотрим три точки $M(x, y)$, $N(x_1, y_1)$, $L(x_2, y_2)$ и запишем элементарное тождество

$$|x-x_1|+|y-y_1| = \\ = |x-x_2+x_2-x_1|+|y-y_2+y_2-y_1|. \quad (12)$$

Если использовать тот факт, что для любых вещественных чисел *a* и *b*, $|a+b| \leqslant |a|+|b|$, то из (12) получается неравенство

$$|x-x_1|+|y-y_1| \leqslant \\ \leqslant |x-x_2|+|x_2-x_1|+|y-y_2|+|y_2-y_1|,$$

что уже после перестановки членов в правой части равносильно искомому соотношению 1

$$\varrho_1(M, N) \leqslant \varrho_1(M, L)+\varrho_1(L, N). \quad (13)$$

Тем самым доказано неравенство треугольника для пространства *l*.

Расстояние $\varrho_1(M, N)$ можно трактовать как минимальный путь, проходимый частицей, движущейся из *M* в *N*, если эта частица может двигаться только по отрезкам, параллельным осям координат. Из рис. 8 видно, что существует много (бесконечно много!) наикратчайших путей для этой частицы.

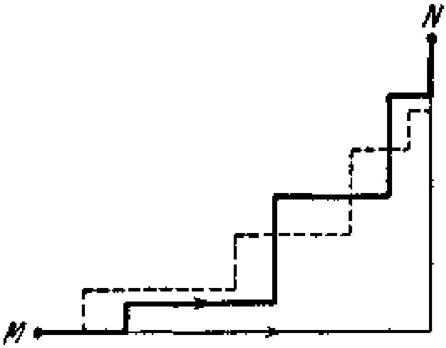


Рис. 8.

Нетрудно показать, что это равносильно тому, что в пространстве *l* существует, вообще говоря, бесконечное множество различных «отрезков», соединяющих точки *M* и *N* (исключением является случай, когда точки *M* и *N* находятся на одной вертикали или горизонтали).

«Отрезком» в пространстве *l*, соединяющим точки *M* и *N*, является любая ломаная, соединяющая эти точки, состоящая из вертикальных и горизонтальных линий и пересекаемая любой вертикальной или

горизонтальной линией не более чем по одному звену. (Это утверждение предоставляем проверить читателю.)

Еще более естественная картина получится, если рассмотреть метрическое пространство *l*, состоящее из всех точек, находящихся в узлах некоторой прямоугольной решетки на плоскости (рис. 9) с расстоянием, определенным формулой (9).

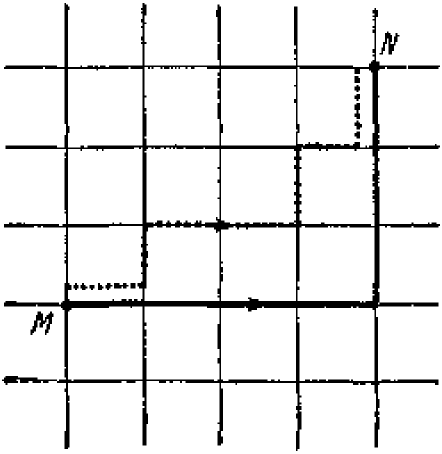


Рис. 9.

Точки этого пространства можно рассматривать как перекрестки улиц бесконечного идеально распланированного города (тоже своего рода математическая абстракция!). Расстояние $\varrho_1(M,\ N)$ есть в этом случае минимальная длина пути, который необходимо пройти по улицам города, чтобы попасть из перекрестка *M* в перекресток *N* (Не пользуясь, разумеется, проходными дворами).

Следующий пример метрического пространства — пространство *C* состоит из точек плоскости с расстоянием, определения. определяемым по формуле

$$\varrho_\infty(M,\ N) = \max(|x-x_1|,\ |y-y_1|), \qquad (14)$$

если точка *M* имеет координаты *x* и *y,* а точка *N*—координаты x_1 и y_1. Геометрически это означает (см. рис. 3), что расстояние $\varrho_\infty(M,\ N)$ равно максимальному катету треугольника *MLN*. Так как даже максимальный катет меньше гипотенузы (или равен ей в вырожденном треугольнике), то

$$\varrho_\infty(M,\ N) \leqslant \varrho(M,\ N), \qquad (15)$$

где $\varrho(M,\ N)$—обычное геометрическое расстояние.

Отсюда при $x_l=y_l=0$ получается алгебраическое неравенство

$$\max(|x|, |y|) \leqslant \sqrt{x^2+y^2}. \qquad (16)$$

Для расстояния $\varrho_\infty(M, N)$ также очевидны свойства 1, 2, 3.

Аксиома треугольника проверяется следующим образом. Вы берем любые три точки $M(x, y)$, $N(x_1, y_1)$ и $L(x_2, y_2)$. Предположим, что $|x-x_1| \geqslant |y-y_1|$ (В противном случае доказательство видоизменяется очевидным образом.). Это значит, что

$$\varrho_\infty(M, N) = \max(|x-x_1|, |y-y_1|) = |x-x_1| = \\ = |x-x_2+x_2-x_1|.$$

Следовательно,

$$\varrho_\infty(M, N) \leqslant |x-x_2| + |x_2-x_1|. \qquad (17)$$

Очевидно также, что

$$\left. \begin{aligned} |x-x_2| &\leqslant \max(|x-x_2|, |y-y_2|) = \varrho_c(M, L), \\ |x_2-x_1| &\leqslant \max(|x_2-x_1|, |y_2-y_1|) = \varrho_c(L, N). \end{aligned} \right\} \qquad (18)$$

Подставляя правые части неравенств (18) в неравенство (17), получаем

$$\varrho_\infty(M, N) \leqslant \varrho_\infty(M, L) + \varrho_\infty(L, N), \qquad (19)$$

что и требовалось доказать.

На примере города, изображенного на рис. 9, расстоянию $\varrho_\infty(M, N)$ можно придать следующий смысл. Предположим, что некто добирается до перекрестка *N*, отсчитывая попорядку номера продольных и поперечных улиц. Предположим, что длины всех кварталов одинаковы (и равны единице), и путешественник может сбиться со счета не более чем на *k* улиц (как продольных, так и поперечных).

Тогда путешественник может попасть на перекресток *M,* отстоящий от *N* на расстоянии $\varrho_\infty(M, N) \leqslant k$.

Мы уже указывали, что в любом метрическом пространстве можно ввести понятие обобщенной сферы. (Слово «обобщенный» мы в дальнейшем будем опускать.) Аналогично этому обобщенным шаром с центром в точке *M* и радиусом *r* мы будем называть множество точек *N,* для которых

$$\varrho(M, N) \leqslant r. \qquad (20)$$

Если $\varrho(M, N)$ — геометрическое расстояние на плоскости, то обобщенный шар является кругом с центром в точке *M* и радиусом *r*.

Для трехмерного пространства с обычным расстоянием шар, определяемый неравенством (20), есть обычный шар с центром в точке *M* и радиусом *r.* (Отсюда и происходит название «обобщенный шар».)

В пространстве *l* обобщенным шаром является квадрат с центром в точке *M*, с диагоналями, равными *2r* и параллельными осям координат. В пространстве *C* шаром является также квадрат с центром в точке *M,* но со сторонами, равными *2r* и параллельными осям координат. На рис. 10 изображены обобщенные шары радиуса *r* в пространствах *l* и *C* и в смысле обычного расстояния.

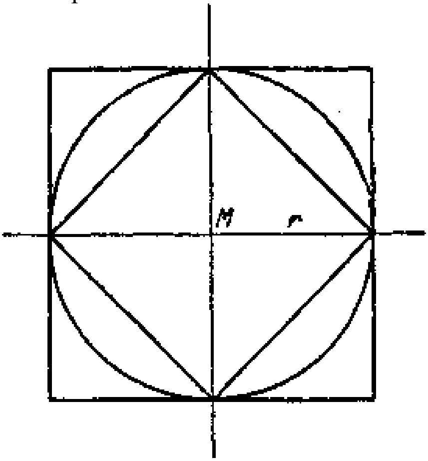


Рис. 10

Доказательство того, что шары в *l* и *C* имеют указанный вид, предоставляется читателю.

Интересный класс метрических пространств получается, если определить расстояние $\varrho_p(M, N)$ по формуле

$$\varrho_p(M, N) = \sqrt[p]{|x - x_1|^p + |y - y_1|^p}, \tag{21}$$

и получить тем самым пространство l_p.

Свойства 1, 2, 3 для этого расстояния очевидны. Аксиома треугольника следует из неравенства Минковского

$$\sqrt[p]{|a + a_1|^p + |b + b_1|^p} \leqslant \leqslant \sqrt[p]{|a|^p + |b|^p} + \sqrt[p]{|a_1|^p + |b_1|^p}, \tag{22}$$

справедливого при $p \geq 1$. Читатель легко получит из (22) доказательство аксиомы треугольника для расстояния (21) при $p \geq 1$, если для точек $M(x, y)$, $N(x_1, y_1)$ и $L(x_2, y_2)$ обозначит

$$a = x - x_2; \quad a_1 = x_2 - x_1; \quad b = y - y_2; \quad b_1 = y_2 - y_1$$

и будет действовать по аналогии с доказательством аксиомы треугольника для пространства *l.* При $p<1$ аксиома треугольника неверна, так как знак неравенства (22) меняется на противоположный.

Нетрудно видеть, что при $p = 1$ расстояние $\varrho_p(M, N) = \varrho_l(M, N)$, а при p=2 расстояние $\varrho_p(M, N)$ совпадает с обычным геометрическим расстоянием $\varrho(M, N)$.

Таким образом, рассмотренное ранее пространство l совпадает с пространством l_1, а плоскость с обычным расстоянием есть пространство l_2.

Покажем, что при $p \to \infty$ расстояние $\varrho_p(M, N)$ стремится к расстоянию $\varrho_\infty(M, N)$.

В самом деле, рассмотрим сначала случай $|x - x_1| > |y - y_1|$.

Тогда $\varrho_\infty(M, N) = |x - x_1|$. С другой стороны, преобразуя (21), имеем

$$\varrho_p(M, N) = |x - x_1| \sqrt[p]{1 + \left|\frac{y - y_1}{x - x_1}\right|^p}. \tag{23}$$

Замечая, что при p>1

$$1 \leqslant \sqrt[p]{1 + \left|\frac{y - y_1}{x - x_1}\right|^p} \leqslant 1 + \left|\frac{y - y_1}{x - x_1}\right|^p$$

и что величина $\left|\frac{y - y_1}{x - x_1}\right|^p \to 0$ при $p \to \infty$, получаем

$$\lim_{p \to \infty} \sqrt[p]{1 + \left|\frac{y - y_1}{x - x_1}\right|^p} = 1.$$

Учитывая этот факт и используя (23), видим, что

$$\lim_{p \to \infty} \varrho_p(M, N) = |x - x_1| = \varrho_\infty(M, N). \tag{24}$$

Аналогично, при $|x - x_1| < |y - y_1|$ можно получить

$$\lim_{p \to \infty} \varrho_p(M, N) = |y - y_1| = \varrho_\infty(M, N). \tag{25}$$

Наконец, рассмотрим случай $|x - x_1| = |y - y_1|$. Тогда

$$\varrho_p(M, N) = \sqrt[p]{2|x - x_1|^p} = |x - x_1| \sqrt[p]{2}.$$

Так как $\lim\limits_{p \to \infty} \sqrt[p]{2} = 1$, то и в этом случае

$$\lim_{p \to \infty} \varrho_p(M, N) = |x - x_1| = \varrho_\infty(M, N). \tag{26}$$

Сопоставляя (24), (25) и (26), получаем, что во всех случаях

$$\lim_{p \to \infty} \varrho_p(M, N) = \varrho_\infty(M, N), \tag{27}$$

что и требовалось доказать.

Таким образом, пространство C можно было бы обозначить символом l_∞, так как расстояние в этом пространстве получается предельным переходом из расстояния в пространстве l_p при $p \to \infty$.

На рис. 11 показаны шары в пространствах l_p с центром в точке M при различных значениях p.

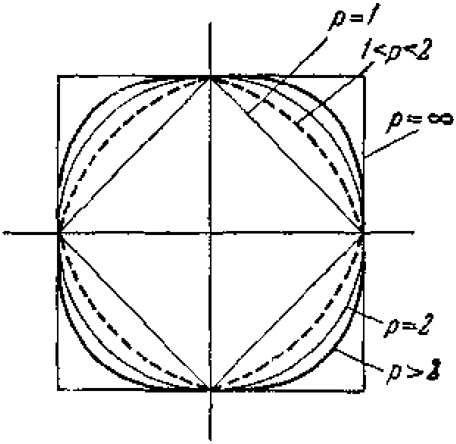


Рис. 11

Метрические пространства l носят общее название пространств Минковского. Ниже мы рассмотрим многомерные пространства Минковского.

В качестве упражнения для читателя предлагается найти вид «отрезков» в пространствах l_p.

Следующий класс метрических пространств на плоскости получится, если расстоянием называть минимальное время, требующееся для прохождения при некоторых заданных условиях пути из M в N.

При этом обычное расстояние получается, если путь от M к N проходится точкой, движущейся с постоянной скоростью, равной единице.

Расстояние в l получится, если этой точке разрешается двигаться и с постоянной скоростью только по отрезкам, параллельным осям координат.

Новый пример получится, если рассмотреть (см. рис. 12) схему московского метрополитена и считать, что путешественник проходит из точки M в N следующим образом.

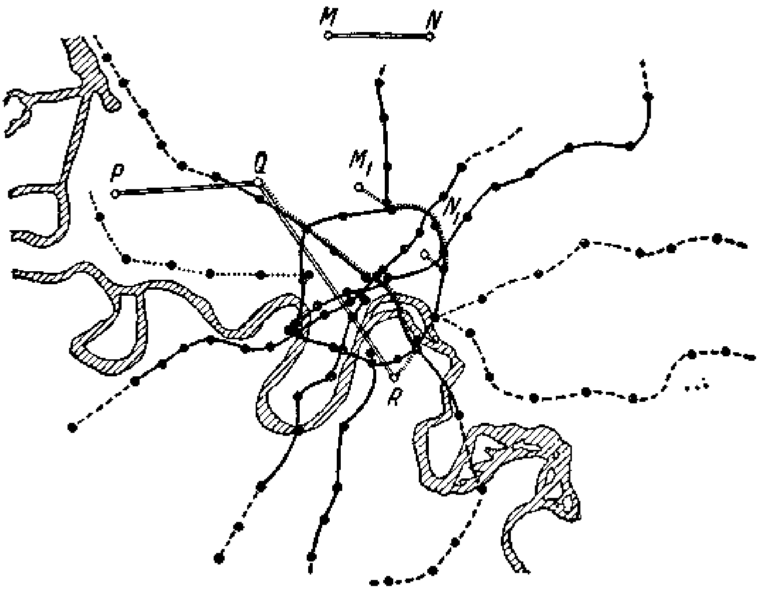


Рис. 12.

Если существует одна и та же ближайшая от обеих точек станция метро, то путешественник идет пешком но наикратчайшему пути. Если такой станции нет, то путешественник идет пешком (кратчайшим путем) до станции, ближайшей к исходной точке *M*, едет скорейшим путем к станции, ближайшей к точке *N*, и оттуда идет пешком к *N*. Если для точек *M* или *N* есть несколько ближайших станций метро, то выбирается вариант маршрута с наименьшим временем пути. На рис. 12 есть точка *M,* из которой надо двигаться в точку *N* пешим способом, и точка M_1, из которой надо двигаться в точку N_1 с помощью метро. Скажем, если человеку, живущему посредине между станциями «Рижская» и «Ботаническая», нужно добраться в район Земляного вала, то ему надо сесть на «Ботанической» и доехать до «Лермонтовской» или «Курской». Нетрудно видеть, что определенное таким образом расстояние $\varrho_t(M, N)$ существенно отличается от обычного геометрического расстояния. В самом деле, если точка *Q* расположена около вертолетной станции, т. е. между станциями «Динамо» и «Аэропорт», точка *P* в районе Волоколамского шоссе, а точка *R* в районе Валовой улицы (возле станции метро Павелецкая) (см. рис. 12), то в смысле обычного расстояния точка *P* гораздо ближе к Q, чем точка *R*,

$$\varrho(P, Q) < \varrho(Q, R).$$

Однако из карты на рис. 12 очевидно, что

$$\varrho_t(P, Q) > \varrho_t(P, R).$$

Действительно, если не пользоваться услугами такси, то от вертолетной станции до Валовой можно добраться быстрее, чем до Волоколамского шоссе.

Для расстояния $\varrho_t(M, N)$ нетривиальной является аксиома симметричности (свойство 1).

Равенство

$$\varrho_t(M, N) = \varrho_t(N, M)$$

означает, что время, затрачиваемое в обе стороны кратчайшего пути *MN*, одинаково. Это более или менее справедливо, если пользоваться метро и пешим способом сообщения. Но уже при пользовании такси это не так, ибо одно дело ждать такси недалеко от таксомоторного парка, а совсем другое — получить такси в каком-либо отдаленном районе, или у Курского вокзала в часы прихода поездов.

Свойства 2 и 3 для расстояния $\varrho_t(M, N)$ очевидны. Доказательство свойства 4 (аксиомы треугольника) читатель легко проведет самостоятельно, если вспомнит, как это свойство доказывалось для расстояния $\varrho_\Pi(M, N)$.

Для исследования свойств метрических пространств введем следующее понятие.

Пусть *E*—некоторое метрическое пространство, а $L_1, L_2, \ldots, L_k$— точки этого пространства.

Назовем *областью Дирихле точки* L_i множество D_i, состоящее из точек *N*, для которых

$$\varrho(L_i, N) \leqslant \varrho(L_j, N) \qquad (28)$$

для всех $j \neq i$. Таким образом, область Дирихле D_i состоит из точек пространства, расположенных ближе к точке L_i, чем к остальным выбранным точкам. Ясно, что определение области Дирихле зависит от набора точек $L_1, L_2, \ldots, L_k$ и выбранной точки L_i. Рассмотрим примеры областей Дирихле в различных метрических пространствах.

Для начала возьмем плоскость с обычным расстоянием $\varrho_2(M, N)$ и две точки L_1 и L_2. Соединим эти точки отрезком L_1, L_2 (рис. 13) и проведем через середину этого отрезка перпендикуляр.

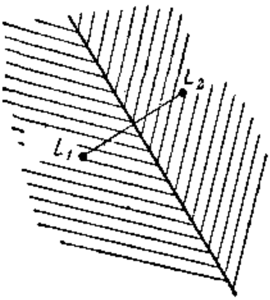


Рис. 13.

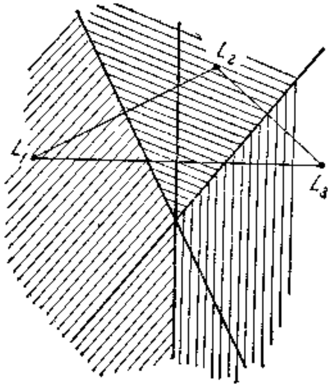


Рис. 14.

Этот перпендикуляр разбивает плоскость на две полуплоскости, которые и будут областями Дирихле для точек L_1 и L_2.

Возьмем теперь три точки L_1, L_2 и L_3 на плоскости с обычным расстоянием. На рис. 14 построены и отмечены штриховкой области Дирихле для этих трех точек. Способ построения ясен из рисунка.

Рассмотрим теперь две точки L_1 и L_2 на плоскости с расстоянием $\varrho_1(M, N)$ (т. е. в пространстве *l*). Для наглядности мы опять рассматриваем город, разделенный на кварталы. Области Дирихле состоят из тех перекрестков, от которых путь по городу будет короче, чем до точки L_l (соответственно L_2). Эти области отделены на рис. 15 жирной чертой. На рис. 16 приведено аналогичное разбиение для пространства *C*.

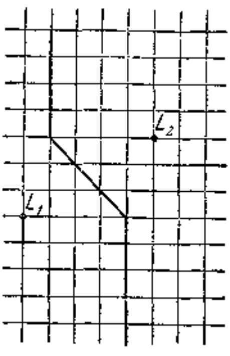


Рис. 15.

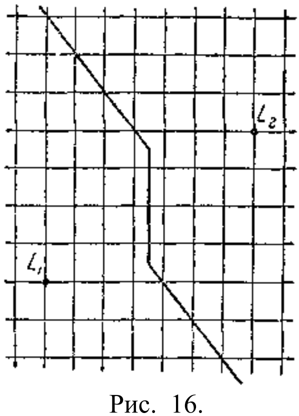


Рис. 16.

Предоставляем читателю вывести общее правило построения областей Дирихле в пространствах *l* и *C,* для пар точек L_1 и L_2, а затем для троек L_1, L_2, L_3.

Обращаясь снова к рис. 12, мы видим, что если построить разбиение пространства на области Дирихле для пары точек *P* и *R,* то точка *Q* попадает в области Дирихле точки *R*.

Предоставляем опять же читателю нарисовать это разбиение на области Дирихле. Важно, что это разбиение сильно отличается от разбиения на области Дирихле, показанного на рис. 13, 15 и 16.

2.5. Пространство сообщений

Под словом «сообщение» обычно понимается какая-то конкретная форма передачи информации. Сообщениями в широком смысле слова являются книги, письма, телеграммы, записи музыкальных произведений на различные носители или с помощью нот, записи с помощью которых информация о решаемой задаче вводится в вычислительные машины, сигналы, управляющие взлетом и посадкой космических кораблей, молекулы дезоксирибонуклеиновой кислоты (ДНК), передающие наследственную информацию от родителей к потомству, и т. п. Вопросы передачи и кодирования информации рассматривают в теории информации. В теории информации разработаны способы определения количества информации, содержащейся в данном сообщении. Оказывается, что определенное количество информации может быть закодировано в виде сообщения разной величины. С этим обстоятельством мы часто встречаемся в жизни. Так, при составлении телеграммы мы стараемся сократить количество слов, не нарушая смысла (т. е. сохраняя количество информации).

Обратная ситуация возникает, когда плохо подготовленный студент на экзамене или семинаре начинает тянуть время, пытаясь небольшое количество имеющейся у него по данному вопросу информации выразить в виде достаточно внушительного количества слов.

Однако избыточная величина сообщения по сравнению с количеством передаваемой информации не всегда вредна. Эту избыточность можно обратить на пользу, когда при передаче сообщения возникают помехи.

Например, при плохой слышимости по телефону мы вынуждены повторять отдельные слова. При передаче незнакомого и трудного для восприятия названия или фамилии используется побуквенная система. Так, сообщая по телефону собственное имя «Павсикакий», мы будем произносить следующее: «Петр, Андрей, Василий, Семен, Илья, Константин, Анна, Кирилл, Инна, Иван краткий».

В этом и следующем параграфах мы, не вдаваясь в специфические вопросы теории информации, будем изучать некоторые приемы «помехоустойчивого кодирования» информации. Иными словами, мы будем изучать способы записи сообщений, позволяющие автоматически исправлять допущенные ошибки, если этих ошибок не слишком много. **Эти приемы тесно связаны с возможностью определения расстояния в так называемом пространстве сообщений, которое изучается в настоящем параграфе.**

Идею этих приемов все мы часто используем в обиходе, читая книги с опечатками и получая телеграммы с ошибками. Действительно, если в книге мы читаем слово «кистрюля», то не нужно заглядывать в перечень опечаток, чтобы догадаться о его значении. Очень мало шансов, что автор в этом месте написал слово «телеграф». В самом деле, в этом случае мы имели бы дело с восемью опечатками подряд, в то время как из слова «кастрюля» тот же результат получается вследствие одной опечатки.

Правда, существуют курьезные примеры, когда из-за ошибки в одной букве могут возникнуть различные слова. Скажем, слово «корона» может произойти и из «корова», и из «ворона».

На этом обстоятельстве основан известный анекдот, как одна провинциальная газета поместила в сообщении о короновании Николая Второго фразу: «митрополит возложил на голову Его Величества ворону». На следующий день было помещено исправление: «митрополит возложил на голову Его Величества корову».

Легко видеть, что и в этом случае истинный смысл сообщения довольно легко установить по контексту.

Аналогично ошибка в потной записи часто может быть обнаружена по фальшивому звучанию и исправлена по законам гармонии.

Надо сказать, что точно так же ошибки могут возникать не только при передаче информации, но и при длительном ее хранении, например в памяти электронных вычислительных машин.

Проблема восстановления правильного сообщения имеет одинаковый вид и при передаче и при хранении информации.

Каждый вид сообщения записывается с помощью каких-то символов. Множество используемых символов образует алфавит $\mathfrak{A}$. Мы предположим, что этот алфавит задан заранее и состоит из конечного множества символов. Например, алфавит может состоять из всех русских букв, символа пропуска буквы и знаков препинания. В этом алфавите можно писать любые русские фразы. Другой пример алфавита — множество десятичных цифр, алгебраических знаков, знаков препинания и латинских и греческих букв. С помощью такого алфавита можно записывать разнообразнейшие математические формулы.

Еще один пример — это двоичный алфавит — множество из двух символов $\mathfrak{A}_2\{0,\ 1\}$. С помощью такого алфавита можно записывать числа в двоичной системе счисления.

Действительно, можно легко показать, что всякое целое число *x* записывается в виде

$$x=\varepsilon_n 2^n+\varepsilon_{n-1}2^{n-1}+\ldots+\varepsilon_1 2+\varepsilon_0, \qquad (29)$$

где величины ε_i- принимают значения 0 или 1.

Таким образом, для того чтобы передать информацию о значении числа *x*, достаточно передать последовательность таков алфавита $\mathfrak{A}_2$: $\varepsilon_n\ \varepsilon_{n-1}\ \ldots\ \varepsilon_1\ \varepsilon_0$. На самом деле, чтобы разделять сообщения о разных числах, надо либо ввести специальный знак для конца числа, либо передавать только последовательности двоичных знаков стандартной длины.

Последний способ фактически применяется в вычислительных машинах, где запоминаемые двоичные последовательности обычно имеют числовые массивы переменной длины. Формула (29) аналогична хорошо известной формуле:

$$x=\alpha_n 10^n+\alpha_{n-1}10^{n-1}+\ \ldots+\alpha_1 10+\alpha_0, \qquad (30)$$

где $\alpha_n,\ \alpha_{n-1},\ \ldots,\ \alpha_1,\ \alpha_0$—цифры в десятичном представлении числа *x*. Легко распространить равенство (29) на не целые числа, подобно тому как вводятся десятичные дроби.

Установим связь между числом *n* и значением *x* в (29). Ясно, что если несколько первых слагаемых в правой части (29) равны нулю, то их можно отбросить. Поэтому мы будем предполагать, что $\varepsilon_n=1$. Перенося все члены, кроме первого, из правой части в левую, мы получим

$$x-\varepsilon_{n-1}2^{n-1}-\varepsilon_{n-2}2^{n-2}\ldots-\varepsilon_1 2^1-\varepsilon_0=\varepsilon_n 2^n.$$

Отсюда ясно, что

$$2^n\leqslant x$$

или

$$n\leqslant\log_2 x. \qquad (31)$$

С другой стороны, верна следующая оценка:

$$\varepsilon_n 2^n+\varepsilon_{n-1}2^{n-1}+\varepsilon_{n-2}2^{n-2}+\ldots+\varepsilon_1 2^1+\varepsilon_0\leqslant$$
$$\leqslant 2^n+2^{n-1}+2^{n-2}+\ldots+2^1+1=2^{n+1}-1.$$

Из последнего соотношения и из (29) следует

$$x\leqslant 2^{n+1}-1$$

или

$$x<2^{n+1},$$

что можно записать и так:

$$n+1>\log_2 x. \qquad (32)$$

Сопоставляя (31) и (32), получаем двойное неравенство

$$n\leqslant \log_2 x\leqslant n+1. \qquad (33)$$

Неравенство (33) означает в других терминах, что

$$n=[\log_2 x],$$

т. е. *n* равно целой части $\log_2 x$ (читается: «*n* равно антье от $\log_2 x$»). Вышеизложенное можно записать еще так: количество двоичных знаков, необходимых для кодирования всех целых чисел в диапазоне $0\leqslant x\leqslant a$, равно

$$1+[\log_2 a]=1+n. \qquad (34)$$

Единица здесь добавляется потому, что двоичные знаки включают еще и нулевой: ε_0.

С помощью двоичного алфавита $\mathfrak{A}_2$ в памяти вычислительных машин записывается любая информация (числа, команды, логические соотношения и др.).

Сообщением в данном алфавите $\mathfrak{A}$ мы будем называть конечную упорядоченную последовательность символов этого алфавита. Иногда удобно делить сообщения на некоторые стандартные сообщения, которые называются словами.

Вообще говоря, возможны также бесконечные алфавиты и сообщения, но мы их здесь рассматривать не будем.

Сообщение, записанное в одном алфавите, можно различными способами записать в другом алфавите.

Так, мы уже видели, что целое число, заданное десятичными цифрами, можно записать и в двоичном алфавите. Один из важных способов такой перекодировки состоит в следующем. Пусть дан алфавит $\mathfrak{A}$. Возьмем все слова, составленные в этом алфавите, длиной не более *k* и будем считать это множество слов новым алфавитом $\mathfrak{A}'$. Ясно, что всякое сообщение в алфавите $\mathfrak{A}$ можно разбить на последовательность слов длины не более *k*, а значит, закодировать в новом алфавите $\mathfrak{A}'$.

Сходную идею можно было бы провести для сообщений на русском языке, записанных в русском алфавите, дополненном знаком пропуска и знаками препинания. Для этого нужно было бы взять полный словарь русского языка и каждое слово в нем обозначить иероглифом (используя, к примеру, комбинации китайских и египетских иероглифов). Если еще при этом ввести иероглифы так, что по ним можно различать падежи и спряжения глаголов, то можно будет закодировать любое сообщение на русском языке.

Вместо иероглифов можно было бы использовать шестизначные десятичные числа. Первых пяти знаков такого числа достаточно для кодирования слов (так как заведомо можно удовлетвориться словарем из 100 000 русских слов.); шестой знак можно использовать для кодирования грамматических признаков.

Здесь мы впервые столкнулись с важным понятием кодирования и перекодирования сообщений. **Под *кодированием* понимается, вообще говоря, формирование в заданном алфавите сообщений, несущих заданную информацию или преобразование сообщений, записанных в одном алфавите, в сообщения, записанные в другом алфавите.** При этом наибольший интерес представляет взаимно однозначное перекодирование, когда по сообщению, записанному в любом из двух алфавитов, можно однозначно восстановить вид сообщения в другом алфавите. Нетрудно видеть, что рассмотренное перекодирование русских фраз из буквенного в «иероглифический» алфавит обладает указанным свойством.

Рассмотренный способ «пословного» кодирования практически используется при шифровке сообщений наряду с «по-буквенной» шифровкой.

Возможна и обратная картина, когда символы исходного алфавита $\mathfrak{A}$ кодируются в виде слов, записанных в более простом алфавите $\mathfrak{A}'$. Например, пусть алфавит $\mathfrak{A}'$ состоит из трех символов $\{\cdot, —, *\}$; читается так: точка, тире, конец буквы. Тогда любая русская (или латинская) буква и знак препинания могут быть записаны в так называемой азбуке Морзе как слово из семи — не более— символов алфавита $\mathfrak{A}'$.

Азбука Морзе *)

Символы Морзе	Русские буквы	Символы Морзе	Русские буквы
· —	а	· · — ·	ф
— · · ·	б	· · · ·	х
· — —	в	— · — ·	ц
— — ·	г	— — — ·	ч
— · ·	д	— — — —	ш
·	е	— — · —	щ
· · · —	ж	— · — —	ы
— — · ·	з	· · — —	ю
· ·	и	· — · —	я
— · —	к	· — — —	й
· — · ·	л	— · · —	ъ, ь
— —	м	· · — · ·	э
— ·	н	· — · · —	, (запятая)
— — —	о	· · · · · ·	. (точка)
· — — ·	п	— · — · — ·	; (точка с запятой)
· — ·	р	— — — · · ·	: (двоеточие)
· · ·	с	· · — — · ·	? (вопр. знак)
—	т	— — · · — —	! (воскл. знак)
· · —	у		

*) Знак конца буквы ∗ кодируется временным интервалом и поэтому не входит в приведенный в таблице код Морзе.

Таким образом, слова русского языка могут быть вместо русского алфавита записаны в алфавите Морзе.

Пример. Русская фраза «Что такое расстояние?» записывается в алфавите Морзе следующим образом:

— — — · ∗ — ∗ — — — ∗ — ∗ · — ∗ — · — ∗ — — — ∗ ·
∗ · — · ∗ · — ∗ · · · ∗ · · · ∗ — ∗ — — — ∗ · — · — ∗ — · ∗ · · ∗
· ∗ · · — — · · ∗

Оказывается (и этим часто пользуются в вычислительной технике), сообщения в любом конечном алфавите могут быть перекодированы с помощью двоичного алфавита $\mathfrak{A}_2\ \{0,\ 1\}$, состоящего только из двух символов «0» и «1».

Действительно, любое целое неотрицательное число может быть представлено в виде (29), т. е. в так называемой двоичной системе счисления. Так, каждое целое число x может быть закодировано в виде слова в двоичном алфавите.

Если целые числа меняются в диапазоне $0 \leqslant x < a$, то последовательность двоичных знаков $x \sim \varepsilon_n \varepsilon_{n-1} \varepsilon_{n-2} \ldots \varepsilon_1 \varepsilon_0$, необходимая для записи любого из чисел этого диапазона, может быть выбрана из не более чем $n+1=1+[\log_2 a]$ знаков, как это следует из (34).

Теперь, если имеется произвольный алфавит $\mathfrak{A}$ из *m* символов, то каждому этому символу можно сопоставить порядковый номер от 0 до *m*—1. Значит, каждому символу алфавита $\mathfrak{A}$ можно сопоставить двоичное слово, соответствующее, согласно (29), номеру этого символа. При этом, согласно (24), можно обойтись словами длины *n,* где

$$\log_2 (m-1) < n \leqslant 1+\log_2 (m-1). \tag{35}$$

Таким образом, в принципе можно было бы в любом случае ограничиться словами в двоичном алфавите.

Теперь введем понятие *пространства сообщении.* Рассмотрим произвольный алфавит $\mathfrak{A}$ и множество сообщений, состоящих ровно из *n* символов алфавита $\mathfrak{A}$.

Назовем *расстоянием между двумя сообщениями* ξ и η, $\varrho(\xi, \eta)$ количество позиций, в которых у сообщений ξ и η стоят различные символы. Полученное метрическое пространство $E(n, \mathfrak{A})$ называется ***n*-мерным пространством сообщений в алфавите** $\mathfrak{A}$.

Пример I. $\mathfrak{A}$—русский алфавит, *n* =6. Пусть ξ = корова; η = ворона. Не совпадают первая и пятая буквы, следовательно, $\varrho(\xi, \eta)=2$.

Пример II. $\mathfrak{A}_2$—двоичный алфавит, *n* =12 и ξ = 000110101010; η = 010101101011. Не совпадают 2-й, 5-й, 6-й и 12-й двоичные знаки, $\varrho(\xi, \eta)=4$.

Заметим, что можно было бы сравнивать любые слова длиной не более *n*, если условиться слова более короткие, чем *n,* дополнять определенным знаком (в двоичном алфавите нулем) до *n* символов.

Проверим, что определенное выше расстояние $\varrho(\xi, \eta)$ удовлетворяет всем необходимым свойствам.

Симметричность $\varrho(\xi \quad \eta)=\varrho(\eta, \xi)$ следует из определения, в котором оба слова ξ и η участвуют равноправно. Очевидно, что $\varrho(\xi, \eta)\geqslant 0$, причем $\varrho(\xi, \eta)=0$, только если, все соответствующие символы в сообщениях ξ и η совпадают, т. е. совпадают слова ξ и η.

Аксиома треугольника проверяется так.

Пусть даны три слова ξ, η и ζ по *n* символов каждое.

Предположим, что в *k*-й позиции совпадают символы у слов ξ и ζ и у слов ζ и η). Ясно, что тогда в этой позиции совпадают символы и у слов ξ и η).

Действительно, пусть ξ_k — *k*-й символ сообщения ξ, ζ_k — *k*-й символ сообщения ζ, а η_k—*k*-й символ сообщения η. Тогда, если $\xi_k=\zeta_k$, а $\zeta_k=\eta_k$, то, очевидно, $\xi_k=\eta_k$. Значит, если $\xi_k\neq\eta_k$, то либо $\xi_k\neq\zeta_k$, либо $\zeta_k\neq\eta_k$.

Таким образом, у слов ξ и η могут быть не совпадающие символы только на тех местах, где либо у ξ и ζ, либо у ζ и η не совпадают символы. Это значит, что общее количество не совпадающих символов у слов ξ и η не превосходит суммарного числа не совпадающих символов у ξ и ζ и ζ и η. Но общее число символов, не совпадающих у ξ и ζ, плюс общее число символов, не совпадающих у ζ и η, есть $\varrho(\xi,\zeta)+\varrho(\zeta,\ \eta)$. Иными словами,

$$\varrho(\xi,\ \eta)\leqslant\varrho(\xi,\ \zeta)+\varrho(\zeta,\ \eta), \tag{36}$$

что и требовалось доказать.

Пример. В пространстве $E(6,\ \mathfrak{A})$, где $\mathfrak{A}$—русский алфавит, запишем слова: ξ=картон; η=кортеж; ζ=кордон. Ясно, что

$$\varrho(\xi,\ \eta)=3;\ \varrho(\xi,\ \zeta)=2,\ \varrho(\zeta,\ \eta)=3$$

и

$$\varrho(\xi,\ \eta)\leqslant\varrho(\xi,\ \zeta)+\varrho(\zeta,\ \eta).$$

С помощью так определенного понятия расстояния можно сформулировать общий принцип построения кодов, позволяющих автоматически исправлять группы из не более чем l ошибок. Этот принцип впервые предложил Р. Хэмминг. Он будет рассмотрен в следующем параграфе.

2.6. Автоматическое исправление ошибок в сообщениях

В этом параграфе мы рассматриваем пространство сообщений $E(n,\ \mathfrak{A})$, т. е. пространство сообщений длины n в алфавите $\mathfrak{A}$. Как мы видели ранее, фактически можно ограничиться двоичными сообщениями, т. е. сообщениями в алфавите $\mathfrak{A}_2$. Все содержательные примеры будут рассматриваться именно в этом алфавите.

Рассмотрим следующую общую схему передачи информации (рис. 17).

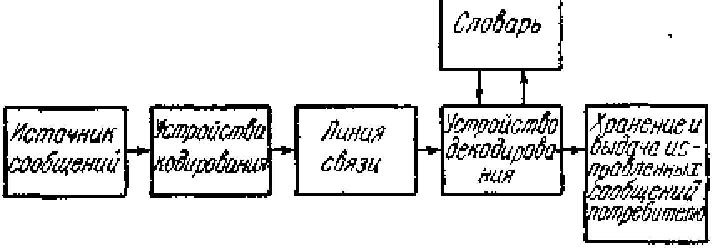


Рис. 17.

Сообщения, передаваемые источником, перекодируются в помехоустойчивый код устройством кодирования. Затем эти сообщения передаются по линиям связи, в которых возможны искажения сообщений. После этого сообщения на приемном конце исправляются в декодирующем устройстве и снова перекодируются в исходный код, если это необходимо.

Совершенно аналогично происходит автоматическое обнаружение и исправление ошибок при хранении сообщений в машинной памяти (рис. 18).

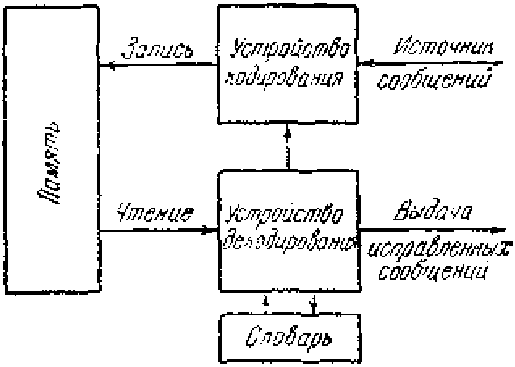


Рис. 18.

При записи информации в память сообщения кодируются в помехоустойчивом коде. При чтении сообщения проходят соответствующую дешифровку с обнаружением и исправлением ошибок, допущенных при хранении. На рис. 18 видна дополнительная возможность, состоящая в том, что информация периодически считывается, проходит дешифровку с исправлением, а затем снова шифруется и записывается в память. Эта возможность бывает полезна, ибо она обеспечивает проверку хранимой информации не реже, чем один раз за *T сек,* где *T* надо выбрать так, чтобы за это время не могло возникнуть слишком много искажений в хранимой информации. Иначе говоря, чтобы расстояние $\varrho(\xi, \xi')$ между записанным сообщением ξ, и считанным сообщением ξ' не успело стать слишком большим.

Пусть в пространстве $E(n, \mathfrak{A})$ выделено подмножество $H_k \subset E(n, \mathfrak{A})$, обладающее тем свойством, что для любых не совпадающих слов ξ и η множества H_k расстояние удовлетворяет условию

$$\varrho(\xi, \eta) \geqslant k. \tag{37}$$

Множество H_k назовем множеством осмысленных слов. Предположим, что при передаче или хранении осмысленного слова $\xi \in H_k$ допущено l ошибок $(l \leqslant k-1)$, т. е. неверно передано l символов из слова ξ. Это ошибочно переданное слово обозначим ξ'. По определению расстояния $\varrho(\xi, \xi') = l$. Ясно, что слово ξ' не осмысленно, так как в противном случае $\varrho(\xi, \xi')$ было бы больше l (согласно (37)).

Таким образом, проверив переданное слово ξ' и убедившись, что оно не осмысленно (для этого можно его, к примеру, сравнить со всеми словами из H_k — на рис. 17 и 18 эта возможность обеспечивается наличием словаря), мы обнаруживаем ошибочность передачи. При хранении слов в машинной памяти такую проверку можно осуществлять периодически, выбирая период T из того условия, что за время T мало шансов для возникновения более чем k—1 ошибок в слове. Это рассуждение показывает, что у нас уже есть общий принцип для обнаружения ошибок.

Оказывается, можно сделать больше, а именно: исправить допущенные ошибки. Для этого предположим, что число ошибок $l \leqslant \frac{k-1}{2}$.

Пусть η —произвольное осмысленное слово, не совпадающее с ξ, а ξ', как и ранее, ошибочно переданное сообщение. Запишем неравенство треугольника

$$\varrho(\xi, \eta) \leqslant \varrho(\xi, \xi') + \varrho(\xi', \eta)$$

и подставим сюда $\varrho(\xi, \xi') = l$ и оценку (37):

$$k \leqslant l + \varrho(\xi', \eta).$$

Отсюда видно, что

$$\varrho(\xi', \eta) \geqslant k - l \geqslant k - \frac{k-1}{2} = \frac{k+1}{2}, \qquad (38)$$

так как по условию $l \leqslant \frac{k-1}{2}$.

Из (38) видно, что ошибочное слово ξ' отстоит от любого осмысленного слова $\eta \neq \xi$ не менее чем на $\frac{k+1}{2}$, а от ξ не более чем на $\frac{k-1}{2}$. Следовательно, найдя ближайшее к ξ' осмысленное слово ξ, мы тем самым восстановим правильное сообщение.

Иными словами, нужно для осмысленных сообщений построить их области Дирихле.

Для каждого проверяемого (после передачи или хранения) сообщения ξ' определяется, чьей области Дирихле оно принадлежит. «Владелец» этой области Дирихле — осмысленное сообщение ξ — и считается правильным сообщением.

В этом и состоит идея Хэмминга. Она приводит к тому, что для передачи сообщений в условиях возможных искажений надо пользоваться не всеми возможными сочетаниями символов алфавита, а некоторым множеством осмысленных слов. Так как в русском языке используются в качестве осмысленных слов только некоторые сочетания букв, то восстановление смысла искаженного сообщения часто может быть выполнено без специальных дополнительных приемов кодирования. Это положение уже иллюстрировалось.

Мы сейчас рассмотрим несколько конкретных примеров построения множества осмысленных сообщений $H_k \in E(n, \mathfrak{A})$, т. е. выясним, как фактически, исходя из общего принципа Хэмминга, строить конкретные помехоустойчивые коды. Все эти примеры мы будем строить для случая двоичного алфавита $\mathfrak{A}_2$. Как мы уже видели, такое условие на самом деле не является ограничением, так как в двоичном алфавите можно записывать любые сообщения.

Задачу о помехоустойчивом кодировании можно поставить следующим образом. Пусть имеется пространство s-значных двоичных сообщений $E(s, \mathfrak{A}_2)$. Требуется каждому такому сообщению поставить в соответствие сообщение из некоторого $H_k \subset E(n, \mathfrak{A}_2)$. Иными словами, нужно построить некоторое множество осмысленных сообщений H_k, устойчивое относительно l-кратных ошибок, с помощью которого можно кодировать исходные сообщения из $E(s, \mathfrak{A}_2)$. Величина $\frac{n-s}{n}$ называется *избыточностью* кода.

В действительности точная постановка этой задачи должна исходить из заданной статистики искажений в передающем канале. При этом нужно строить такой код (множество $H_k \subset E(n, \mathfrak{A}_2)$), что вероятность получить в слове из n двоичных символов более чем l ошибок будет достаточно мала. Такая постановка задачи изучается в теории информации, но мы ею здесь заниматься не будем.

Для построения этих кодов удобно использовать тот факт, что над величинами 0 и 1 можно выполнять особую операцию сложения по модулю 2 согласно правилам:

$$0 \oplus 0 = 0, \quad 1 \oplus 0 = 1,$$
$$0 \oplus 1 = 1, \quad 1 \oplus 1 = 0.$$

Знак операции сложения обведен кружком, чтобы показать, что эта операция несколько отличается от обычного сложения. Расстояние между двумя двоичными словами $\xi = x_1 x_2 \ldots x_n$ и $\eta = y_1 y_2 \ldots y_n$ (x_i и y_i принимают значения 0 или 1) можно с помощью этой операции записать так:

$$\rho(\xi, \eta) = (x_1 \oplus y_1) + (x_2 \oplus y_2) + \ldots + (x_n \oplus y_n).$$

Действительно, в скобках здесь будут появляться единицы, когда соответствующие символы у обоих слов не совпадают, и нули в противном случае. Общее количество единиц как раз равно количеству не совпадающих символов в словах ξ и η.

Рассмотрим пространство сообщений $E(s, \mathfrak{A}_2)$ и сопоставим каждому слову $\xi \in E(s, \mathfrak{A}_2)$ слово ξ' длины s+1, образозанное по следующему правилу.

Первые s символов слова ξ' совпадают со словом ξ.

Последний, (s+l)-й символ слова ξ' выбирается равным нулю или единице с тем, чтобы общее количество единиц в слове ξ' было четным. Если положить $\xi' = x_1x_2\ldots x_sx_{s+1}$, то это условие означает, что

$$x_1 \oplus x_2 \oplus \ldots x_s \oplus x_{s+1} = 0. \quad (39)$$

Отсюда можно выразить x_{s+1}:

$$x_{s+1} = x_1 \oplus x_2 \oplus \ldots \oplus x_s. \quad (40)$$

Например, если $\xi = 001\,011\,101$, то $\xi' = 001\,011\,101\,1$.

Определенные таким образом слова ξ' образуют множество осмысленных сообщений $H_2 \subset E(s+1, \mathfrak{A}_2)$.

В самом деле, расстояние между двумя не совпадающими осмысленными сообщениями ξ' и ξ'' должно быть четным, так как, если бы ξ' отличалось от ξ'' в нечетном числе позиций, то сумма единиц, хотя бы в одном из слов ξ' или ξ'', была бы нечетной. А это невозможно по построению этих слов. Так как минимальное четное число, не равное нулю, есть 2, то минимальное расстояние между различными осмысленными сообщениями равно двум.

Следовательно, этот код позволяет обнаруживать единичные ошибки. Процесс обнаружения состоит просто в том, что после передачи слова ξ' подсчитывается четность единиц в нем (или, что равносильно, проверятся (39)). Если критерий четности не удовлетворяется ((39) не выполнено), то принятое слово считается ошибочным. Этот процесс проверки широко известен как «проверка на четность» и очень часто применяется ввиду его простоты. Избыточность в этом случае равна $\frac{1}{s+1}$.

Опишем теперь принадлежащий Хэммингу пример кода, способного исправлять единичные ошибки.

Пусть $\xi \in E(s, \mathfrak{A}_2)$—двоичное слово длины s. Образуем слово $\xi' \in E(n, \mathfrak{A}_2)$ по следующему правилу. Среди n-позиций, которые занимает слово ξ', мы выделим 1-ю, 2-ю, 4-ю,..., 2^k-ю позицию для контрольных символов, которые определяются по слову ξ. Между этими позициями мы будем последовательно записывать символы, образующие слово:

$$\xi = 10011100100110111111101100010010.$$

В написанном слове показано взаимное расположение контрольных (выделенных жирным шрифтом) и смысловых позиций для случая $s=25$, $n=31$, $k=4$. Для того чтобы разместить нужным образом s смысловых позиций, необходимо и достаточно, чтобы выполнялось соотношение

$$2^k \leqslant s+k<2^{k+1}. \tag{41}$$

Избыточность данного кода равна $\frac{k}{s+k}$. Контрольная позиция с номером 2^i заполняется по следующему правилу. Каждая позиция слова ξ' определяется номером l, отсчитываемым от начала слова. Рассмотрим двоичное представление этого номера:

$$l=l_k 2^k+l_{k-1}2^{k-1}+\ldots+l_1 2+l_0$$

(число двоичных разрядов в представлении номера l определяется тем, что согласно (41) $l<2^{k+1}$).

Возьмем теперь множество П всех тех позиций, для номеров которых l_i = 1. В множество П$_i$ входит одна контрольная позиция, а именно позиция с номером $l = 2^i$. Заполним ее так, чтобы сумма всех единиц в позициях из П$_i$ была четной.

В следующей таблице приведен пример слова ξ' (его можно прочесть по вертикали во втором столбце таблицы), указаны двоичные номера позиций и отмечена знаком ∗ принадлежность каждой позиции к множеству П$_i$. Слова ξ', построенные по этому правилу, назовем осмысленными.

№ позиции	Содержимое позиции	$П_0$	$П_1$	$П_2$	$П_3$	$П_4$
00 001		*				
00 010			*			
00 011	0	*	*			
00 100				*		
00 101	1	*		*		
00 110	1		*	*		
00 111	0	*	*	*		
01 000					*	
01 001	1	*			*	
01 010	0		*		*	
01 011	0	*	*		*	
01 100	1			*	*	
01 101	1	*		*	*	
01 110	0		*	*	*	
01 111	1	*	*	*	*	
10 000						*
10 001	1	*				*
10 010	1		*			*
10 011	1	*	*			*
10 100	1			*		*
10 101	0	*		*		*
10 110	1		*	*		*
10 111	1	*	*	*		*
11 000	0				*	*

Продолжение

№ позиции	Содержимое позиции	$П_0$	$П_1$	$П_2$	$П_3$	$П_4$
11 001	0	*			*	*
11 010	0		*		*	*
11 011	1	*	*		*	*
11 100	0			*	*	*
11 101	0	*		*	*	*
11 110	1		*	*	*	*
11 111	0	*	*	*	*	*

Покажем, что расстояние между любыми двумя осмысленными словами ξ' и η' не менее чем 3, то есть осмысленные слова образуют множество $H_3 \subset E(s+k, \mathfrak{A}_2)$.

I случай. Пусть соответствующие слова ξ и η из $E(s, \mathfrak{A}_2)$ отличаются не менее чем в трех позициях. Ясно, что слова ξ' и η' также отличаются не менее чем в трех позициях, а следовательно, $\varrho(\xi', \eta') \geqslant 3$.

II случай. Пусть слова ξ и η отличаются в двух позициях. Тем самым слова ξ' и η' отличаются в двух смысловых позициях с номерами l и l'. Выберем множество позиций Π_i, содержащее позицию l и не содержащее позиции l'. Для этого нужно номер i выбрать так, чтобы i-й двоичный знак числа l, l_i=1, а i-й двоичный знак числа l', $l'_i=0$ (так как $l \neq l'$, то у них обязательно есть не совпадающие двоичные знаки).

Таким образом, у слов ξ' и η' количество единиц в смысловых позициях множества Π_i отличается на одну. Так как общее число единиц во множестве позиций Π_i должно быть четным и у ξ', и у η', то слова ξ' и η' должны отличаться еще и в контрольной позиции множества Π_i (з позиции с номером 2^i). Тем самым ξ' и η' отличаются не менее чем в трех позициях, $\varrho(\xi', \eta') \geqslant 3$.

III случай. Пусть слова ξ и η отличаются в одной позиции. Тогда слова ξ' и η' отличаются ровно в одной смысловой позиции с номером l. Этот номер не может быть точной степенью двойки, так как номера вида 2^i закреплены за контрольными позициями. Таким образом, у номера l есть по крайней мере два ненулевых двоичных знака $l_i=1$ и $l_j=1$. Следовательно, позиция с номером l входит в два множества Π_i и Π_j. Так как количество единиц в этих множествах позиций и для ξ', и для η' должно быть четным, то кроме l-й позиции слова ξ' и η' обязательно должны отличаться и в контрольных позициях с номерами 2^i и 2^j.

Всего, таким образом, слова ξ' и η' отличаются не менее чем в трех позициях, $\varrho(\xi', \eta') \geqslant 3$.

Итак, доказано, что множество осмысленных сообщений есть H_3 и, следовательно, позволяет принципиально восстанавливать единичные ошибки.

В данном случае этот процесс восстановления может быть сравнительно просто выполнен. В основе этого лежит такое соображение. Для исправления слова в двоичном алфавите достаточно узнать номер позиции, где допущена ошибка и переменить содержимое в этой позиции (если был «0», поставить «1» и наоборот).

В рассматриваемом коде номер неисправной позиции устанавливается путем следующих проверок.

После передачи сообщения ξ', в результате которой возможно искажение в одной позиции $\xi' \rightarrow \xi^*$, проверяется четность единиц для каждого множества позиций Π_i.

Иными словами, вычисляются контрольные величины

$$\alpha_l = \xi_1^* \oplus \xi_2^* \oplus \xi_3^* \oplus \ldots,$$

т. е. α_i равна сумме по модулю 2 всех символов в позициях множества Π_i принятого сообщения ξ^*.

Если все $\alpha_i = 0$, то ξ^* является осмысленным сообщением (то есть если и была допущена ошибка, то не менее чем в трех позициях, и ее нельзя исправить с помощью данного кода.). Если же некоторое $\alpha_i = 1$, то ошибка допущена в позиции, принадлежащей множеству Π_i, а значит, номер l этой позиции имеет i-й двоичный знак, равный 1, l_i=1.

Таким образом, контрольные величины α_i равны знакам двоичного разложения номера ошибочной позиции l, т. е.

$$l = \alpha_k 2^k + \alpha_{k-1} 2^{k-1} + \ldots + \alpha_1 2 + \alpha_0. \tag{42}$$

Это значит, что после всех контрольных вычислений номер l определен и дает возможность восстановить правильное сообщение ξ'.

Для примера возьмем слово ξ', записанное в предыдущей таблице, и исказим его в 19-й позиции.

Получится слово

$$\xi^* = 100111001001101101 0 \quad 10110001001 0;$$

осуществляя все проверки, получаем

$$\alpha_4 = 1,\ \alpha_3 = 0,\ \alpha_2 = 1,\ \alpha_1 = 0,\ \alpha_0 = 1,$$

т. е. l= 10 101, что соответствует десятичному числу 19. Переменив в 19 позиции слова ξ^* 0 на 1, получаем передававшееся слово ξ'.

Более простой код с исправлением одиночных ошибок будет, если слово ξ' получается трехкратным повторением слова $\xi \in E(s, \mathfrak{A}_2)$. Если ξ и η отличаются в r позициях, то соответствующие ξ' и η' отличаются в $3r$ позициях.

Таким образом, $\varrho(\xi', \eta') \geqslant 3$, если $\xi' \neq \eta'$. Проверка переданного слова осуществляется так.

Берутся тройки позиций с номерами l, $l+s$, $l+2s$. Если символы в этих позициях совпадают, то соответствующий символ считается правильно переданным. Если совпадают два из этих символов, то их общее значение считается правильным и заносится в третью позицию.

Если все три символа не совпадают, то ошибку исправить нельзя.

Таким образом, этот код может исправлять единичные ошибки в каждой тройке соответствующих позиций и обнаруживать двойные ошибки в каждой тройке.

Недостатком этого кода является его большая избыточность, равная $\frac{2s}{3s} = \frac{2}{3}$. Избыточность предыдущего кода примерно равна $\frac{\log_2 s}{s}$, что при больших s (длинных сообщениях) близко к нулю.

Коды, позволяющие исправлять ошибки при передаче и хранении сообщений, очень важны для различных устройств автоматического управления.

2.7. Расстояния и нормы в многомерном пространстве

В этом параграфе рассматривается так называемое n-мерное векторное пространство E_n и различные определения расстояния в нем, превращающие E_n в различные метрические пространства. Векторное пространство E_n служит обобщением понятий прямой, плоскости и трехмерного пространства из элементарной геометрии. Прийти к разумному определению n-мерного пространства можно следующим образом.

Рассмотрим плоскость с выбранной системой декартовых координат. Каждая точка M на плоскости однозначно определяется своими координатами (x, y).

Каждой точке M однозначно соответствует вектор (см. рис. 19), соединяющий начало координат с этой точкой.

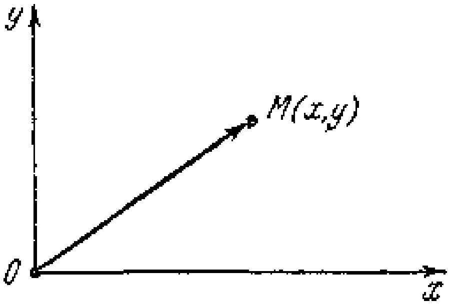


Рис. 19.

Таким образом, существует взаимное однозначное соответствие между следующими объектами:

$$\text{точка } M \leftrightarrow \text{вектор } \overline{OM} \leftrightarrow \text{пара координат } (x,\ y).$$

Это значит, что можно с одинаковым успехом говорить о плоскости как о множестве точек, как о множестве векторов или как о множестве пар чисел $(x_y\ y)$ (то есть говорить об одном и том же математическом объекте — плоскости — на языке точек, на языке векторов или на языке числовых пар. Любое утверждение может быть переведено с каждого языка на каждый.). Аналогично, о трехмерном пространстве можно говорить как о множестве троек чисел *(x, y, z)*. Теперь уже сами собой напрашиваются такие определения.

n-мерным вектором называется упорядоченная совокупность из *n* вещественных чисел

$$\xi=(x_1,\ x_2,\ \dots,\ x_n).$$

Числа $x_1, x_2, \dots, x_n$ называются *координатами вектора* ξ. Множество всех *n*-мерных векторов называется *n-мерным векторным пространством* E_n.

Ясно, что векторное пространство E_3 есть обычное трехмерное пространство, пространство E_2 есть плоскость, а пространство E_1— прямая.

Над векторами в *n*-мерном пространстве можно выполнять операции сложения, вычитания и умножения на вещественное число. Эти операции определяются следующим образом.

Суммой векторов $\xi=(x_1,\ x_2,\ \dots,\ x_n)$ и $\eta=(y_1,\ y_2,\dots\ \dots,y_n)$ называется вектор

$$\zeta=\xi+\eta=(x_1+y_1,\ x_2+y_2,\dots,\ x_n+y_n),$$

у которого координаты равны суммам соответствующих координат слагаемых.

Аналогично, *разностью тех же векторов* ξ и η называется вектор

$$\theta=\xi-\eta=(x_1-y_1,\ x_2-y_2,\dots,\ x_n-y_n),$$

у которого координаты равны разностям соответствующих координат.

Произведением вектора $\xi=(x_1,\ x_2,\dots,\ x_n)$ на число *a* называется вектор

$$\varphi=a\,\xi=(ax_1,\ ax_2,\ \dots,ax_n).$$

Иными словами, чтобы умножить вектор на число, нужно умножить на это число все его координаты. На плоскости и в трехмерном пространстве эти операции имеют простой геометрический смысл. Для наглядности рассмотрим два вектора ξ и η на плоскости (см. рис. 20).

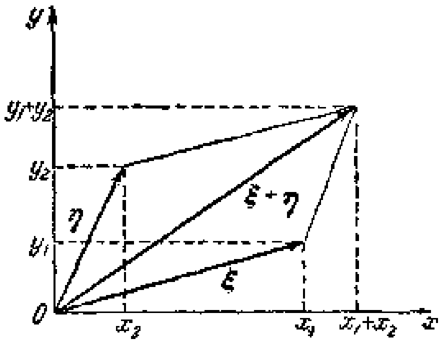


Рис. 20.

Из рисунка видно, что сумма $\zeta = \xi + \eta$ есть вектор, образованный диагональю параллелограмма, построенного на векторах ξ и η. Такое определение суммы векторов возникает в физике при сложении сил.
Разность векторов ξ и η (см. рис. 21) есть вектор, направленный из конца вектора η в конец вектора ξ.

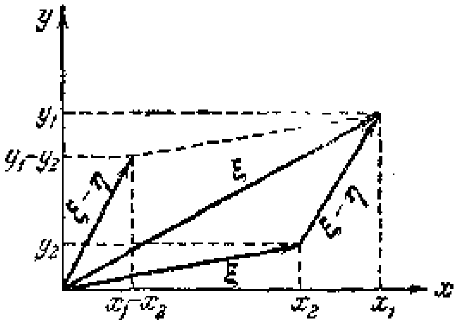


Рис. 21.

Произведение вектора ξ на положительное число *a* есть вектор, имеющий то же направление, а длину, в *a* раз большую, чем ξ. (Ясно, что при $a< 1$ длина вектора $a\xi$ меньше длины вектора ξ.) Чтобы умножить вектор ξ на отрнцательное число *a*, надо умножить его на $|\ a\ |$, а затем изменить направление на противоположное. Все эти случаи показаны на рис. 22.

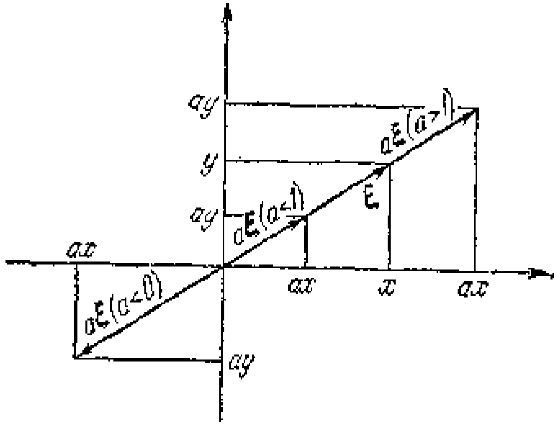


Рис. 22.

Легко показать, что операции над *n*-мерными векторами удовлетворяют следующим свойствам, аналогичным свойствам соответствующих числовых операций.

1. $\xi+\eta=\eta+\xi$ (коммутативность),
2. $\xi+(\eta+\zeta)=(\xi+\eta)+\zeta$ (ассоциативность),
3. $\xi-\xi=\mathbf{0}$,
4. $\mathbf{0}+\xi=\xi$,
5. $\lambda(\xi+\eta)=\lambda\xi+\lambda\eta$ (дистрибутивность),
6. $(\lambda+\mu)\xi=\lambda\xi+\mu\xi$ (дистрибутивность),
7. $\lambda(\mu\xi)=(\lambda\mu)\xi$ (ассоциативность умножения),
8. $0\cdot\xi=\mathbf{0}$,
9. $\lambda\cdot\mathbf{0}=\mathbf{0}$,
10. $1\cdot\xi=\xi$.

Здесь использовано обозначение для нулевого вектора $\mathbf{0}=(0, 0, \ldots, 0)$. Ясно, что два вектора ξ и η совпадают в том и только том случае, когда $\xi-\eta=0$.

Рассмотрим несколько примеров многомерных пространств, естественно возникающих в геометрии.

Пример I. Множество всех шаров в пространстве. Каждый шар определяется четырьмя параметрами *(x, y, z, R)*, где *(x, y, z)* — координаты центра, а *R* — радиус.

Пример II. Множество всех треугольников. Каждый. треугольник определяется девятью параметрами

$(x_1, y_1, z_1, x_2, y_2, z_2, x_3, y_3, z_3)$, где каждая тройка (x_i, y_i, z_i) состоит из координат некоторой вершины. Предоставляем читателю убедиться, что в обоих случаях умножение всех координат (соответственно четырех- и девятимерного вектора) на число λ равносильно подобному преобразованию с центром подобия в начале координат.

Теперь рассмотрим различные способы определения расстояния в E_n, превращающие E_n в различные метрические пространства.

Пространство $l_2^{(n)}$ получается, если расстояние определить формулой

$$\varrho_2(\xi, \eta) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2}, \quad (43)$$

где

$$\xi = (x_1, x_2, \ldots, x_n), \text{ а } \eta = (y_1, y_2, \ldots, y_n).$$

В трехмерном пространстве и на плоскости расстояние $\varrho_2(\xi, \eta)$ совпадает с обычным геометрическим расстоянием. Свойства 1, 2, 3 для этого расстояния очевидны.

Пространство $l_1^{(n)}$ определяется с помощью расстояния

$$\varrho_1(\xi, \eta) = |x_1 - y_1| + |x_2 - y_2| + \ldots + |x_n - y_n|. \quad (44)$$

В случае плоскости это расстояние совпадает с расстоянием $\varrho_1(M, N)$, определенным ранее. Опять-таки, свойства 1, 2 н 3 для расстояния $\varrho_1(\xi, \eta)$ очевидны.

Пространство $C^{(n)}$ возникает, если ввести расстояние по правилу

$$\varrho_\infty(\xi, \eta) = \max(|x_1 - y_1|, |x_2 - y_2|, \ldots |x_n - y_n|), \quad (45)$$

т. е. как максимальное отклонение координат соответствующих векторов. Свойства 1, 2 и 3 очевидны и для этого расстояния, которое на плоскости совпадает с введенным ранее расстоянием $\varrho_\infty(M, N)$.

Докажем аксиому треугольника для пространства $l_1^{(n)}$.

Пусть

$$\xi = (x_1, x_2, \ldots, x_n); \ \eta = (y_1, y_2, \ldots, y_n); \ \zeta = (z_1, z_2, \ldots, z_n).$$

Тогда, очевидно,

$$\begin{aligned}\varrho_1(\xi, \eta) &= |x_1 - y_1| + |x_2 - y_2| + \ldots + |x_n - y_n| = \\ &= |x_1 - z_1 + z_1 - y_1| + |x_2 - z_2 + z_2 - y_2| + \ldots + \\ &+ |x_n - z_n + z_n - y_n| \leqslant |x_1 - z_1| + |z_1 - y_1| + |x_2 - z_2| + \\ &+ |z_2 - y_2| + \ldots + |x_n - z_n| + |z_n - y_n| = \\ &= \varrho_1(\xi, \zeta) + \varrho_1(\zeta, \eta),\end{aligned}$$

что и требовалось доказать.

Для пространства $C^{(n)}$ аксиома треугольника доказывается так. Пусть $|x_k - y_k|$ самая большая из соответствующих разностей координат, т. е.

$$\varrho_\infty(\xi, \eta) = \max(|x_1 - y_1|, |x_2 - y_2|, \ldots, |x_n - y_n|) = \\ = |x_k - y_k| = |x_k - z_k + z_k - y_k| \leqslant \\ \leqslant |x_k - z_k| + |z_k - y_k|. \qquad (46)$$

Очевидно, что

$$\left.\begin{aligned} |x_k - z_k| \leqslant \max(|x_1 - z_1|, |x_2 - z_2|, \ldots, |x_n - z_n|) = \\ = \varrho_\infty(\xi, \zeta), \\ |z_k - y_k| \leqslant \max(|z_1 - y_1|, |z_2 - y_2|, \ldots, |z_n - y_n|) = \\ = \varrho_\infty(\zeta, \eta). \end{aligned}\right\} \qquad (47)$$

Сравнивая (46) и (47), получаем искомое соотношение

$$\varrho_\infty(\xi, \eta) \leqslant \varrho_\infty(\xi, \zeta) + \varrho_\infty(\zeta, \eta).$$

Более общий вид метрических пространств получится, если ввести расстояние по формуле

$$\varrho_p(\xi, \eta) = \sqrt[p]{(x_1 - y_1)^p + (x_2 - y_2)^p + \ldots + (x_n - y_n)^p},$$

где $p \geq 1$, и получить таким образом пространство $l_p^{(n)}$.

Свойства 1, 2 и 3 проверяются здесь столь же легко. Неравенство треугольника (свойство 4) получается из неравенства Минковского:

$$\sqrt[p]{(a_1 - b_1)^p + (a_2 - b_2)^p + \ldots + (a_n - b_n)^p} \leqslant \\ \leqslant \sqrt[p]{a_1^p + a_2^p + \ldots + a_n^p} + \sqrt[p]{b_1^p + b_2^p + \ldots + b_n^p}.$$

Нетрудно видеть, что при $p = 1$ и $p=2$ получаются выше определенные пространства $l_1^{(n)}$ и $l_2^{(n)}$. При $p \to \infty$ расстояние $\varrho_p(\xi, \eta)$ переходит в расстояние $\varrho_\infty(\xi, \eta)$, т. е. $l_\infty^{(n)} = C^{(n)}$.

В качестве упражнения читателю предлагается показать, что обобщенный шар радиуса r в пространстве $l_1^{(3)}$ имеет вид октаэдра (рис. 23), а в пространстве $C^{(3)}$ имеет вид куба (рис. 24).

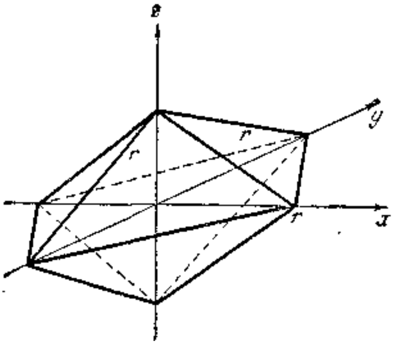


Рис. 23.

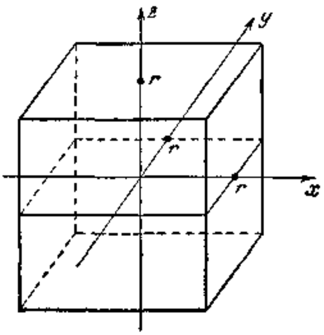


Рис. 24.

Пространства $l_p^{(n)}$, так же как и соответствующие пространства l_p на плоскости, называются пространствами Минковского. Эти пространства допускают обобщения на случай бесконечного количества координат у вектора.

Более общий вид расстояния можно ввести с помощью понятия выпуклого тела.

Введем несколько новых определений. Каждый вектор из векторного пространства E_n мы будем геометрически интерпретировать как точку, соответствующую концу этого вектора, считая сам вектор выходящим из начала координат. Поэтому каждому факту, относящемуся к векторам, соответствует некоторое свойство точек n-мерного пространства. Назовем подмножество V пространства E_n *выпуклым,* если вместе с каждыми двумя векторами $\xi \in V$ и $\eta \in V$ оно содержит любой вектор вида $a\xi + (1 - a)\eta$, где a — любое число между нулем и единицей $(0 \leqslant a \leqslant 1)$. Геометрически (в E_2 и в E_3) это означает, что множество V вместе с любой парой точек содержит и весь отрезок, соединяющий эти точки.

Назовем подмножество $V \subset E_n$ *ограниченным,* если существует такое общее для всех векторов число K, что координаты всех векторов $\xi \in V$ удовлетворяют неравенствам

$$|x_1| < K; \ |x_2| < K; \ \ldots \ |x_n| < K.$$

На рис. 25 изображено выпуклое, но не ограниченное множество на плоскости.

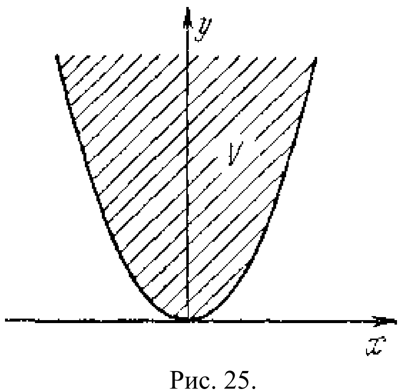


Рис. 25.

А на рис. 26 — выпуклое и ограниченное множество.

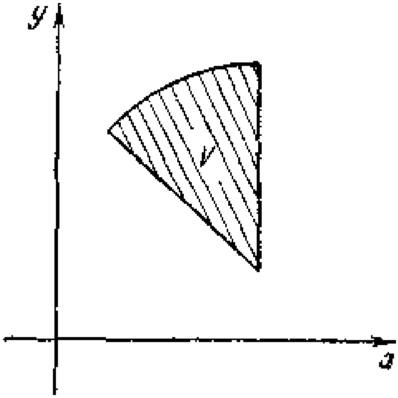


Рис. 26.

Вектор ξ, принадлежащий множеству V, называется *внутренним* для множества V, если для всякого вектора η можно подобрать такое положительное число a, что век тор $\xi + a\eta$ также входит в множество V. Это значит, что если из конца вектора ξ двигаться по любому направлению, то некоторое время мы будем еще находиться во множестве V.

Если V—плоская фигура, расположенная в пространстве E_3, то никакая его точка не является внутренней. Действительно (см. рис. 27), если $\xi \in V$, а вектор η перпендикулярен плоскости, где лежит V, то при любом a вектор $\xi + a\eta$ лежит вне этой плоскости, а следовательно, и вне фигуры V.

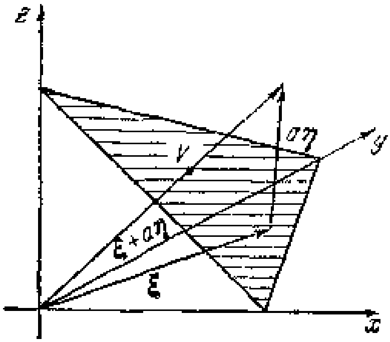


Рис. 27.

Наличие во множестве *V* внутренних точек делает его *«полноразмерным»*.

Назовем, наконец, *выпуклым телом* выпуклое ограниченное множество *V*, имеющее хотя бы одну внутреннюю точку.

(Читателю предоставляется доказать, что в этом случае множество *V* обязательно имеет бесконечно много внутренних точек.)

Рассмотрим выпуклое тело *V*, имеющее *O* своей внутренней точкой и симметричное относительно этой точки. Последнее означает, что вместе с каждым вектором ξ в теле *V* содержится и противоположный вектор $-\xi$.

С помощью этого выпуклого тела можно ввести расстояние $\varrho_V(\xi, \eta)$. Для этого используем такую конструкцию. Образуем разность векторов $\zeta = \xi - \eta$. Так как *O* является внутренней точкой тела *V*, то существует такое положительное число *a*, при котором $a\zeta \in V$. С другой стороны, из ограниченности тела *V* следует, что при достаточно больших *a* вектор $a\zeta$, лежит вне тела *V*, если сам вектор $\zeta \neq 0$. Определим расстояние $\varrho_V(\xi, \eta)$ как нижнюю грань величин $\frac{1}{a}$, $a > 0$, для которых вектор $a\zeta$, еще принадлежит телу *V*,

$$\varrho_V(\xi, \eta) = \inf_{a\zeta \in V} \frac{1}{a}. \tag{48}$$

При $\zeta = 0$, т. е. когда векторы ξ и η совпадают, допустимо любое *a* (так как всегда $a\zeta = 0 \in V$) и нижняя грань в (48) равна нулю. При несовпадающих ξ и η вектор $\zeta \neq 0$ и допустимые значения *a* ограничены сверху. С другой стороны, при достаточно малых $a > 0$, $a\zeta \in V$; поэтому значения величины $\frac{1}{a}$ также ограничены сверху. Следовательно, в этом случае величина $\varrho_V(\xi, \eta)$ конечна, положительна и не равна нулю.

Из симметричности выпуклого тела *V* следует, что векторы $a\zeta$ и $a(-\zeta)$ принадлежат телу *V* при одних и тех же значениях *a*. Так как $-\zeta = \eta - \xi$, то отсюда следует симметричность расстояния

$$\varrho_V(\xi, \eta) = \varrho_V(\eta, \xi).$$

Итак, для определенного нами с помощью выпуклого тела *V* расстояния доказаны свойства 1, 2 и 3. Осталось доказать неравенство треугольника.

Это можно сделать следующим образом.

Рассмотрим три вектора ξ, η и ζ. Выберем два положительных числа *a* и *b* такие, что $a(\xi - \zeta) \in V$ и $b(\zeta - \eta) \in V$.

Обозначим теперь через *a* число

$$\alpha = \frac{\frac{1}{a}}{\frac{1}{a}+\frac{1}{b}} = \frac{b}{a+b}.$$

Ясно, что $0 < \alpha < 1$ и

$$1-\alpha = \frac{a}{a+b}.$$

В силу выпуклости тела *V* вектор

$$\theta = \alpha\left[a(\xi-\zeta)\right] + (1-\alpha)\left[b(\zeta-\eta)\right] \qquad (49)$$

также содержится в *V*. Преобразуем теперь выражение (49), подставляя вместо *a* и 1—*a* их значения и пользуясь свойствами операций над векторами:

$$\theta = \frac{b}{a+b}\cdot a(\xi-\zeta) + \frac{a}{a+b}\cdot b(\zeta-\eta) =$$

$$= \frac{ab}{a+b}\left[\xi-\zeta+\zeta-\eta\right] = \frac{ab}{a+b}(\xi-\eta) = c(\xi-\eta). \qquad (50)$$

Так как вектор $\theta = c(\xi-\eta) \in V$, то число

$$\frac{1}{c} = \frac{a+b}{ab} = \frac{1}{b}+\frac{1}{a}$$

может быть только больше расстояния $\varrho_V(\xi, \eta)$:

$$\varrho_V(\xi, \eta) \leqslant \frac{1}{a}+\frac{1}{b}. \qquad (51)$$

Однако в силу определения чисел *a* и *b* величины $\frac{1}{a}$ и $\frac{1}{b}$ можно сделать сколь угодно близкими, соответственно к расстояниям $\varrho_V(\xi, \zeta)$ и $\varrho_V(\zeta, \eta)$. Так как при переходе к пределу знак неравенства сохраняется, то из (51) получается искомое неравенство

$$\varrho_V(\xi, \eta) \leqslant \varrho_V(\xi, \zeta) + \varrho_V(\zeta, \eta), \qquad (52)$$

справедливое для любых троек векторов.

Определение расстояния с помощью симметричного выпуклого тела *V* можно изложить в несколько иных терминах.

Назовем *нормой вектора* ξ величину

$$\|\xi\|_V = \inf_{a\xi\in V}\frac{1}{a}. \qquad (53)$$

Ясно, что определенное выше расстояние $\varrho_V(\xi, \eta)$ равно норме разности векторов ξ и η

$$\varrho_V(\xi, \eta) = \|\xi-\eta\|_V. \qquad (54)$$

Можно показать (оставляем доказательство читателю), что норма удовлетворяет следующим свойствам:

1. $\|\xi\| \geqslant 0$;
2. $\|\xi\|_V = 0$ тогда и только тогда, когда $\xi=\mathbf{0}$;
3. $\|a\xi\|_V = |a|\,\|\xi\|_V$, если *a* — любое вещественное число;
4. $\|\xi+\eta\|_V \leqslant \|\xi\|_V + \|\eta\|_V$.

Можно было бы подойти к понятию нормы иным, более абстрактным способом. Назовем *нормой* в *n*-мерном векторном пространстве E_n число $\|\xi\|$, поставленное в соответствие каждому вектору ξ, так что при этом выполнены следующие свойства:

1. $\|\xi\| \geqslant 0$;
2. $\|\xi\| = 0$ тогда и только тогда, когда $\xi = 0$;
3. $\|a\xi\| = |a|\,\|\xi\|$;
4. $\|\xi+\eta\| \leqslant \|\xi\| + \|\eta\|$.

Векторное пространство E_n с определенной в нем нормой называется *n-мерным пространством Минковского.* Оказывается, что норма всегда может быть определена с помощью некоторого симметричного выпуклого тела *V*. Проверим это утверждение. Рассмотрим множество *V*, состоящее из всех векторов ξ, для которых $\|\xi\| \leqslant 1$. Покажем, что это множество является выпуклым симметричным телом в E_n.

Выпуклость множества *V* проверяется так. Пусть $\xi \in V$ и $\eta \in V$, и пусть $0 \leqslant a \leqslant 1$. Тогда по свойствам нормы 3 и 4

$$\|a\xi + (1-a)\eta\| \leqslant \|a\xi\| + \|(1-a)\eta\| = a\|\xi\| + (1-a)\|\eta\|.$$

Так как $\|\xi\| \leqslant 1$ и $\|\eta\| \leqslant 1$, то

$$\|a\xi + (1-a)\eta\| \leqslant a + (1-a) = 1,$$

т. е. вектор $a\xi + (1-a)\eta$ также принадлежит множеству *V*, что и означает выпуклость последнего.

Проверим, что *O* является внутренней точкой множества *V*. Действительно, если ξ любой вектор, не равный нулю, то, выбрав число $a = \frac{1}{\|\xi\|}$, мы убедимся, что

$$\|a\xi\| = a\|\xi\| = \frac{1}{\|\xi\|}\|\xi\| = 1,$$

т. е. $a\xi \in V$. Если же $\xi = \mathbf{0}$, то для любого числа *a*, $a\xi \in V$, так как по свойству 2 $\|a\xi\| = \|\mathbf{0}\| = 0 \leqslant 1$.

Симметричность множества *V* следует из свойства 3. Действительно, если $\xi \in V$, то $\|\xi\| \leqslant 1$, но тогда $\|-\xi\| = |-1|\,\|\xi\|$, т. е. $-\xi \in V$.

Ограниченность множества *V* доказывается несколько более громоздко, и мы это доказательство опустим.

Итак, множество V есть выпуклое тело и с помощью него можно определить норму $\|\xi\|_V$. Покажем, что эта норма совпадает с нормой $\|\xi\|$. Пусть a — такое положительное число, при котором $a\xi \in V$. Это значит, что $\|a\xi\| \leqslant 1$, а значит, $a\|\xi\| \leqslant 1$ и $\frac{1}{a} \geqslant \|\xi\|$. Нижняя грань величины $\frac{1}{a}$ — достигается, если положить $a = \frac{1}{\|\xi\|}$, таким образом:

$$\inf_{a\xi \in V} \frac{1}{a} = \|\xi\|. \tag{55}$$

Сравнивая равенства (55) и (53), убеждаемся, что

$$\|\xi\| = \|\xi\|_V, \tag{56}$$

что и требовалось доказать.

В случае трехмерного пространства E_3 норме, определяемой с помощью произвольного выпуклого тела V, можно дать простое физическое истолкование.

Предположим, что имеется некоторая анизотропная среда, в которой звук распространяется с разной скоростью по разным направлениям. При этом скорость звука в любых двух противоположных направлениях одинакова.

Будем теперь из начала координат откладывать по каждому направлению вектор, равный величине скорости звука c вдоль этого направления. Предположим, что совокупность концов этих векторов ограничивает некоторое выпуклое тело V. Нетрудно видеть, что тело V ограничено, центрально-симметрично относительно начала координат и имеет последнее своей внутренней точкой. Тем самым в E_3 определяется норма $\|\xi\|_V$ и расстояние

$$\rho_V(\xi, \eta) = \|\xi - \eta\|_V.$$

Читателю предоставляется проверить, что это расстояние равно времени, за которое звуковой сигнал проходит из конца вектора ξ, в конец вектора η по прямой, соединяющей эти концы.

Кроме конечномерных пространств Минковского, в математике рассматриваются и аналогичные **бесконечномерные— так называемые банаховы пространства.**

Банаховым пространством **называется множество элементов (векторов), над которыми можно выполнять операции сложения, вычитания и умножения на число и для которых определено понятие нормы.**

При этом нужно, чтобы операции над векторами удовлетворяли всем условиям, перечисленным ранее, а норма обладала свойствами, указанными ранее.

Пример бесконечномерного банахова пространства можно построить следующим образом.

Назовем вектором любую непрерывную функцию *f(t)*, определенную при $0 \leqslant t \leqslant 1$. Суммой векторов назовем сумму функций $f(t) + g(t)$. Произведением вектора *f(t)* на число *а* назовем функцию $af(t)$, получаемую умножением всех ее значений на число *а*. Нулевым вектором является функция $f_0(t) \equiv 0$, тождественно равная нулю. Нормой функции назовем максимум ее модуля

$$\|f(t)\| = \max_{0 \leqslant t \leqslant 1} |f(t)|.$$

Читатель легко сумеет убедиться, что все свойства операций над векторами и все свойства нормы здесь выполнены.

Это пространство обычно обозначается как $C_{[0,\,1]}$, или просто *С*.

Банаховы пространства, в которых векторами являются функции, играют большую роль в математике.

То обстоятельство, что метрическое пространство, точками которого являются функции, является бесконечномерным, можно пояснить следующим рассуждением.

Рассмотрим отрезок [0, 1] и проведем через какие-то из точек этого отрезка вертикальные прямые (рис. 28).

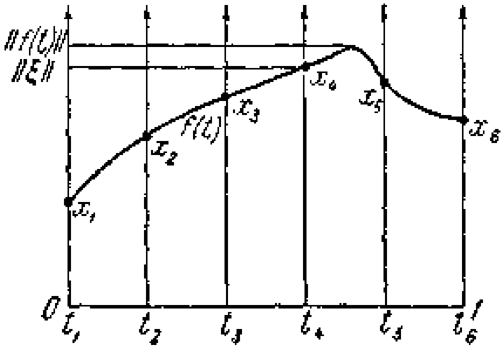


Рис. 28.

Возьмем вектор $\xi = (x_1, \ x_2, \ \ldots, \ x_n)$, принадлежащий *n*-мерному пространству, и будем откладывать его координаты на этих вертикальных прямых.

Полученные точки на вертикалях образуют график некоторой функции, заданной в соответствующих точках отрезка [0, 1]. Отсюда видно, что при $n \to \infty$ эти наборы точек могут приближаться к графикам непрерывных функций. (При этом надо выбирать только «гладкие» наборы точек, т. е. такие, что ординаты на близких вертикалях близки.) Если в E_n взять норму $C^{(n)}$, т. е. $\|\xi\| = \max(|x_1|, |x_2|, \ldots\ldots, |x_n|)$, то «в пределе» эта норма переходит в норму C:

$$\|f(t)\| = \max_{0 \leqslant t \leqslant 1} |f(t)|.$$

2.8. Сглаживание ошибок экспериментальных измерений

При наблюдении той или иной физической величины результатом эксперимента обычно является серия измеренных значений этой величины $(x_1, x_2, \ldots, x_n)$.

Сама эта величина может быть стационарной или переменной. В последнем случае значения $x_1, x_2, \ldots, x_n$ теоретически должны получаться изменяющимися по какому-то закону, а в первом случае постоянными. Кроме того, в величинах $x_1, x_2, \ldots, x_n$ содержатся еще ошибки измерений, искажающие картину.

Эту ситуацию можно было трактовать как получение некоего сообщения от природы, на которое из-за несовершенства эксперимента налагаются искажения.

Задача математической обработки измерений состоит в том, чтобы в какой-то степени восстановить правильное сообщение. К решению этой задачи можно применить соображения, уже использованные для автоматического исправления ошибок в дискретных сообщениях.

Если измеряемые величины могут принимать любые вещественные значения, то **в качестве пространства сообщений можно взять *n*-мерное векторное пространство *E_n*.** Расстояние между точками этого пространства $\varrho(\xi, \eta)$ определяется из конкретных свойств проводимого эксперимента. Наиболее часто используется расстояние вида

$$\varrho(\xi, \eta) = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}, \qquad (57)$$

при котором пространство сообщений является пространством l_2^n.

Теоретически возможные сообщения образуют подмножество H пространства сообщений.

В качестве правильного сообщения выбирается вектор $\eta \in H$, «ближайший» к принятому сообщению ξ, т. е. такой, что

$$\varrho(\xi, \eta) = \min. \qquad (58)$$

Для расстояния вида (57) этот принцип широко известен под названием метода наименьших квадратов и впервые предложен великим немецким математиком К. Гауссом.

Рассмотрим теперь конкретный пример подмножества теоретически возможных сообщений. Предположим, что измеряемая величина изменяется во времени по линейному закону:

$$y = kt + b$$

(Для простоты бучем считать, что ошибками в определении моментов времени t_i можно пренебречь.)

Это значит, что каждый вектор $\eta \in H$ имеет вид $\eta = (y_1, y_2, \ldots, y_n)$, где

$$\begin{aligned} y_1 &= kt_1 + b, \\ y_2 &= kt_2 + b, \\ &\ldots\ldots \\ y_n &= kt_n + b. \end{aligned}$$

Вектор $\eta \in H$ определяется двумя параметрами k и b.

Пусть вектор, полученный в результате измерений этой величины, равен $\xi = (x_1, x_2, \ldots, x_n)$. Основное условие (58) записывается теперь так:

$$F(k, b) = (kt_1 + b - x_1)^2 + (kt_2 + b - x_2)^2 + \ldots + \\ + (kt_n + b - x_n)^2 = \min.$$

В выражении $F(k, b)$ неизвестными являются параметры искомого теоретически возможного сообщения k и b; величины $t_1, t_2, \ldots, t_n$ и $x_1, x_2, \ldots, x_n$ известны из опыта.

Для отыскания минимума величины $F(k, b)$ используем критерий, известный из дифференциального исчисления:

$$\frac{\partial F}{\partial k} = 0; \qquad \frac{\partial F}{\partial b} = 0, \tag{59}$$

который в данном случае положительной квадратичной функции $F(k, b)$ является необходимым и достаточным для минимума.

Вычислим частные производные:

$$\left.\begin{aligned} \frac{\partial F}{\partial k} &= 2t_1(kt_1 + b - x_1) + \ldots + 2t_n(kt_n + b - x_n), \\ \frac{\partial F}{\partial b} &= 2\ \ (kt_1 + b - x_1) + \ldots + 2\ \ (kt_n + b - x_n). \end{aligned}\right\} \tag{60}$$

Обозначим для удобства

$$[t^2] = t_1^2 + t_2^2 + \ldots + t_n^2,$$
$$[t] = t_1 + t_2 + \ldots + t_n,$$
$$[tx] = t_1x_1 + t_2x_2 + \ldots + t_nx_n,$$
$$[x] = x_1 + x_2 + \ldots + x_n,$$
$$[1] = 1 + 1 + \ldots + 1 = n.$$

Выражения (60) можно записать тогда в виде

$$\left.\begin{aligned} \frac{\partial F}{\partial k} &= 2[t^2]k + 2[t]b - 2[tx], \\ \frac{\partial F}{\partial b} &= 2[t]k + 2[1]b - 2[x]. \end{aligned}\right\} \qquad (61)$$

Приравнивая согласно (59) эти выражения нулю, сокращая на два и перенося свободные члены в правую часть, получаем основные уравнения метода наименьших квадратов в симзолической форме:

$$\left.\begin{aligned} [t^2]k + [t]b &= [tx], \\ [t]k + [1]b &= [x]. \end{aligned}\right\} \qquad (62)$$

На рис. 29 изображены измерения значения x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8 и определенная по методу наименьших квадратов прямая $y = kt + b$.

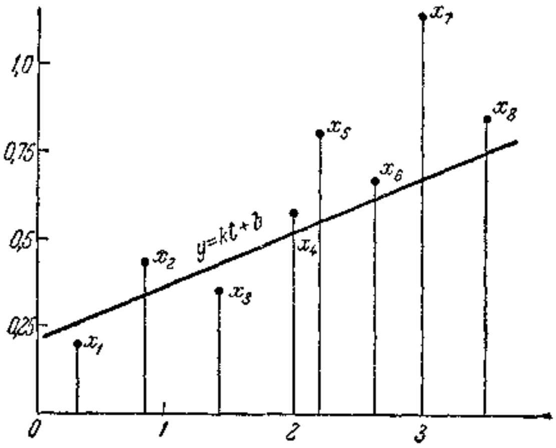


Рис. 29.

Из этого рисунка понятен смысл термина «сглаживание» ошибок. Следующая таблица показывает рациональный порядок выполнения вычислений.

№ п. п.	x_i	t_i	$t_i x_i$	t_i^2
1	0,20	0,30	0,06	0,09
2	0,43	0,91	0,39	0,83
3	0,35	1,50	0,53	2,25
4	0,52	2,00	1,04	4,00
5	0,81	2,20	1,78	4,84
6	0,68	2,62	1,79	6,86
7	1,15	3,00	3,45	9,00
8	0,85	3,30	2,81	10,89
$\sum$	$[x]=4,79$	$[t]=15,83$	$[tx]=11,85$	$[t^2]=38,76$

Система (62) имеет в этом случае вид

$$\left.\begin{aligned} 38,76k+15,83b&=11,85,\\ 15,83k+8b&=4,79.\end{aligned}\right\}$$

Решением являются $k=0,319$, $b=-0,032$. Искомое теоретическое сообщение $y=0,319t-0,032$.

Аналогично, если множество теоретически возможных сообщений H состоит из параболических зависимостей вида $y=at^2+bt+c$, то основное условие (53) запишется в виде

$$F(a,\ b,\ c)=(at_1^2+bt_1+c-x_1)^2+\ldots$$
$$\ldots+(at_n^2+bt_n+c-x_n)^2=\min.$$

Минимизация функции $F(a,\ b,\ c)$ сводится к решению системы трех линейных уравнений с тремя неизвестными параметрами *a, b* и *c*.

Основные приемы метода наименьших квадратов легко переносятся на случай расстояния вида

$$\varrho(\xi,\eta)=\alpha_1(x_1-y_1)^2+\alpha_2(x_2-y_2)^2+\ldots+\alpha_n(x_n-y_n)^2, \quad (63)$$

где $\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_n$—положительные числа (веса). Неравные веса приходится использовать, если отдельные измерения в эксперименте не равноточны и более грубым измерениям приходится приписывать наименьший вес.

Основной принцип сглаживания ошибок (58) применяется также для расстояний в пространствах $C^{(n)}$ и $l_1^{(n)}$. Однако при этом методы восстановления теоретически возможного сообщения оказались более сложными.

2.9. Обобщения понятия расстояния

Как мы уже говорили, возможны различные обобщения понятия расстояния. Одно из наиболее радикальных обобщений используется в теории относительности. В этой теории рассматривается пространственно-временной мир, состоящий из точек вида (x, y, z, t), где x, y и z — пространственные координаты, а t — момент времени. Расстояние (пространственно-временной интервал) между двумя такими точками определяется по формуле

$$\varrho(\xi, \eta) = \sqrt{c^2(t-t_1)^2-(x-x_1)^2-(y-y_1)^2-(z-z_1)^2}, \quad (64)$$

где c — скорость света. Ясно, что расстояние (64) может принимать и мнимые значения.

Можно было бы обобщить понятие расстояния, считая, что функция $\varrho(M, N)$, удовлетворяя всем свойствам расстояния 1, 2, 3 и 4, может принимать бесконечные значения (точнее, значение $+\infty$.). Однако в этом случае множество точек пространства может быть разбито на отдельные подмножества без общих точек, каждое из которых является уже метрическим пространством в обычном смысле. Следовательно, такое обобщение не слишком интересно. Идея доказательства этого факта состоит в следующем.

Рассмотрим такое «обобщенное» пространство E. Выберем некоторую точку $N_1 \in E$ и рассмотрим множество E_1, состоящее из всех точек M пространства E, для которых расстояние $\varrho(N_1, M)$ конечно.

Пусть M и M'—произвольные точки множества E_1, тогда по аксиоме треугольника

$$\varrho(M, M') \leqslant \varrho(M, N_1) + \varrho(N_1, M').$$

Следовательно, $\varrho(M, M')$ конечно, а, значит, множество E_1 образует метрическое пространство с конечным расстоянием. Выкинем из E множество E_1. Если E совпадает с E_1, то само E есть обычное метрическое пространство. Если же нет, то среди оставшихся элементов выберем наугад элемент N_2 и образуем множество E_2 из элементов M, для которых расстояние $\varrho(N_2, M)$ конечно.

Продолжая этот процесс, можно как раз и получить разбиение обобщенного пространства E на обычные метрические пространства с конечным расстоянием. Этот пример показывает, что получить содержательное обобщение какого-либо математического понятия не так-то просто. Во всяком случае, такое **обобщение должно идти от содержательного изучения объектов, а не от чисто формального изменения аксиом.** В частности, существует ряд вполне содержательных обобщений метрического пространства. Одно из таких обобщений мы и рассматриваем далее. Именно, понятие расстояния можно обобщить, отказавшись от условия симметрии (условия 1). Получающийся при этом класс пространств оказывается связанным с некоторыми существенными для математики объектами и поэтому заслуживает рассмотрения.

Мы будем называть *обобщенным метрическим пространством* множество *E,* для любых двух элементов которого ξ и η определено вещественное число $\varrho(\xi, \eta)$ со следующими свойствами:

1. $\varrho(\xi, \eta) \geqslant 0$.

2. Двойное равенство $\varrho(\xi, \eta) = \varrho(\eta, \xi) = 0$ выполняется тогда и только тогда, когда элементы ξ и η совпадают:

$$\xi = \eta.$$

3. Для любой тройки элементов ξ, η, ζ

$$\varrho(\xi, \eta) \leqslant \varrho(\xi, \zeta) + \varrho(\zeta, \eta).$$

Ясно, что обычное расстояние удовлетворяет этим условиям. Однако этим свойствам 1, 2,3 может удовлетворять и функция $\varrho(\xi, \eta)$, несимметричная относительно своих аргументов. Пример с несимметричным расстоянием у нас встретился фактически в конце п.2.4 в связи с определением расстояния как минимального времени, затрачиваемого на переход из одной точки в другую. В связи с тем, что путь в противоположные концы может занимать разное время, это расстояние, вообще говоря, несимметрично, хотя аксиома треугольника проверяется легко.

Другой пример пространства с несимметричным расстоянием получается, если рассмотреть множество из 10 точек, являющихся вершинами ломаной, изображенной на рис. 30.

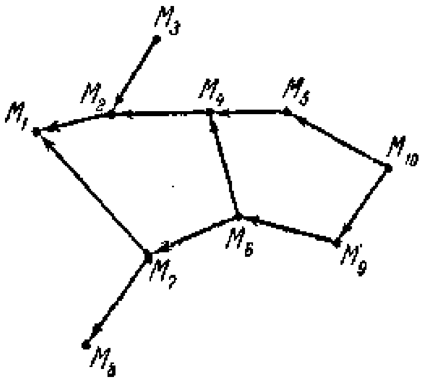


Рис. 30.

Расстояние между точками M_i и M_j, $\varrho(M_i, M_j)$ определяется как минимальное число отрезков, проходимых против стрелки в пути, соединяющем M_i и M_j.

Например,

$$\varrho(M_1, M_{10}) = 4; \ \varrho(M_{10}, M_1) = 0;$$
$$\varrho(M_8, M_9) = 3; \ \varrho(M_9, M_8) = 1 \text{ и т. д.}$$

Ясно, что $\varrho(M_i, M_j) \geqslant 0$. Условие $\varrho(M_i, M_j) = \varrho(M_j, M_i) = 0$ (равенство нулю обоих расстояний) означает, что и точку M_i с M_j, и точку M_j с M_i можно соединить отрезками, направленными по стрелке, т. е. M_i и M_j принадлежат замкнутой ломаной с согласованными направлениями отрезков. Так как таких петель на рис. 30 нет, то равенство $\varrho(M_i, M_j) = \varrho(M_j, M_i) = 0$ влечет за собой совпадение точек M_i и M_j.

Таким образом, мы проверили выполнение условий 1 и 2 для несимметричного расстояния. Аксиома треугольника проверяется следующим рассуждением. Рассмотрим путь с наименьшим количеством отрезков, направленных против стрелки, идущий из M_i в M_k, и аналогичный путь из M_k в M_j. Объединяя эти пути, мы получим путь из M_i в M_j с количеством отрезков, направленных против стрелки, равным $\varrho(M_i, M_k) + \varrho(M_k, M_j)$. Так как в «кратчайшем» пути из M_i в M_j количество таких отрезков может быть только меньше, то

$$\varrho(M_i, M_j) \leqslant \varrho(M_i, M_k) + \varrho(M_k, M_j). \qquad (65)$$

В этом примере можно было бы определить новое расстояние по правилу

$$\varrho^*(M_i, M_j) = \varrho(M_i, M_j) + \varrho(M_j, M_i). \qquad (66)$$

Это расстояние равно просто минимальному количеству отрезков в пути, соединяющем точки M_i и M_j. Отсюда ясно, что $\varrho^*(M_i, M_j)$ обладает свойствами обычного расстояния.

Аналогичный факт справедлив для любого обобщенного метрического пространства.

Теорема. *Если в обобщенном метрическом пространстве с расстоянием* $\varrho(\xi, \eta)$ *ввести новое расстояние*

$$\varrho^*(\xi, \eta) = \varrho(\xi, \eta) + \varrho(\eta, \xi), \qquad (67)$$

то оно обладает всеми свойствами обычного расстояния и пространств0 превращается в обычное метрическое пространство.

Доказательство. Симметричность расстояния $\varrho^*(\xi, \eta)$ следует из того, что правая часть (67) не меняется при перестановке ξ и η. Так как $\varrho(\xi, \eta) \geqslant 0$ и $\varrho(\eta, \xi) \geqslant 0$, то $\varrho^*(\xi, \eta) \geqslant 0$. Равенство

$\varrho^*(\xi, \eta)=0$ равносильно двойному равенству $\varrho(\xi, \eta)=\varrho(\eta, \xi)=0$, т. е. равносильно совпадению точек ξ и η. Наконец, складывая два неравенства

$$\varrho(\xi, \eta)\leqslant\varrho(\xi, \zeta)+\varrho(\zeta, \eta)$$

и

$$\varrho(\eta, \xi)\leqslant\varrho(\eta, \zeta)+\varrho(\zeta, \xi),$$

получаем

$$\varrho(\xi, \eta)+\varrho(\eta, \xi)\leqslant\varrho(\xi, \zeta)+\varrho(\zeta, \xi)+\varrho(\zeta, \eta)+\varrho(\eta, \zeta)$$

или

$$\varrho^*(\xi, \eta)\leqslant\varrho^*(\xi, \zeta)+\varrho^*(\zeta, \eta),$$

т. е. для нового расстояния выполнено неравенство треугольника. Пример обобщенного метрического пространства с несимметричным расстоянием получается с помощью важного понятия частично упорядоченного множества.

Множество называется *частично упорядоченным,* если для некоторых пар точек *M* и *N* этого множества имеет смысл отношение $M \prec N$ (читается: *M* предшествует *N)* и при этом выполняются свойства:

1. Если $M \prec N$, то невозможно отношение $N \prec M$.
2. Если $M \prec N$, а $N \prec L$, то $M \prec L$.

Примером частично упорядоченного множества является множество точек на рис. 30. Для этого нужно положить $M_j \prec M_i$, если из точки M_i можно достичь точку M_j, двигаясьтолько по отрезкам вдоль стрелки.

Например, $M_8 \prec M_{10}, M_1 \prec M_3,\ M_2 \prec M_6$.

Второй пример получится, если взять множество точек на числовой прямой и положить $x \prec y$, если число $x < y$.

В этом случае для каждой пары различных точек *x* и *y* справедливо либо отношение $x \prec y$, либо $y \prec x$ (такие множества называются упорядоченными).

Во всяком частично упорядоченном множестве *E* можно ввести понятие *непосредственного предшествования.*

Точка *M* непосредственно предшествует точке *N* (пишется: $M \bigcirc N$), если $M \prec N$ и не существует третьей, отличной от *M* и *N* точки *L,* лежащей между ними: $M \prec L \prec N$.

Так, в примере на рис. 30 $M_2 \bigcirc M_3,\ M_2 \bigcirc M_4,\ M_1 \bigcirc M_7$ и т. д.

Во множестве вещественных чисел непосредственно предшествующих точек нет вообще. Действительно, если $x \prec y$, то всегда $x \prec \frac{x+y}{2} \prec y$, так как $x < \frac{x+y}{2} < y$.

Теперь рассмотрим конечное частично упорядоченное множество *E*. Предположим, что множество *E* обладает свойством *связности.* Это значит, что для любых точек *M* и *N* из *E* существует последовательность точек $L_1 = M,\ L_2, \ldots\ldots, L_k = N$, так что для любой пары $L_i,\ L_{i+1}$ либо $L_i \Subset L_{i+1}$, либо $L_{i+1} \Subset L_i$. Так, для точек M_3 и M_8 на рис. 30 эту последовательность можно составить так:

$$L_1 = M_3; \qquad L_2 = M_2; \qquad L_3 = M_7; \qquad L_4 = M_8,$$

так как

$$M_2 \Subset M_3; \qquad M_1 \Subset M_2; \qquad M_8 \Subset M_7.$$

Можно проверить, что это множество является связным. Множество точек на рис. 31 не является связным, так как для точек M_8 и M_3 такой последовательности нельзя найти.

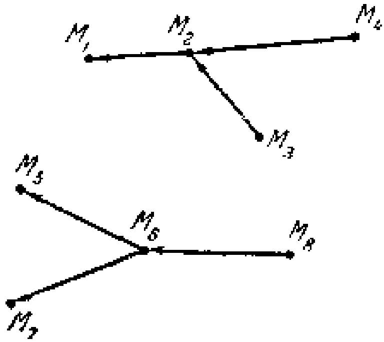


Рис. 31

Однако оба подмножества $E_1\{M_1,\ M_2,\ M_3,\ M_4\}$ и $E_2\{M_5,\ M_6,\ M_7,\ M_8\}$ являются связными. Можно легко проверить, что всякое конечное частично упорядоченное множество распадается на связные подмножества без общих точек.

Введем теперь в конечное связное частично упорядоченное множество *E* расстояние $\varrho(M,\ N)$, определенное по следующему правилу. Назовем путем из *M* в *N* цепочку из точек $L_1 = M; L_2,$ $L_3,\ \ldots,\ L_k = N$ таких, что всегда или $L_i \bigcirc L_{i+1}$, или $L_{i+1} \bigcirc L_i$. Длиной пути мы назовем коли чество точек L_i в нем, для которрых $L_{i+1} \bigcirc L_i$. Тогда расстоянием $\varrho(M,\ N)$ мы назовем минимальную длину пути из *M* в *N*.

Проверим, что так определенное расстояние обладает всеми необходимыми свойствами. Неотрицательность расстояния $\varrho(M,\ N)$ следует из определения. Для доказательства второго свойства расстояния заметим, что при несовпадающих *M* и *N* из условия

$\varrho(M, N)=0$ следует, что $M \prec N$. Действительно, условие $\varrho(M, N)=0$ означает, что существует цепочка точек $L_1=M, L_2, .., L_k= N$ такая, что всегда $L_i \bigcirc L_{i+1}$, а следовательно, и $L_i \prec L_{i+1}$. Но тогда по второму свойству частичной упорядоченности получается, что $M \prec N$. Аналогично, из $\varrho(M, N)=0$ вытекает, что $N \prec M$. Значит, если M и N — разные точки и $\varrho(M, N)=\varrho(N, M)=0$, то одновременно $M \prec N$ и $N \prec M$. Так как последнее невозможно по первому свойству частично упорядоченных множеств, то M совпадает с N. Итак, второе свойство расстояния доказано.

Для доказательства третьего свойства (аксиомы треугольника) используем надежно послуживший нам ранее прием. Возьмем «кратчайший» путь из M в Q: $M = L_1, L_2$ .... .. , $L_k = Q$ и «кратчайший» путь из Q в N: $Q = L_1', L_2'$, ... , $L'_p= N$. Объединяя их, мы получим цепочку точек

$$M=L_1,\ L_2,\ \ldots\ ,\ L_k=Q=L_1',\ L_2',\ \ldots\ ,\ L_p'=N,$$

образующую путь из M в N. Общее число смежных точек этой цепочки, для которых $L_{i+1} \bigcirc L_i$ $(L_{i+1}' \bigcirc L_i')$, равно сумме расстояний $\varrho(M, Q)+\varrho(Q, N)$. Ясно, что число таких пар для «кратчайшего» пути из M в N может быть только меньше этого значения:

$$\varrho(M, N) \leqslant \varrho(M, Q)+\varrho(Q, N). \qquad (68)$$

Итак, доказано, что **всякое конечное связное частично упорядоченное множество *E* является обобщенным метрическим пространством, если определить расстояние по указанному рецепту**.

В качестве упражнения предоставляем читателю убедиться, что для всякой пары точек M и N таких, что $M \prec N$, расстояние $\varrho(M, N)=0$.

Оказывается, что верно и в некотором смысле обратное утверждение. Именно в каждом обобщенном метрическом пространстве E можно ввести частичную упорядоченность, определивправило предшествования так: $M \prec N$, если

$$\varrho(M, N)=0.$$

Покажем, что при этом выполнены оба свойства частичной упорядоченности. Пусть M и N — разные точки и $M \prec N$, т. е. $\varrho(M, N)=0$. Тогда, если бы было одновременно $M \prec N$, то было бы и $\varrho(N, M)=0$. Отсюда по второму свойству расстояния вытекало бы, что M и N совпадают. Так как последнее неверно по предположению, то, значит, из $M \prec N$ следует невозможность отношения $N \preccurlyeq M$. Первое свойство частичной упорядоченности доказано.

Пусть теперь $M \preccurlyeq L$ и $L \preccurlyeq N$. Это значит, что $\varrho(M, L) = 0$ и $\varrho(L, N) = 0$. Тогда, по аксиоме треугольника,

$$\varrho(M, N) \leqslant \varrho(M, L) + \varrho(L, N) = 0,$$

т. е. $\varrho(M, N) = 0$ и $M \preccurlyeq N$.

Мы доказали, таким образом, что из $M \preccurlyeq L$ и $L \preccurlyeq N$ следует $M \preccurlyeq N$.

Полученное частично упорядоченное множество, вообще говоря, не обязано быть связным.

Рассмотрим, например, частично упорядоченное множество *Е* точек на рис. 31. Между точками подмножества $E_1\{M_1, M_2, M_3, M_4\}$ можно определить расстояние с помощью «наикратчайших» путей. То же самое можно сделать и для точек подмножества $E_2\{M_5, M_6, M_7, M_8\}$.

Определим теперь расстояние между любой точкой $M_i \in E_1$ и любой точкой $M_j \in E_2$, просто положив

$$\varrho(M_i, M_j) = \varrho(M_j, M_i) = 100. \tag{69}$$

Нетрудно проверить, что мы получим обобщенное метрическое пространство, в котором равенство $\varrho(M, N) = 0$ равносильно отношению $M \preccurlyeq N$ в частично упорядоченном множестве *Е*. Однако *Е* не является, как уже говорилось, связным частично упорядоченным множеством.

Можно ввести понятие *связного обобщенного метрического пространства.* Именно, пространство *Е* называется *связным,* если для любой пары несовпадающих точек *М* и *N* из *Е* существует цепочка точек $L_1 = M, L_2, \ldots, L_k = N$ такая, что для любой смежной пары точек L_i и L_{i+1}, либо $\varrho(L_i, L_{i+1}) = 0$, либо $\varrho(L_{i+1}, L_i) = 0$. Читатель легко проверит, что связным метрическим пространствам соответствуют связные частично упорядоченные множества.

Конечное частично упорядоченное множество (а значит, и соответствующее метрическое пространство) можно довольно просто изобразить геометрически. Для этого будем изображать элементы частично упорядоченного множества точками в трехмерном пространстве, обозначая их теми же буквами, что и соответствующие элементы. Каждую пару точек *М* и *N*, для которых $M \bigcirc N$, соединим линией, направленной из *N* в *М,* и укажем это направление стрелкой. Полученная геометрическая фигура, состоящая из точек (вершин этой фигуры) и соединяющих их направленных линий, называется *графом.* Примеры графов мы уже видели на рис. 30 и 31.

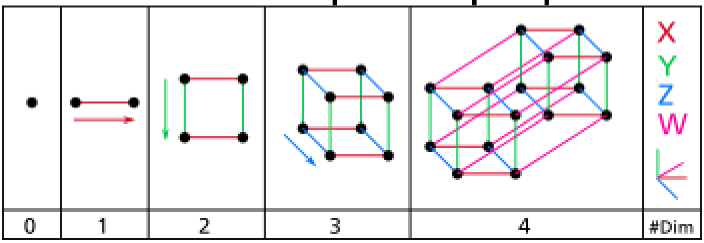
Нетрудно видеть, что если $M \preccurlyeq N$, то из *N* в *М* можно попасть, двигаясь только по направлению стрелки.

Заметим в заключение, что пространства с несимметричным расстоянием связаны также с важным для математики понятием **дискретных топологических пространств**.

На этом мы заканчиваем изучение понятия расстояния. Мы установили, что это понятие имеет много различных аспектов и связано не только с чисто математическими задачами, но и с такими вполне прикладными проблемами, как построение помехоустойчивых кодов. Эта особенность — многозначность применений и сложные логические связи — характерна и для других существенных математических понятий. Основной смысл создания таких понятий как раз и состоит в возможности установления связей и аналогий между, казалось бы, очень далекими областями и в выяснении скрытых пружин, от которых зависят математические свойства различных объектов.

3. Размерность пространств и их метричность

3.1. Размерность пространств



3.1.1.Определения

- Теория размерности — часть топологии, в которой изучаются размерности — числовые топологические инварианты определённого типа.
- Размерность пространства — количество независимых параметров, необходимых для описания состояния объекта

В математике существует несколько различных подходов к определению размерности, например

- Размерность векторного пространства
- Комбинаторная размерность множества определяется на основании его комбинаторных свойств и может быть произвольным неотрицательным числом[1].
  - Размерность Лебега, или топологическая размерность.
  - Хаусдорфова размерность метрического пространства.
  - Размерность Минковского допускает обобщение на фракталы, при этом их размерность может быть произвольным неотрицательным числом.

Ве́кторное (или **лине́йное**) **простра́нство** — математическая структура, которая представляет собой набор элементов, называемых векторами, для которых определены операции сложения друг с другом и умножения на число — скаляр. Эти операции подчинены восьми аксиомам. Скаляры могут быть элементами вещественного, комплексного или любого другого поля чисел. Частным случаем подобного пространства является обычное трехмерное евклидово пространство, векторы которого используются, к примеру, для представления физических сил. При этом следует отметить, что вектор как элемент векторного пространства не обязательно должен быть задан в виде направленного отрезка. Обобщение понятия «вектор» до элемента векторного пространства любой природы не только не вызывает смешения терминов, но и позволяет уяснить или даже предвидеть ряд результатов, справедливых для пространств произвольной природы.

Векторные пространства являются предметом изучения линейной алгебры. Одна из главных характеристик векторного пространства — его размерность. Размерность представляет собой максимальное число линейно независимых элементов пространства, то есть, прибегая к грубому геометрическому описанию, число направлений, невыразимых друг через друга посредством только операций сложения и умножения на скаляр. Векторное пространство можно наделить дополнительными структурами, например, нормой или скалярным произведением. Подобные пространства естественным образом появляются в математическом анализе, преимущественно в виде бесконечномерных функциональных пространств (*англ.*), где в качестве векторов выступают функции. Многие проблемы анализа требуют выяснить, сходится ли последовательность векторов к данному вектору. Рассмотрение таких вопросов возможно в векторных пространствах с дополнительной структурой, в большинстве случаев — подходящей топологией, что позволяет определить понятия близости и непрерывности. Такие топологические векторные пространства, в частности, банаховы и гильбертовы, допускают более глубокое изучение.

Кроме векторов, линейная алгебра изучает также тензоры более высокого ранга (скаляр считается тензором ранга 0, вектор — тензором ранга 1).

Первые труды, предвосхитившие введение понятия векторного пространства, относятся к XVII веку. Именно тогда своё развитие получили аналитическая геометрия, учения о матрицах, системах линейных уравнений, евклидовых векторах.

Размерность Лебега или **топологическая размерность** — размерность, определённая посредством покрытий, важнейший инвариант топологического пространства.

Размерность Хаусдорфа, **хаусдорфова размерность** — естественный способ определить размерность подмножества в метрическом пространстве. Размерность Хаусдорфа согласуется с нашими обычными представлениями о размерности в тех случаях, когда эти обычные представления есть. Например, в трёхмерном евклидовом пространстве хаусдорфова размерность конечного множества равна нулю, размерность гладкой кривой — единице, размерность гладкой поверхности — двум и размерность множества ненулевого объёма — трём. Для более сложных (фрактальных) множеств размерность Хаусдорфа может не быть целым числом.

Размерность Минковского или **грубая размерность** ограниченного множества в метрическом пространстве равна

$$\lim_{\varepsilon \to 0} \frac{\ln(N_\varepsilon)}{\ln(\varepsilon)},$$

где N_ε — минимальное число множеств диаметра ε , которыми можно покрыть наше множество. Если предел не существует, то можно рассматривать верхний и нижний предел и говорить соответственно о верхней и нижней размерности Минковского.

Близким к размерности Минковского понятием является размерность Хаусдорфа. Во многих случаях эти размерности совпадают, хотя существуют множества, для которых они различны.

Фракта́льная разме́рность (англ. *fractal dimension*) — один из способов определения размерности множества в метрическом пространстве. Фрактальную размерность *n*-мерного множества можно определить с помощью формулы:

$D = -\lim\limits_{\varepsilon \to 0} \dfrac{\ln(N_\varepsilon)}{\ln(\varepsilon)}$, где N_ε — минимальное число *n*-мерных «шаров» радиуса ε, необходимых для покрытия множества.

Фрактальная размерность может принимать не целое числовое значение.

Основная идея «дробной» (англ. *fractured*) размерности имеет долгую историю в области математики, но именно сам термин введён в оборот Бенуа Мандельбротом в 1967 году в его статье о самоподобии, в которой он описал «дробную» (англ. *fractional*) размерность. В этой статье Мандельброт ссылался на предыдущую работу Льюиса Фрайя Ричардсона, описывающую противоречащую здравому смыслу идею о том, что измеренная длина береговой линии зависит от длины мерной палки (шеста) (см. рис. 1).

11.5 х 200 = 2300 км

28 x 100 = 2800 км

70 x 50 = 3500 км

Рис. 1. Общая длина береговой линии Великобритании возрастает, когда длина измерительной палки (шеста) уменьшается.

Следуя этому представлению, фрактальная размерность береговой линии соответствует отношению числа шестов (в определенном масштабе), нужных для измерения длины береговой линии, к выбранному масштабу шеста. Есть несколько формальных математических определений фрактальной размерности, которые строятся на этой базовой концепции, о изменении в элементе с изменением в масштабе. Одним из элементарных примеров является фрактальная размерность снежинки Коха. Её топологическая размерность равна 1, но это ни в коем случае не спрямляемая кривая, поскольку длина кривой между любыми двумя точками снежинки Коха — бесконечность. Никакая сколько угодно малая часть кривой не является отрезком прямой. Скорее, снежинка Коха состоит из бесконечного числа сегментов, соединённых под разными углами. Фрактальную размерность кривой можно объяснить интуитивно, предполагая, что **фрактальная линия — это объект слишком**

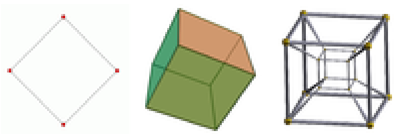
детальный (подробный), чтобы быть одномерным, но недостаточно сложный, чтобы быть двумерным. Поэтому её размерность лучше описывать не обычной топологической размерностью 1, но её фрактальной размерностью, равной в этом случае числу, лежащему в интервале между 1 и 2.

Размерность в физике — количество независимых параметров, необходимых для описания состояния объекта, или количество степеней свободы физической системы, положенными в основу системы единиц; записывается в виде произведения символов соответствующих основных величин, возведенных в определенные степени, которые называются показателями размерности. Размерность является физической величиной, показывающей связь данной величины с физическими величинами, положенными в основу системы единиц; записывается в виде произведения символов соответствующих основных величин, возведенных в определенные степени, которые называются показателями размерности.

3.1.2. Пространственные измерения

Классические физические теории описывают трёхмерные физические измерения.

Примеры



Квадрат->Куб->Тессеракт

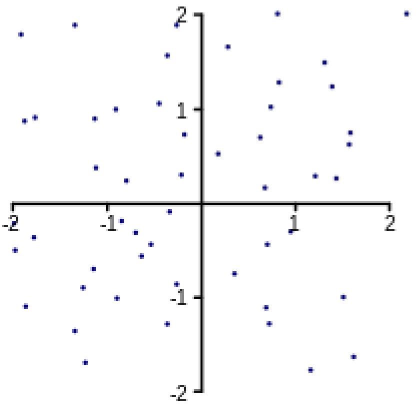
- Для того, чтобы описать положение окружности на плоскости, достаточно трёх параметров: двух координат центра и радиуса, то есть: *пространство окружностей на плоскости* — трёхмерно; *пространство точек на той же поверхности* — двумерно; тем не менее сама окружность — *пространство точек на окружности* — одномерна: любая её точка может быть описана одним параметром.
- В рамках ходовых моделей поверхности нашей планеты для определения положения города (город при этом

рассматривается не как двумерный объект, а как точка) на поверхности Земли достаточно двух параметров, а именно: географической широты и географическая долготы. Соответственно: пространство в таких моделях является двумерным (сокращённо — 2D, от англ. *dimension*).

- В рамках ходовых моделей нашей физической реальности для определения положения некоего объекта, к примеру — самолёта (самолёт при этом рассматривается не как трёхмерный объект, а — как точка), требуется указать три координаты — дополнительно к широте и долготе нужно знать высоту, на которой он находится. Соответственно: пространство в таких моделях является трёхмерным (3D). К этим трём координатам может быть добавлена четвёртая (время) для описания не только текущего положения самолёта, но и момента времени. Если добавить в модель ориентацию (крен, тангаж, рыскание) самолёта, то добавятся ещё три координаты и соответствующее абстрактное пространство модели станет семимерным.

Точка (нульмерное пространство)

У этого термина существуют и другие значения.



Набор точек на плоскости

В геометрии, топологии и близких разделах математики **то́чкой** называют абстрактный объект в пространстве, не имеющий ни объёма, ни площади, ни длины, ни каких-либо других измеримых характеристик. Таким образом, точкой называют нульмерный объект. Точка является одним из фундаментальных понятий в математике; любая геометрическая фигура считается состоящей из точек.

Точка в Евклидовой геометрии

Евклид определил точку как объект, не имеющий измерений. В современной аксиоматике геометрии точка является первичным понятием, задаваемым лишь перечнем его свойств — аксиомами.

3.1.3.Одномерное пространство

Одномерное пространство — геометрическая модель материального мира, в которой положение точки возможно охарактеризовать всего одним числом. Также одномерным пространством считается *n-мерное пространство*, где **n=1**.

Геометрия одномерного пространства

Единственным политопом, существующим в одномерном пространстве, является отрезок.

(**Политоп** — это подмножество Евклидова пространства, которое представимо в виде объединения конечного числа симплексов.

Свойства

- Любой политоп допускает триангуляцию; то есть, может быть представлен как объединение конечного множества симплексов $\mathcal{S}$ таких что
    - с любым из симплексов из $\mathcal{S}$, в $\mathcal{S}$ входят все его грани;
    - любые два симплекса либо вообще не имеют общей точки, либо они пересекаются только по целой грани какой-то размерности.

- Пересечение и объединение конечного числа политопов является политопом.

Вариации и обобщения

- Топологческий политоп — топологического пространства гомеоморфное некоторому политопу.)

Гиперсфера в одномерном пространстве — это пара точек, расположенных на расстоянии друг от друга, равном

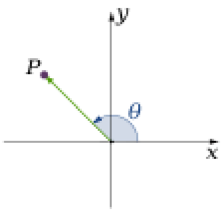
$$L = 2r,$$

где r— радиус сферы.

Примером системы координат в одномерном пространстве является числовая прямая, на которой располагаются точки и отрезки, имеющие только одну пространственную характеристику — протяжённость или длину. Одномерным пространством можно считать также угол. Обычную линию, на которой поставлена точка с координатой 0 как точка отсчёта, нельзя считать одномерным пространством, хотя простую линию без каких-либо точек можно считать таковым

- 

  Числовая прямая

- 

  Угол

3.1.4. Двумерное пространство

Двумéрное прострáнство (иногда говорят **двухмéрное пространство**) — геометрическая модель плоской проекции физического мира, в котором мы живём. Двумерным пространством считается *n-мерное пространство*, где **n=2**.

Примером двумерного пространства является плоскость. Точки данного пространства возможно задать всего двумя числами. Например, любую точку можно задать парой чисел: (*x*, *y*). Плоские объекты характеризуются не только длиной, но и шириной.

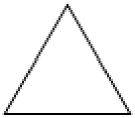
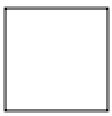
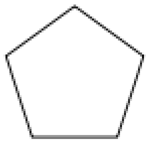
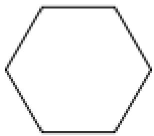
Геометрия двумерного пространства

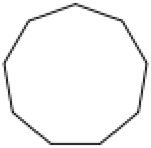
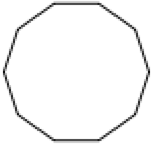
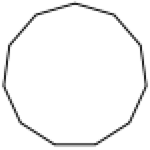
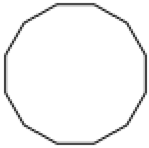
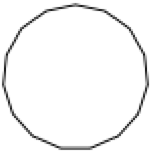
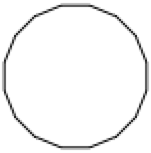
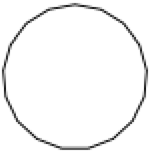
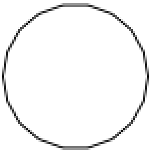
Многогранники

В двумерном пространстве существует бесконечно много правильных многогранников: правильные многоугольники. Примеры последних приведены ниже:

Выпуклые

Символ {p} (символ Шлефли) обозначает правильный *p*-угольник.

Название	**Треугольник (2-симплекс)**	**Квадрат (2-куб)**	**Пятиугольник**	**Шестиугольник**	**Сем**
Символ Шлефли	{3}	{4}	{5}	{6}	{7}
Вид					
Название	**Девятиугольник**	**Десятиугольник**	**11-угольник**	**12-угольник**	**13-у**
Символ	{9}	{10}	{11}	{12}	{13

Шлефли					
Вид					
Название	**15-угольник**	**16-угольник**	**17-угольник**	**18-угольник**	**19-у**
Символ Шлефли	{15}	{16}	{17}	{18}	{19
Вид					

Гиперсфера

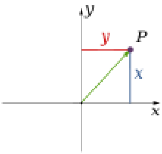
Гиперсферой в двумерном пространстве является окружность, которую иногда называют **1-сфера**, потому что её поверхность является одномерной. Площадь части плоскости, заключённой внутри гиперсферы (площадь круга) равна:

$$A = \pi r^2,$$

где r — радиус окружности.

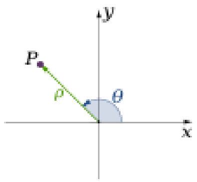
Системы координат в двумерном пространстве

Наиболее распространённые координатные системы — прямоугольная (Декартова) система координат, полярная система координат и географическая координатная система.



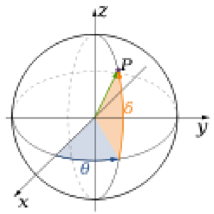
•

Прямоугольная система координат



•

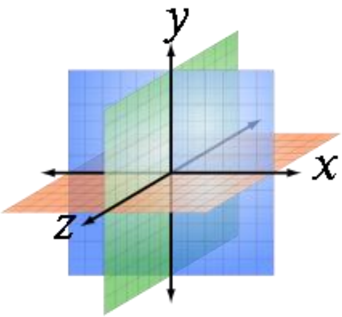
Полярная система координат



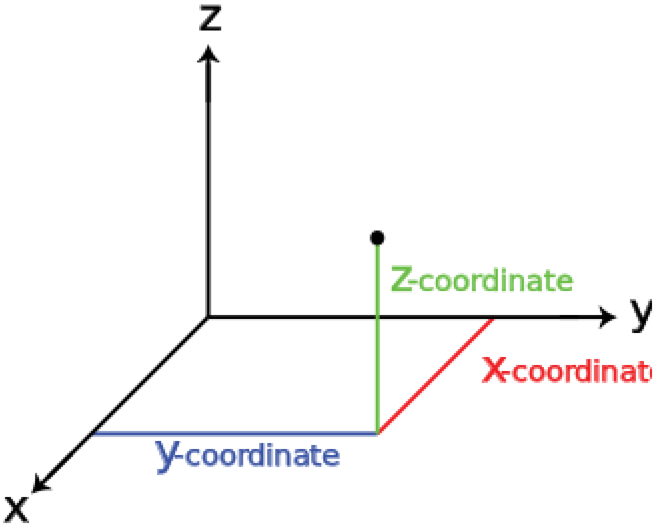
•

Географические координаты

3.1.5. Трёхмерное пространство



Трёхмерная метрика пространства



Трёхмерная система координат с осью X направленной к читателю.

Трёхме́рное простра́нство — геометрическая модель материального мира, в котором мы находимся. Это пространство называется трёхмерным, так как оно имеет три однородных измерения — высоту, ширину и длину, то есть трёхмерное пространство описывается тремя единичными ортогональными векторами.

Понимание трёхмерного пространства людьми, как считается, развивается ещё в младенчестве, и тесно связано с координацией движений человека. Визуальная способность воспринимать окружающий мир органами чувств в трёх измерениях называется глубиной восприятия.

В аналитической геометрии каждая точка трёхмерного пространства описывается как набор из трёх величин — координат. Задаются три взаимно перпендикулярных координатных оси, пересекающихся в начале координат. Положение точки задаётся относительно этих трёх осей заданием упорядоченной тройки чисел. Каждое из этих чисел задаёт расстояние от начала отсчёта до точки, измеренное вдоль соответствующей оси, что равно расстоянию от точки до плоскости, образованной другими двумя осями.

Также существуют другие системы координат, наиболее часто используются цилиндрическая и сферическая системы.

Другой взгляд даёт линейная алгебра, где важную роль играет понятие линейной независимости. Пространство трёхмерно по той причине, что высота коробки не зависит от её длины и ширины. На языке линейной алгебры пространство трёхмерно, потому что каждая точка может быть задана комбинацией из трёх линейно независимых векторов. В этих терминах пространство-время четырёхмерно, потому что положение точки во времени не зависит от её положения в пространстве.

Трёхмерное пространство имеет несколько свойств, которые отличают его от пространств другой размерности. Например, это пространство наименьшей размерности, в котором можно завязать узел на куске верёвки. Многие законы физики, например многие законы обратных квадратов связаны с тем что размерность нашего пространства три.

Нульмерное, одномерное и двухмерное пространства могут рассматриваться как располагающиеся в трёхмерном пространстве; само оно может считаться частью модели четырёхмерного пространства (четвёртым измерением континуума, как правило, называют время — неоднородное качество по отношению к пространственной мерности).

3.1.6.Четырёхмерное пространство

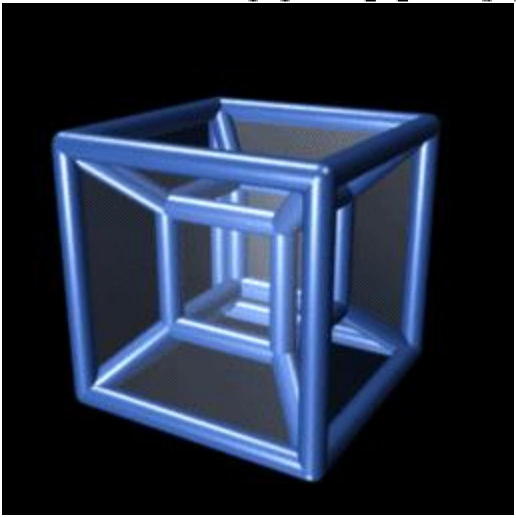
Четвёртое измерение, или **четырёхмерное («4D») пространство** — в математике абстрактное понятие, производимое путём обобщения правил трёхмерного пространства. Оно изучалось математиками и философами на протяжении почти двух столетий как ради простого интереса, так и ради возможностей, которые это понятие открывает в математике и смежных областях.

Алгебраически оно получено путём применения правил векторов и координатной геометрии к пространству с четырьмя измерениями. В частности, вектор с четырьмя компонентами может быть использован для представления позиции в четырёхмерном пространстве. Это Евклидово пространство, поэтому имеет метрику и норму, и таким образом все измерения рассматриваются одинаково: дополнительное измерение неотличимо от трёх других.

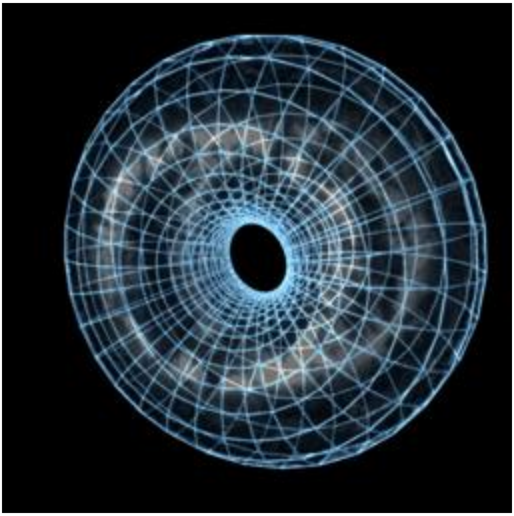
В современной физике пространство и время объединены в единый четырёхмерный континуум, называемый пространством Минковского, метрика которого рассматривает временное измерение иначе, чем пространственные измерения. Таким образом, пространство Минковского является псевдоевклидовым, а не евклидовым.

Аналогично 2- и 3-мерным пространствам, скалярное произведение векторов вычисляется по формуле:

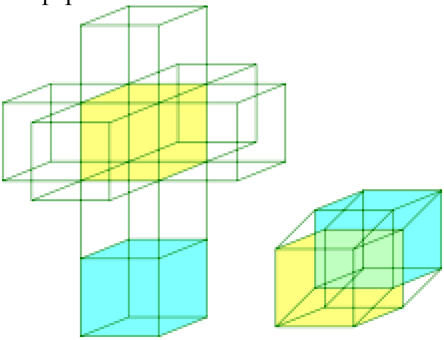
$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3 + a_4 b_4 .$$



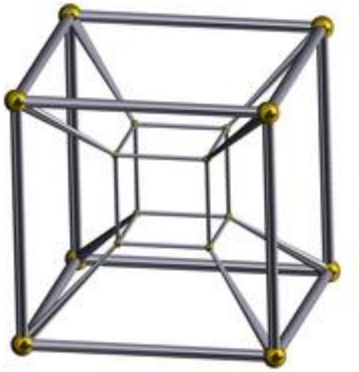
3D проекция тессеракта, простое вращение



Стереографическая проекция тора Клиффорда: множество точек (cos(a), sin(a), cos(b), sin(b)), который является подмножеством 3-сферы.



Развертка тессеракта



3.1.7. Проективное пространство

Проекти́вное простра́нство над телом K — пространство, состоящее из прямых (одномерных подпространств) некоторого линейного пространства $L(K)$ над данным телом. Прямые пространства $L(K)$ называются *точками* проективного пространства.

Если L имеет размерность $n+1$, то размерностью проективного пространства называется число n, а само проективное пространство обозначается KP^n и называется ассоциированным с L (чтобы это указать, принято обозначение $P(L)$).

Переход от векторного пространства $L(K)$ размерности $n+1$ к соответствующему проективному пространству KP^n называется **проективизацией** пространства $L(K)$.

Точки KP^n можно описывать с помощью однородных координат.

Проективное пространство может быть также определено системой аксиом типа гильбертовской, что наиболее интересно в случае проективной плоскости. Тогда оказывается, что проективная плоскость, определённая аксиомами, может быть определена как двумерное проективное пространство над некоторым телом тогда и только тогда, когда выполняется т. н. аксиома Дезарга, которая для размерностей больших 2 является теоремой.

1. Связанные определения

- Двумерное проективное пространство называется проективной плоскостью.
- Пусть M есть гиперплоскость в линейном пространстве L. Проективное пространство $P(M)\subset P(L)$ называется проективной гиперплоскостью $P(L)$.

2. Свойства

- На дополнении проективной гиперплоскости $A=P(L)\backslash P(M)$ существует естественная структура аффинного пространства.
- Обратно, взяв за основу аффинное пространство A можем получить проективное пространство как аффинное, к которому добавлены бесконечно удалённые точки. Первоначально проективное пространство и было введено таким образом.

3. Тавтологическое расслоение

Тавтологическим расслоением $\gamma^n : E\to\mathbb{R}P^n$ называется векторное расслоение, пространством расслоения которого является подмножество прямого произведения $\mathbb{R}P^n\times\mathbb{R}^{n+1}$

$$E(\gamma^n):=\big\{(\{\pm\, x\},v)\in\mathbb{R}P^n\times\mathbb{R}^{n+1}:v=\lambda x,\ \lambda\in\mathbb{R}\big\}.$$

а слоем — вещественная прямая $\mathbb{R}$. Каноническая проекция γ^n отображает прямую, проходящую через точки $\pm x\in\mathbb{R}^{n+1}$, в соответствующую точку проективного пространства. При $n\geq 1$ это расслоение не является тривиальным. При $n=1$ пространством расслоения является лента Мёбиуса.

3.2. Метрические пространства

В основе математического анализа лежит понятие предела числовой последовательности и операция предельного перехода. Достаточно сказать, что производная и определённый интеграл определяются через понятие предела. При определении предела используется тот факт, что на числовой прямой определено расстояние между вещественными числами. Но оказывается, что для формулировки многих фундаментальных понятий и доказательства различных теорем анализа важна не природа действительных чисел, а только само понятие расстояния. Обобщением представления о расстоянии между действительными числами на случай произвольного множества является понятие метрики, которое мы сейчас введём.

3.2.1. Определение метрического пространства

Пусть M — некоторое непустое множество, ρ — некое отображение, ставящие в соответствие двум элементам множества M некоторое вещественное число:

$$\rho\colon M \times M \to \mathbb{R},$$

отображение ρ называется **метрикой**, если оно обладает следующими свойствами (аксиомы метрики):

1. Аксиома тождества: $\rho(x, y) = 0 \Leftrightarrow x = y$,
2. Аксиома симметрии: $\rho(x, y) = \rho(y, x)$,
3. Аксиома неравенства треугольника: $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$.

Совокупность множества M и определённой на нём метрики ρ называют **метрическим пространством** и обозначают (M,ρ). Иногда, особенно когда из контекста понятно о какой метрике идёт речь, метрическое пространство обозначают так же, как и само множество M. Элементы метрического пространства обычно называют **точками**.

Одним из простейших (и важнейших) примеров метрического пространства является числовая прямая. Покажем, что множество вещественных чисел с метрикой $\rho(x, y)=|x - y|$ является метрическим пространством. Действительно, рассмотрим три произвольных вещественных числа

$$x, y, z \in \mathbb{R},$$

Все аксиомы метрического пространства выполняются, по свойствам модуля:

$$|x - y| = 0 \Leftrightarrow x = y,$$
$$|x - y| = |-(y - x)| = |-1| \cdot |y - x| = |y - x|,$$
$$|x - y| = |x - z + z - y| \leq |x - z| + |z - y|$$

Пусть (M, ρ) — метрическое пространство, и A — непустое подмножество множества M, тогда (A, ρ) — тоже является метрическим пространством, которое называется **подпространством** метрического пространства (M,ρ).

Например, множество рациональных чисел является подмножеством действительных чисел:

$$\mathbb{Q} \subset \mathbb{R},$$

а следовательно, если взять естественную для вещественных чисел метрику

$$\rho(x, y) = |x - y|,$$

то

$$(\mathbb{Q}, \rho)$$

будет метрическим пространством.

В принципе, любое множество можно рассматривать как метрическое пространство. Действительно, если для элементов произвольного множества ввести так называемую дискретную метрику:

$$\rho(x,y) = \begin{cases} 0, & x = y \\ 1, & x \neq y \end{cases},$$

то получится метрическое пространство, которое называют **пространством изолированных точек**.

На одном и том же множестве можно задавать различные метрики (ниже дан пример), однако не следует считать, что метрику можно задавать произвольно. Дело в том, что при решении практических задач метрика, как правило, является частью постановки задачи.

3.2.2. Свойства метрики

Рассмотрим некоторые свойства метрики, которые могут быть выведены из её определения.

Свойство 1. Метрика является неотрицательной функцией:

$$\rho(x,y) \geq 0.$$

Доказательство:

По аксиоме тождества

$$\forall x \in M : \rho(x,x) = 0.$$

С другой стороны, по аксиоме неравенства треугольника:

$$\forall y \in M : \rho(x,x) \leq \rho(x,y) + \rho(y,x).$$

В силу аксиомы симметрии:

$$\rho(x,y)=\rho(y,x),$$

поэтому

$$\forall y\in M: 0=\rho(x,x)\leq 2\rho(x,y).$$

Откуда и получается, что

$$\rho(x,y)\geq 0.$$

Свойство 2 (Неравенство многоугольника). Для любой конечной системы элементов

$$x_1, x_2, ..., x_n$$

множества *M* имеет место неравенство

$$\rho(x_1,x_n)\leq \rho(x_1,x_2)+...+\rho(x_{n-1},x_n)=\sum_{i=1}^{n-1}\rho(x_i,x_{i+1})$$

.

Доказательство проводится с помощью метода математической индукции, база индукции — аксиома треугольника.

Предположим, утверждение верно для некоторого целого числа m. Рассмотрим систему из $m+1$ элемента. По аксиоме треугольника:

$$\rho(x_1,x_{m+1})\leq \rho(x_1,x_m)+\rho(x_m,x_{m+1}).$$

Но по предположению:

$$\rho(x_1,x_m)\leq \rho(x_1,x_2)+...+\rho(x_{m-1},x_m),$$

следовательно:

$$\rho(x_1, x_{m+1}) \leq \rho(x_1, x_2) + ... + \rho(x_{m-1}, x_m) + \rho(x_m, x_{m+1})$$

Таким образом утверждение доказано для всех целых чисел $n > 1$.

Свойство 3 (Неравенство четырёхугольника). Для любых четырёх элементов

$$x, y, z, u \in M$$

имеет место неравенство

$$|\rho(x,z) - \rho(y,u)| \leq \rho(x,y) + \rho(z,u).$$

Доказательство основано на применении неравенства многоугольника для $n = 4$:

$$\rho(x,z) \leq \rho(x,y) + \rho(y,u) + \rho(u,z) \Rightarrow \rho(x,z) - \rho(y,u) \leq \rho(x,y) + \rho(u,z),$$
$$\rho(y,u) \leq \rho(x,y) + \rho(x,z) + \rho(u,z) \Rightarrow \rho(y,u) - \rho(x,z) \leq \rho(x,y) + \rho(u,z).$$

Сравнивая два эти неравенства, получим

$$|\rho(x,z) - \rho(y,u)| \leq \rho(x,y) + \rho(z,u).$$

Свойство 3а (Второе неравенство треугольника). Для любых трёх элементов

$$x, y, z \in M$$

имеет место неравенство

$$|\rho(x,z) - \rho(z,y)| \leq \rho(x,y).$$

Для доказательства нужно положить $u = z$ в неравенстве четырёхугольника.

3.2.3. Важные неравенства

Для проверки аксиомы неравенства треугольника для различных пространств полезны следующие леммы.

Лемма 1 (неравенство Коши-Буняковского):

Для любых двух конечных наборов вещественных чисел $(a_1, \ldots, a_n)$ и $(b_1, \ldots, b_n)$ имеет место неравенство:

$$\left|\sum_{i=1}^{n} a_i b_i\right| \leq \sqrt{\sum_{i=1}^{n} a_i^2} \cdot \sqrt{\sum_{i=1}^{n} b_i^2}.$$

Доказательство.

Определим функцию вещественного переменного $F(t)$ следующим образом

$$F(t) = \sum_{i=1}^{n} (a_i t + b_i)^2 \geq 0.$$

Применим формулу квадрата суммы:

$$F(t) = t^2 \sum_{i=1}^{n} a_i^2 + 2t \sum_{i=1}^{n} a_i b_i + \sum_{i=1}^{n} b_i^2.$$

Пусть сначала все a_i равны нулю. В этом случае

$$\sum_{i=1}^{n} a_i^2 = 0,$$

$$\sum_{i=1}^{n} a_i b_i = 0.$$

Так как $0 \leq 0$, то в этом случае неравенство Коши-Буняковского действительно имеет место.

Теперь будем считать, что

$$\sum_{i=1}^{n} a_i^2 > 0.$$

Поскольку

$$F(t) \geq 0$$

как сумма квадратов, то дискриминант Dквадратичной относительно t функции $F(t)$должен быть меньше либо равен нулю. Так как

$$D = 4\left(\sum_{i=1}^{n} a_i b_i\right)^2 - 4\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2,$$

то

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 - \sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2 \leq 0$$

и следовательно

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2.$$

Лемма доказана.

Лемма 2 (неравенство Минковского):

Для любых двух конечных наборов вещественных чисел $(a_1, \ldots, a_n)$ и $(b_1, \ldots, b_n)$ имеет место неравенство:

$$\sqrt{\sum_{i=1}^{n}(a_i + b_i)^2} \le \sqrt{\sum_{i=1}^{n} a_i^2} + \sqrt{\sum_{i=1}^{n} b_i^2}.$$

Доказательство.

По формуле квадрата суммы и в силу неравенства Коши-Буняковского:

$$\sum_{i=1}^{n}(a_i + b_i)^2 = \sum_{i=1}^{n} a_i^2 + 2\sum_{i=1}^{n} a_i b_i + \sum_{i=1}^{n} b_i^2 \le \sum_{i=1}^{n} a_i^2 + 2\sqrt{\sum_{i=1}^{n} a_i^2} \cdot \sqrt{\sum_{i=1}^{n} b_i^2} + \sum_{i=1}^{n} b_i^2.$$

Правая часть этого неравенства может быть записана в виде квадрата суммы:

$$\sum_{i=1}^{n} a_i^2 + 2\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2} + \sum_{i=1}^{n} b_i^2 = \left(\sqrt{\sum_{i=1}^{n} a_i^2} + \sqrt{\sum_{i=1}^{n} b_i^2}\right)^2.$$

Таким образом

$$\sum_{i=1}^{n}(a_i + b_i)^2 \le \left(\sqrt{\sum_{i=1}^{n} a_i^2} + \sqrt{\sum_{i=1}^{n} b_i^2}\right)^2.$$

Неравенство Минковского получается после извлечения квадратного корня из правой и левой части данного неравенства.

Лемма 3 (Интегральное неравенство Коши-Буняковского).

$$\int_a^b f(t)g(t)dt \le \sqrt{\int_a^b f^2(t)dt} \cdot \sqrt{\int_a^b g^2(t)dt}.$$

Доказательство.

Если одна из функций равна нулю на всём $[a; b]$, то левая и правая части нестрого неравенства равны нулю и лемма доказана. Теперь будем считать, что обе функции не равны тождественно нулю на всём $[a; b]$.

Рассмотрим неотрицательную функцию

$$F(\lambda) = \int_a^b [f(t)\lambda + g(t)]^2 dt.$$

По свойствам интеграла и формуле квадрата суммы:

$$F(\lambda) = \lambda^2 \int_a^b f^2(t)dt + 2\lambda \int_a^b f(t)g(t)dt + \int_a^b g^2(t)dt.$$

Функция $F(\lambda)$ является квадратичной и неотрицательной, значит её дискриминант должен быть меньше либо равен нулю:

$$4\left(\int_a^b f(t)g(t)dt\right)^2 - 4\int_a^b f^2(t)dt \cdot \int_a^b g^2(t)dt \le 0,$$

откуда и следует утверждение леммы.

Лемма 4 (Интегральное неравенство Минковского).

$$\sqrt{\int_a^b [f(t)+g(t)]^2 dt} \le \sqrt{\int_a^b f^2(t)dt} + \sqrt{\int_a^b g^2(t)dt}.$$

Доказательство.

По свойствам интеграла:

$$\int_a^b [f(t)+g(t)]^2 dt = \int_a^b f^2(t)dt + 2\int_a^b f(t)g(t)dt + \int_a^b g^2(t)dt$$

Воспользуемся интегральным неравенством Коши-Буняковского:

$$\int_a^b [f(t)+g(t)]^2 dt \le \int_a^b f^2(t)dt + 2\sqrt{\int_a^b f^2(t)dt} \cdot \sqrt{\int_a^b g^2(t)dt} + \int_a^b g^2(t)dt,$$

откуда следует

$$\int_a^b [f(t)+g(t)]^2 dt \le \left(\sqrt{\int_a^b f^2(t)dt} + \sqrt{\int_a^b g^2(t)dt}\right)^2.$$

Интегральное неравенство Минковского получается после извлечения квадратного корня из правой и левой части данного неравенства.

3.2.4. Примеры метрических пространств

Мы уже рассмотрели два метрических пространства: множество вещественных чисел и множество рациональных чисел. Ниже приведены ещё некоторые примере метрических пространств, все они играют важную роль в математическом анализе и алгебре.

Проверка первых двух аксиом является, как правило, тривиальной задачей, основные трудности связаны с доказательством справедливости аксиомы треугольника.

3.2.4.1. Арифметическое евклидово пространство

Множество $\mathbb{R}^n$ с метрикой

$$\rho(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}$$

является метрическим пространством.

Действительно, рассмотрим любые три элемента из множества $\mathbb{R}^n$:

$$x=(x_1,...,x_n),$$
$$y=(y_1,...,y_n),$$
$$z=(z_1,...,z_n).$$

Тогда

1. $\rho(x,y)=0\Leftrightarrow\sqrt{\sum_{k=1}^{n}(x_i-y_i)^2}=0\Leftrightarrow\forall i=\overline{1,n}\quad(x_i-y_i)^2=0\Leftrightarrow\ \forall i=\overline{1,n}\quad x_i=y_i\Leftrightarrow x=y$.

2 $\rho(x,y)=\sqrt{\sum_{k=1}^{n}(x_i-y_i)^2}=\sqrt{\sum_{k=1}^{n}((-1)(y_i-x_i))^2}=\sqrt{\sum_{k=1}^{n}(y_i-x_i)^2}=\rho(y,x)$.

Перейдём к проверке третьей аксиомы.
3.

$$\rho(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}=\sqrt{\sum_{i=1}^{n}(x_i-z_i+z_i-y_i)^2}.$$

По неравенству Минковского (Лемма 2):

$$\rho(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-z_i+z_i-y_i)^2}\leq\sqrt{\sum_{i=1}^{n}(x_i-z_i)^2}+\sqrt{\sum_{i=1}^{n}(z_i-y_i)^2}=\rho(x,z)+\rho(z,y),$$

то есть аксиома действительно выполняется.

Таким образом, $(\mathbb{R}^n,\rho)$ — метрическое пространство.

3.2.4.2. Метрика Хэмминга

Снова рассмотрим множество $\mathbb{R}^n$, но расстояние в нём определим как сумму расстояний между координатами:

$$\rho_1(x,y)=\sum_{k=1}^{n}|x_i-y_i|.$$

Метрика такого вида называется метрикой Хэмминга.

Метрическое пространство (R^n,ρ_1)обозначают $\mathbb{R}^n_1$.

3.2.4.3. Равномерная метрика

На множестве $\mathbb{R}^n$можно ввести ещё одну метрику

$$\rho_\infty(x,y)=\max_{1\leq k\leq n}|x_k-y_k|.$$

Пространство с данной метрикой обозначают $\mathbb{R}^n_\infty$.

Таким образом, три рассмотренных примера показывают, что на основе одного и того же множество можно, задавая различные метрики, строить различные метрические пространства.

3.2.4.4. Комплексные числа

Множество комплексных чисел $\mathbb{C}$ с метрикой

$$\rho(z_1, z_2) = |z_1 - z_2|$$

является метрическим пространством. Справедливость аксиом следует из свойств модуля комплексного числа. Действительно, если $z_1 = x_1 + iy_1$, а $z_2 = x_2 + iy_2$, то

$$|z_1 - z_2| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2},$$

таким образом, с точки зрения теории метрических пространств, множество комплексных чисел эквивалентно двумерному арифметическому евклидову пространству (геометрическая интерпретация комплексных чисел).

3.2.4.5. Непрерывные функции

Множество непрерывных на отрезке $[a, b]$ функций $C[a; b]$ с метрикой

$$\rho(f, g) = \max_{x\in[a;b]} |f(x) - g(x)|$$

является метрическим пространством.

Если $f = g$, то очевидно, что $\rho(f, g) = 0$. Наоборот, если $\rho(f, g) = 0$, то по определению метрики для данного пространства:

$$\max_{x\in[a;b]} |f(x) - g(x)| = 0,$$

так как

$$|f(x) - g(x)| \le \max_{x \in [a;b]} |f(x) - g(x)| = 0$$,

то функции f и g равны друг другу на отрезке $[a, b]$. Аксиома тождества доказана.

Аксиома симметрии:

$$\rho(x,y) = \max_{x \in [a;b]} |f(x) - g(x)| = \max_{x \in [a;b]} |-(f(x) - g(x))| = \max_{x \in [a;b]} |g(x) - f(x)| = \rho(y,x)$$.

Аксиома симметрии тоже выполняется.

Докажем теперь аксиому треугольника. Для любых трёх функций

$$f, g, h \in C[a;b]$$,

в силу неравенства треугольника для модуля, выполняется неравенство

$$|f(x) - g(x)| = |f(x) - h(x) + h(x) - g(x)| \le |f(x) - h(x)| + |h(x) - g(x)|$$.

Возьмём максимальное значение левой и правой части:

$$\max_{x \in [a;b]} |f(x) - g(x)| \le \max_{x \in [a;b]} \{|f(x) - h(x)| + |g(x) - h(x)|\}$$.

Так как на отрезке $[a; b]$ для любых двух функций $w_1, w_2 \in C[a;b]$, в силу определения наибольшего значения, имеет место неравенство

$$w_1(x) + w_2(x) \le \max_{x \in [a;b]} w_1(x) + \max_{x \in [a;b]} w_2(x)$$,

а следовательно

$$\max_{x\in[a;b]}\{w_1(x)+w_2(x)\} \le \max_{x\in[a;b]} w_1(x) + \max_{x\in[a;b]} w_2(x),$$

то есть наибольшее значение суммы функций не превосходит суммы их наибольших значений.

Используем последнее неравенство, положив

$$w_1(x) = |f(x) - h(x)|,\ w_2(x) = |h(x) - g(x)|,$$

получим

$$\max_{x\in[a;b]}\{|f(x)-h(x)|+|h(x)-g(x)|\} \le \max_{x\in[a;b]} |f(x)-h(x)| + \max_{x\in[a;b]} |h(x)-g(x)|.$$

А значит:

$$\rho(f,g) = \max_{x\in[a;b]} |f(x)-g(x)| \le \max_{x\in[a;b]} |f(x)-h(x)| + \max_{x\in[a;b]} |h(x)-g(x)| = \rho(f,h)+\rho(h,g).$$

Все аксиомы действительно выполняются.

На множестве непрерывных функций метрику можно определить иначе, например

$$\rho(x,y) = \sqrt{\int_a^b [x(t)-y(t)]^2 dt},$$

полученное метрическое пространство обозначают $C_2[a;b]$.

3.2.4.6. Пространства числовых последовательностей

Рассмотрим множество всевозможных числовых последовательностей вида

$$x = \{x_1, ..., x_n, ...\},$$

удовлетворяющих условию

$$\sum_{k=1}^{\infty} x_k^2 < \infty$$
.

Если на этом множестве ввести расстояние

$$\rho(x,y) = \sqrt{\sum_{k=1}^{\infty}(x_k - y_k)^2}$$
,

то получим метрическое пространство, которое обозначают l_2. Ряд

$$\sum_{k=1}^{\infty}(x_k - y_k)^2$$

сходится, если сходятся ряды

$$\sum_{k=1}^{\infty} x_k^2, \quad \sum_{k=1}^{\infty} y_k^2$$
,

а значит введённая метрика имеет смысл для любых последовательностей из l_2.

Доказательство аксиом тождества и симметрии не представляет труда, доказательство справедливости аксиомы треугольника для указанной метрики можно с помощью предельного перехода в неравенстве Минковского.

Приведённые примеры показывают, что понятия метрики и метрического пространства позволяют рассматривать с единых позиций такие непохожие на первый взгляд объекты, как вещественные и комплексные числа, вектора, непрерывные функции и числовые последовательности.

4. Линейные пространства.

4.1. Определение линейного пространства. Изоморфизм

Определение n-мерного векторного пространства, которое было приведено раньше, начиналось из определения n-мерного вектора как упорядоченной системы чисел. Для n-мерных векторов были введены затем сложение и умножение на числа, что и привело к понятию n-мерного векторного пространства. Первыми примерами векторных пространств являются совокупности векторов-отрезков, которые выходят из начала координат на плоскости или в трехмерном пространстве. Однако, встречаясь с этими примерами в курсе геометрии, мы не всегда считаем необходимым задавать векторы их компонентами в некоторой фиксированной системе координат, так как и сложение векторов и их умножение на скаляр определяются геометрически, независимо от выбора системы координат. Именно, сложение векторов на плоскости или в просторанстве выполняется по правилу параллелограмма, а умножение вектора на число α означает растяжение этого вектора в α раз (с изменением направления вектора на противоположный, если α отрицательно). Целесообразно и в общем случае дать «бескоординатное» определение векторного просторанства, т.е. определение, не требующее задання векторов упорядоченными системами чисел. Приведем такое определение. Это определение есть аксиоматическим: в нем ничего не будет сказано о свойствах отдельного вектора, но будут перечислены те свойства, которыми должны обладать операции над векторами.

Пусть дано множество M. Пусть, далее, в множестве M определены *операция сложения,* которая ставит в соответствие всякой паре элементов $\boldsymbol{a}$, $\boldsymbol{b}$ из M однозначно определенный элемент $\boldsymbol{a}+\boldsymbol{b}$ из M, называемый их *суммой,* и *операция умножения на действительное число,* причем *произведение* $\alpha\boldsymbol{a}$ *элемента* $\boldsymbol{a}$ на число α однозначно определено и принадлежит к M.

Элементы множества M будут называться *векторами,* а само M - *действительным линейным* (или *векторным,* или *аффинным) пространством,* если указанные операции обладают следующими свойства **I—VIII:**

I. Сложение коммутативно, $\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{b}+\boldsymbol{a}$.

II. Сложение ассоциативно, $(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}=\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c})$.

III. В M существует *нулевой, элемент* 0, удовлетворяющий условию:

$\boldsymbol{a}+0=\boldsymbol{a}$ для всех $\boldsymbol{a}$ из M.

Легко доказать, используя I, *единственность нулевого элемента:* если 0_1 и 0_2 — два нулевых элемента, то

$$0_1 + 0_2 = 0_1.$$

$$0_1 + 0_2 = 0_2+0_1 = 0_2.$$

откуда $0_1 = 0_2$.

IV. Для всякого элемента $\boldsymbol{a}$ в M существует *противоположный элемент* $-\boldsymbol{a}$, который удовлетворяет условию: $\boldsymbol{a}+(-\boldsymbol{a}) = 0$.

Легко проверяется, ввиду II и I, *единственность противоположного элемента:* если $(-\boldsymbol{a})_1$ и $(-\boldsymbol{a})_2$ — два противоположных элемента для $\boldsymbol{a}$, то

$$(-\boldsymbol{a})_1+[\boldsymbol{a}+(-\boldsymbol{a})_2] = (-\boldsymbol{a})_1+0 = (-\boldsymbol{a})_1;$$

$$[(-\boldsymbol{a})_1 + \boldsymbol{a}] + (-\boldsymbol{a})_2 = 0 + (-\boldsymbol{a})_2 = (-\boldsymbol{a})_2,$$

откуда $(-\boldsymbol{a})_1=(-\boldsymbol{a})_2$.

Из аксиом I—IV выводится *существование и единственность разности* $\boldsymbol{a} - \boldsymbol{b}$, т.е. такого элемента, который удовлетворяет уравнению

$$\boldsymbol{b}+\boldsymbol{x}=\boldsymbol{a}. \tag{4.1}$$

Действительно, можно положить

$$\boldsymbol{a}-\boldsymbol{b}=\boldsymbol{a}+(-\boldsymbol{b}),$$

так как

$$\boldsymbol{b} + [\boldsymbol{a} + (-\boldsymbol{b})] = [\boldsymbol{b} + (-\boldsymbol{b})] + \boldsymbol{a} = 0 + \boldsymbol{a} = \boldsymbol{a}.$$

Если же существует еще такой элемент **с**, который удовлетворяет уравнению (4.2), т.е.

$$\boldsymbol{b}+\boldsymbol{c} = \boldsymbol{a},$$

то, прибавляя к обеим частям этого равенства элемент $-\boldsymbol{b}$, получаем, что

$$\boldsymbol{c} = \boldsymbol{a} + (-\boldsymbol{b}).$$

Дальнейшие аксиомы **V—VIII** связывают умножение на число со сложением и с операциями над числами. Именно, для любых элементов $\boldsymbol{a}$, $\boldsymbol{b}$ из M, для любых действительных чисел α, β и для действительного числа 1 должны иметь место равенства:

V. $\alpha (\boldsymbol{a}+\boldsymbol{b}) = \alpha\boldsymbol{a} + \alpha\boldsymbol{b}$;

VI. $(\alpha + \beta)\boldsymbol{a} = \alpha\boldsymbol{a} + \beta\boldsymbol{a}$;

VII. $(\alpha\beta)\boldsymbol{a} =\alpha(\beta\boldsymbol{a})$;

VIII. $1\bullet\boldsymbol{a} = \boldsymbol{a}$

Укажем некоторые простейшие следствия из этих аксиом.

1) $\alpha\bullet 0 =0$.

Действительно, для некоторого $\boldsymbol{a}$ из M

$$\alpha\boldsymbol{a} =\alpha(\boldsymbol{a}+0) =\alpha\boldsymbol{a}+\alpha 0,$$

т.е.

$$\alpha\bullet 0=\alpha\boldsymbol{a}- \alpha\boldsymbol{a}=\alpha\boldsymbol{a}+[-(\alpha\boldsymbol{a})] = 0.$$

2) $0\bullet \boldsymbol{a} = 0$,

где слева стоит число нуль, а справа - нулевой элемент из M.

Для доказательства возьмем любое число α. Тогда

$$\alpha\boldsymbol{a}=(\alpha+0)\boldsymbol{a} = \alpha\boldsymbol{a}+0\bullet \boldsymbol{a},$$

откуда

$$0\bullet \boldsymbol{a} = \alpha\boldsymbol{a} - \alpha\boldsymbol{a} = 0.$$

3) Если $\alpha\boldsymbol{a}=0$, то или $\alpha=0$, или $\boldsymbol{a}=0$.

Действительно, если $\alpha\neq 0$, т.е. число α^{-1} существует, то

$$\alpha=1\bullet \alpha=(\alpha^{-1}\alpha)\boldsymbol{a} = \alpha^{-1}(\alpha\boldsymbol{a}) = \alpha^{-1}\bullet 0=0.$$

4) $\alpha(-\boldsymbol{a}) = - \alpha\boldsymbol{a}$.

В самом деле,

$$\alpha\boldsymbol{a} + \alpha(-\boldsymbol{a}) = \alpha\,[\boldsymbol{a}+(-\boldsymbol{a})]=\alpha\bullet 0 = 0,$$

т. е. элемент $\alpha(-\boldsymbol{a})$ противоположен элементу $\alpha\boldsymbol{a}$.

5) $(-\alpha)\boldsymbol{a} = -\alpha\boldsymbol{a}$.

Действительно,

$$\alpha\boldsymbol{a}+(-\alpha)\boldsymbol{a} = [\alpha+(-\alpha)]\boldsymbol{a} = 0\bullet \boldsymbol{a} = 0,$$

т.е. элемент $(-\alpha)\boldsymbol{a}$ противоположен элементу $\alpha\boldsymbol{a}$.

6) $\alpha(\boldsymbol{a} - \boldsymbol{b})=\alpha\boldsymbol{a} - \alpha\boldsymbol{b}$

Действительно, по (4.8),

$$\alpha(\boldsymbol{a} - \boldsymbol{b}) = \alpha[\boldsymbol{a} + (-\boldsymbol{b})] = \alpha\boldsymbol{a}+\alpha(-\boldsymbol{b}) = \alpha\boldsymbol{a}+(-\alpha\boldsymbol{b})=\alpha\boldsymbol{a} - \alpha\boldsymbol{b}.$$

7) $(\alpha-\beta)\boldsymbol{a} = \alpha\boldsymbol{a} - \beta\boldsymbol{a}$.

В самом деле,

$$(\alpha - \beta)\boldsymbol{a}= [\alpha + (-\beta)]\,\boldsymbol{a} = \alpha\boldsymbol{a} + (-\beta)\boldsymbol{a} = \alpha\boldsymbol{a} + (-\beta\boldsymbol{a}) = \alpha\boldsymbol{a} - \beta\boldsymbol{a}.$$

Заметим, что вышеперечисленными аксиомами и следствиями из них мы будем пользоваться дальше без специальных оговорок.

Выше дано определения действительного линейного пространства. Если бы мы предположили, что в множестве M определено умножения не только на действительные, но и на любые комплексные числа, то, сохраняя те же аксиомы I—VIII, получили бы определение *комплексного линейного пространства.* Для определенности ниже рассматриваются действительные линейные пространства, однако все, что будет сказано в этом микромодуле, переносится дословно на случай комплексных линейных пространств.

Примеры действительных линейных пространств могут быть легко указанные. Ими будут, прежде всего, те n-мерные действительные векторные пространства, составленные из векторов-строк, которые изучались раньше. Линейными пространствами будут и множества векторов-отрезков, которые выходят из начала координат на плоскости или в трехмерном пространстве, если операции сложения и умножения на число понимать в том геометрическом смысле, который было указан в начале настоящего раздела.

Существуют также примеры линейных пространств, так сказать, «бесконечномерных». Рассмотрим всевозможные последовательности действительных чисел; они имеют вид

$$\boldsymbol{a} = (\alpha_1, \alpha_2, \dots, \alpha_n, \dots).$$

Операции над последовательностями будем производить покомпонентно: если

$$\boldsymbol{b}=(\beta_1, \beta_2, \dots, \beta_n, \dots),$$

то

$$\boldsymbol{a} +\boldsymbol{b}=(\alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots \alpha_n + \beta_n \dots);$$

с другой стороны, для любого действительного числа γ

$$\gamma\boldsymbol{a} = (\gamma\alpha_1, \gamma\alpha_2,\dots,\gamma\alpha_n,\dots)...$$

Все аксиомы I-VIII выполняются, т.е. мы получаем действительное линейное пространство.

Примером бесконечномерного пространства будет также множество всевозможных действительных функций действительного переменного, если сложение функций и их умножение на действительное число понимать так, как это принято в теории функций, т.е. как сложение или умножение на число значений функций при каждом значении независимого переменного.

Изоморфизм. Нашей ближайшей целью будет выделение среди всех линейных пространств тех, которые естественно назвать конечномерными. Введем сначала одно общее понятие.

В определении линейного пространства говорилось о свойствах операций над векторами, но ничего не говорилось о свойствах самих векторов. Ввиду этого может случиться, что хотя векторы некоторых двух данных линейных пространств по своей природе совершенно разные, однако с точки зрения свойств операций эти два пространств неразличимые. Точное определение таково:

Два действительных линейных пространства M и M' называются *изоморфными,* если между их векторами установлено взаимно-однозначное соответствие — всякому вектору $\boldsymbol{a}$ из M сопоставлен вектор $\boldsymbol{a}'$ из M', *образ* вектора $\boldsymbol{a}$, причем различные векторы из M обладают различными образами и всякий вектор из M' служит образом некоторого вектора из M,— и если при этом соответствии образом суммы двух векторов служит сумма образов этих векторов,

$$(\boldsymbol{a} + \boldsymbol{b})'=\boldsymbol{a}' + \boldsymbol{b}', \qquad (4.2)$$

а образом произведения вектора на число служит произведение образа этого вектора на то же число,

$$(\alpha\boldsymbol{a})'=\alpha\boldsymbol{a}'. \qquad (4.3)$$

Отметим, что взаимно-однозначное соответствие между пространствами M и M', удовлетворяющее условиям (4.2) и (4.3), называется *изоморфным соответствием.*

Так, пространство векторов-отрезков на плоскости, которые выходят из начала координат, изоморфно двумерному векторному пространству, составленному из упорядоченных пар действительных чисел: мы получим изоморфное соответствие между этими пространствами, если на плоскости зафиксируем некоторую систему координат и всякому вектору-отрезку сопоставим упорядоченную пару его координат.

Докажем следующее свойство изоморфизма линейных пространств: *образом нуля пространства M при изоморфном соответствии между пространствами M и M' служит нуль пространства M'.*

Пусть, в самом деле, $\boldsymbol{a}$ будет некоторый вектор из M, $\boldsymbol{a}'$ - его образ в M'. Тогда, ввиду (4.2),

$$\boldsymbol{a}' = (\boldsymbol{a} + 0)\,' = \boldsymbol{a}' + 0',$$

т.е. 0' будет нулем пространства M'.

4.2. Конечномерные пространства. Базисы

Как мы заем, те два определения линейной зависимости векторов-строк, которые были даны раньше, равно как и доказательство эквивалентности этих определений, используют лишь операции над векторами и потому могут быть перенесены на случай любых линейных пространств.

В аксиоматически определенных линейных пространствах можно говорить, следовательно, о линейно независимых системах векторов, о максимальных линейно независимых системах, если такие существуют, и т.д.

Если линейные пространства M и M' изоморфны, то система векторов $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k$ *из M тогда и только тогда линейно зависима, если линейно зависима система их образов* $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k$ *в M'.*

Заметим, что если соответствие $\boldsymbol{a} \to \boldsymbol{a}'$ (для всех $\boldsymbol{a}$ из M) является изоморфным соответствием между M и M', то и обратное соответствие $\boldsymbol{a}' \to \boldsymbol{a}$ также будет изоморфным. Поэтому достаточно рассмотреть случай, когда линейно зависимая система $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k$. Пусть существуют такие числа $\alpha_1, \alpha_2, \dots, \alpha_k$, не все равные нулю, что

$$\alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2 + \dots + \alpha_k\boldsymbol{a}_k = 0.$$

Образом правой части этого равенства при рассматриваемом изоморфизме служит, как мы знаем, нуль 0' пространства M'. Беря образ левой части и применяя несколько раз (4.2) и (4.3), получаем

$$\alpha_1\boldsymbol{a}'_1 + \alpha_2\boldsymbol{a}'_2 + \dots + \alpha_k\boldsymbol{a}'_k = 0',$$

т.е. система $\boldsymbol{a}'_1, \boldsymbol{a}'_2, \dots, \boldsymbol{a}'_k$ также оказалась линейно зависимой.

Конечномерные пространства. Линейное пространство M называется *конечномерным,* если в нем можно найти конечную максимальную линейно независимую систему векторов; всякая такая система векторов будет называться *базисом* пространства *M*.

Конечномерное линейное пространство может обладать многими различными базисами. Так, в пространстве векторов-отрезков на плоскости базисом служит любая пара векторов, отличных от нуля и не лежащих на одной прямой. Заметим, что приведенное определение конечномерного пространства не дает пока ответа на вопрос, могут ли в этом пространстве существовать базисы, которые состоят из разного числа векторов. Больше того, можно было бы допустить даже, что в некоторых конечномерных пространствах существуют базисы со сколь угодно большим числом векторов. Сейчас мы приступим к выяснению того, каково же положение есть на самом деле.

Пусть линейное пространство *M* обладает базисом

$$\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n, \qquad (4.4)$$

который состоит из *n* векторов. Если $\boldsymbol{a}$ — произвольный вектор из *M*, то из максимальности линейно независимой системы (4.4) следует, что $\boldsymbol{a}$ линейно выражается через эту систему,

$$\boldsymbol{a} = \alpha_1\boldsymbol{e}_1 + \alpha_2\boldsymbol{e}_2 + \ldots + \alpha_n\boldsymbol{e}_n. \qquad (4.5)$$

С другой стороны, ввиду линейной независимости системы (4.4) выражение (4.5) будет для вектора $\boldsymbol{a}$ единственным: если

$$\boldsymbol{a} = \alpha'_1\boldsymbol{e}_1 + \alpha'_2\boldsymbol{e}_2 + \ldots + \alpha'_n\boldsymbol{e}_n,$$

то

$$(\alpha_1-\alpha'_1)\boldsymbol{e}_1 + (\alpha_2-\alpha'_2)\boldsymbol{e}_2 + \ldots + (\alpha_n-\alpha'_n)\boldsymbol{e}_n = 0,$$

откуда

$$\alpha_i = \alpha'_i, \quad i = 1, 2, \ldots, n.$$

Таким образом, вектору $\boldsymbol{a}$ однозначно соответствует строка

$$(\alpha_1, \alpha_2, \ldots, \alpha_n) \qquad (4.6)$$

коэффициентов его выражения (4.5) через базис (4.4) или, как мы будем говорить, *строка его координат в базисе* (4.4). Обратно, всякая строка вида (4.6), т.е. всякий *n*-мерный вектор, служит строкой координат в базисе (4.4) для некоторого вектора пространства *M*, а именно для вектора, который записывается через базис (4.4) в виде (4.5).

Мы получили, следовательно, взаимно-однозначное соответствие между всеми векторами пространства *M* и всеми векторами *n*-мерного векторного пространства строк. Покажем, что это соответствие, которое зависит от выбора базиса (4.4), является изоморфным.

Возьмем в пространстве *M*, помимо вектора $\boldsymbol{a}$, который выражается через базис (4.4) в виде (4.5), также вектор *b*, выражение которого через базис (4.4) будет

$$\boldsymbol{b} = \beta_1\boldsymbol{e}_1 + \beta_2\boldsymbol{e}_2 + \ldots + \beta_n\boldsymbol{e}_n.$$

Тогда

$$\boldsymbol{a} + \boldsymbol{b} = (\alpha_1 + \beta_1)\boldsymbol{e}_1 + (\alpha_2 + \beta_2)\boldsymbol{e}_2 + \ldots + (\alpha_г + \beta_г)\boldsymbol{e}_n,$$

т.е. *сумме векторов* ***a*** *и* **b** *соответствует сумма строк их координат в базисе* (4.4). С другой стороны,

$$\gamma\boldsymbol{a} = (\gamma\alpha_1)\boldsymbol{e}_1 + (\gamma\alpha_2)\,\boldsymbol{e}_2 + \ldots + (\gamma\alpha_n)\,\boldsymbol{e}_n,$$

т.е. *произведению вектора* ***a*** *на число соответствует произведение строки его координат в базисе* (4.4) *на это же число* γ.

Этим доказана следующая теорема:

Всякое линейное пространство, обладающее базисом из n векторов, изоморфно n-мерному векторному пространству строк.

Как мы знаем, при изоморфном соответствии между линейными пространствами линейно зависимая система векторов переходит в линейно зависимую и обратно, а потому линейно независимая переходит в линейно независимую. Отсюда следует, что *при изоморфном соответствии базис переходит в базис.*

В самом деле, пусть базис $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ пространства *М* переходит при изоморфном соответствии между пространствами *М* и *М'* в систему векторов $\boldsymbol{e}'_1, \boldsymbol{e}'_2, \ldots, \boldsymbol{e}'_n$ пространства *М'*, которая хотя и линейно независима, но не является максимальной. В *М'* можно найти, следовательно, такой вектор ***f'***, что система $\boldsymbol{e}'_1, \boldsymbol{e}'_2, \ldots, \boldsymbol{e}'_n, \boldsymbol{f}'$ остается линейно независимой. Вектор ***f'*** служит, однако, образом при рассмотренном изоморфизме для некоторого вектора ***f*** из *М*.

Мы получаем, что система векторов $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n, \boldsymbol{f}$ должна быть линейно независимой в противоречии с определением базиса.

Мы знаем из векторного анализа, что в *n*-мерном векторном пространстве строк все максимальные линейно независимые системы состоят из *n* векторов, что всякая система из *n*+1 вектора линейно зависима и что всякая линейно независимая система векторов содержится в некоторой максимальной линейно независимой системе. Используя установленные выше свойства изоморфных соответствий, мы приходим к следующим результатам:

Все базисы конечномерного линейного пространства М состоят из одного и того же числа векторов. Если это число равно *n*, то *М* будет называться *n-мерным линейным пространством,* а число *n* - *размерностью* этого пространства.

Всякая система из n+1 векторов n-мерного линейного пространства линейно зависима.

Всякая линейно независимая система векторов n-мерного линейного пространства содержится в некотором базисе этого пространства.

Теперь легко проверить, что указанные выше примеры действительных линейных пространств - пространство последовательностей и пространство функций - не являются конечномерными пространствами: в каждом из этих пространств легко найти линейно независимые системы, которые состоят из сколь угодно большого числа векторов.

Связь между базисами. Объектом изучения являются для нас конечномерные линейные пространства. Понятно, что, изучая *n*-мерные линейные пространства, мы собственно говоря изучаем *n*-мерное векторное пространство строк. Однако при изученные этого пространства в этом пространстве было выделен один базис - а именно базис, составленный из единичных векторов, т.е. векторов, у которых одна координата равна единице, а все остальные координаты равны нулю, - и все векторы пространства задаются строками их координат в этом базисе; теперь же все базисы пространства являются для нас равноправными.

Посмотрим, как много базисов можно найти в *n*-мерном линейном пространстве и как эти базисы связанны друг с другом.

Пусть в *n*-мерном линейном пространстве *M* заданы базисы

$$\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \tag{4.7}$$

и

$$\mathbf{e}'_1, \mathbf{e}'_2, \ldots, \mathbf{e}'_n. \tag{4.8}$$

Каждый вектор базиса (4.8) , как и всякий вектор пространства *M,* однозначно записывается через базис (4.7)

$$\mathbf{e}'_i = \sum_{j=1}^{n} \tau_{ij}\, \mathbf{e}'_j, \quad i = 1, 2, \ldots, n. \tag{4.9}$$

Матрица

$$T = \begin{pmatrix} \tau_{11\ldots}\tau_{1n} \\ \ldots\ldots\ldots\ldots \\ \tau_{n1\ldots}\tau_{nn} \end{pmatrix},$$

строки которой являются строками координат векторов (4.8) в базисе (4.7), называется *матрицей перехода* от базиса (4.7) к базиса (4.8).

Связь между базисами (4.7) и (4.8) и матрицей перехода *T* можно записать, ввиду (4.9), в виде матричного равенства

$$\begin{pmatrix} \boldsymbol{e}_1' \\ \boldsymbol{e}_2' \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{e}_n' \end{pmatrix} = \begin{pmatrix} \tau_{11}\ \tau_{12} \dots \tau_{1n} \\ \tau_{21}\ \tau_{22} \dots \tau_{2n} \\ \dots\dots\dots\dots \\ \tau_{n1}\ \tau_{n2} \dots \tau_{nn} \end{pmatrix} \begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{e}_n \end{pmatrix} \qquad (4.10)$$

или, обозначая базисы (4.7) и (4.8), которые записаны в столбец, соответственно через $\boldsymbol{e}$ и $\boldsymbol{e}'$, в виде

$$\boldsymbol{e}' = T\boldsymbol{e}.$$

С другой стороны, если T' — матрица перехода от базиса (4.8) к базису (4.7), то

$$\boldsymbol{e} = T'\boldsymbol{e}'.$$

Отсюда

$$\boldsymbol{e}=(T'T)\boldsymbol{e},$$
$$\boldsymbol{e}' = (TT')\boldsymbol{e}',$$

т.е., через линейную независимость базисов $\boldsymbol{e}$ и $\boldsymbol{e}'$,

$$T'T=TT'=E,$$

откуда

$$T' = T^{-1}$$

Этим доказано, что *матрица перехода от одного базиса к другому всегда является невырожденной матрицей.*

Всякая невырожденная квадратная матрица порядка n с действительными элементами служит матрицей перехода от данного базиса n-мерного действительного линейного пространства к некоторому другому базису.

Пусть, в самом деле, дан базис (4.7) и невырожденная матрица T порядка n. Возьмем в качестве (4.8) систему векторов, для которых строки матрицы T служат строками координат в базисе (4.7); имеет место, следовательно, равенство (4.10). Векторы (4.8) линейно независимы — линейная зависимость между ними влекла бы за собой линейную зависимость строк матрицы T в противоречие с ее невырожденнотью. Поэтому система (4.8), как линейно независимая система, которая состоит из n векторов, является базисом нашего просторанства, а матрица T служит матрицей перехода от базиса (4.9) к базису (4.10).

Мы видим, что в n-мерном линейном просторанстве можно найти столь же много различных базисов, как много существует различных невырожденных квадратных матриц порядка n. Правда, при этом два базиса, состоящие из одних и тех же векторов, но записанных в различном порядке, считаются различными.

Преобразование координат вектора. Пусть в n-мерном линейном простоpанстве даны базисы (4.7) и (4.8) с матрицей перехода $T=(\tau_{ij})$,

$$\boldsymbol{e}' = T\boldsymbol{e}.$$

Найдем связь между строками координат произвольного вектора $\boldsymbol{a}$ этих базисов. Пусть

$$\boldsymbol{a}=\sum_{j=1}^{n}\alpha_j\ \boldsymbol{e}_j,\quad \boldsymbol{a}=\sum_{i=1}^{n}\alpha'_i\ \boldsymbol{e}'_i. \qquad (4.11)$$

Используя (4.9), получаем:

$$\boldsymbol{a}=\sum_{i=1}^{n}\alpha'_i\left(\sum_{j=1}^{n}\tau_{ij}\boldsymbol{e}_j\right)=\sum_{j=1}^{n}\left(\sum_{i=1}^{n}\alpha'_i\tau_{ij}\right)\boldsymbol{e}_j$$

Сравнивая с (4.11) и используя единственность записи вектора через базис, получаем:

$$\alpha_j=\sum_{i=1}^{n}\alpha'_i\tau_{ij}\ ,\quad j = 1, 2, ... n,$$

т.е. имеет место матричное равенство

$$(\alpha_1, \alpha_2, \dots , \alpha_n) = (\alpha'_1, \alpha'_2, \dots , \alpha'_n)T.$$

Таким образом, *строка координат вектора **a** в базисе **e** равна строке координат этого вектора в базисе **e**', умноженной справа на матрицу перехода от базиса **e** к базису **e**'* .

Отсюда следует равенство

$$(\alpha'_1, \alpha'_2, \dots , \alpha'_n)= (\alpha_1, \alpha_2, \dots , \alpha_n)T$$

Пример. Рассмотрим трехмерное действительное линейное пространство с базисом

$$\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3. \qquad (4.12)$$

Векторы

$$\left.\begin{aligned}\boldsymbol{e}'_1&=5\boldsymbol{e}_1-\boldsymbol{e}_2-2\boldsymbol{e}_3\\ \boldsymbol{e}'_2&=2\boldsymbol{e}_1+3\boldsymbol{e}_2\\ \boldsymbol{e}'_3&=-2\boldsymbol{e}_1+\boldsymbol{e}_2+\boldsymbol{e}_3\end{aligned}\right\} \qquad (4.13)$$

также составляют базис, в этом пространстве, причем матрицей перехода от (4.12) к (4.13) служит матрица

$$\boldsymbol{T}=\begin{pmatrix}5 & -1 & -2\\ 2 & 3 & 0\\ -2 & 1 & 1\end{pmatrix},$$

откуда

$$T^{-1} = \begin{pmatrix} 3 & -1 & 6 \\ -2 & 1 & -4 \\ 8 & -3 & 17 \end{pmatrix}$$

Вектор

$$\boldsymbol{a} = \boldsymbol{e}_1 + 4\boldsymbol{e}_2 - \boldsymbol{e}_3$$

имеет поэтому в базисе (4.13) строку координат

$$(\alpha'_1, \alpha'_2, \alpha'_3) = (1, 4, -1) \begin{pmatrix} 3 & -1 & 6 \\ -2 & 1 & -4 \\ 8 & -3 & 17 \end{pmatrix} = (-13, 6, -27),$$

т.е.

$$\boldsymbol{a} = -13\boldsymbol{e}'_1 + 6\boldsymbol{e}'_2 - 27\boldsymbol{e}'_3.$$

4.3. Линейные преобразования

Пусть дано n-мерное действительное линейное пространство, которое обозначим через M_n. Рассмотрим *преобразование* этого пространства, т.е. отображение, переводящее каждый вектор $\boldsymbol{a}$ пространства M_n в некоторый вектор $\boldsymbol{a}'$ этого же пространства.

Вектор $\boldsymbol{a}'$ называется *образом* вектора $\boldsymbol{a}$ при рассматриваемом преобразовании.

Если преобразование обозначено через φ, то образ вектора $\boldsymbol{a}$ условимся записывать не через $\varphi(\boldsymbol{a})$ или $\varphi\boldsymbol{a}$, а через $\boldsymbol{a}\varphi$. Таким образом,

$$\boldsymbol{a}' = \boldsymbol{a}\varphi.$$

Преобразование φ линейного пространства M_n называется *линейным преобразованием* этого пространства, если сумму любых двух векторов $\boldsymbol{a}$, $\boldsymbol{b}$ оно переводит в сумму образов этих векторов,

$$(\boldsymbol{a}+\boldsymbol{b})\varphi = \boldsymbol{a}\varphi + \boldsymbol{b}\varphi, \qquad (4.14)$$

а произведение любого вектора $\boldsymbol{a}$ на любое число α переводит в произведение образа вектора $\boldsymbol{a}$ на это же число α,

$$(\alpha\boldsymbol{a})\varphi = \alpha(\boldsymbol{a}\varphi). \qquad (4.15)$$

Из этого определения следует, что *линейное преобразование линейного пространства переводит любую линейную комбинацию данных векторов* $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ *в линейную комбинацию (с теми же коэффициентами) образов этих векторов,*

$$(\alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \ldots + \alpha_k \boldsymbol{a}_k)\varphi = \alpha_1(\boldsymbol{a}_1\varphi) + \alpha_2(\boldsymbol{a}_2\varphi) + \ldots + \alpha_k(\boldsymbol{a}_k\varphi). \quad (4.16)$$

Докажем следующее утверждение:

При любом линейном преобразовании φ *линейного пространства* M_n *нулевой вектор* 0 *остается неподвижным,*

$$0\varphi = 0,$$

а образом вектора, противоположного для данного вектора $\boldsymbol{a}$*, служит вектор, противоположный для образа вектора* $\boldsymbol{a}$*,*

$$(-\boldsymbol{a})\varphi = -\boldsymbol{a}\varphi.$$

В самом деле, если $\boldsymbol{b}$ — произвольный вектор, то, ввиду (4.15),

$$0\varphi = (0 \cdot \boldsymbol{b})\varphi = 0 \cdot (\boldsymbol{b}\varphi) = 0.$$

С другой стороны,

$$(\boldsymbol{a})\varphi = [(-1)\,\boldsymbol{a}]\varphi = (-1)(\boldsymbol{a}\varphi) = -\boldsymbol{a}\varphi.$$

Понятие линейного преобразования линейного пространства возникло как обобщение известного из курса аналитической геометрии понятия аффинного преобразования плоскости или трехмерного пространства; действительно, условия (4.14) и (4.15) для аффинных преобразований выполняются. Эти условия выполняются и для проекций векторов на плоскости или в трехмерном пространстве на некоторую прямую (или на некоторую плоскость). Таким образом, например, в двумерном линейном пространстве векторов-отрезков, которые выходят из начала координат плоскости, преобразование, которое переводит всякий вектор в его проекцию на некоторую ось, проходящую через начало координат, будет линейным преобразованием.

Примерами линейных преобразований в произвольном пространстве M_n служат *тождественное преобразование* ε, которое оставляет всякий вектор $\boldsymbol{a}$ на месте,

$$\boldsymbol{a}\varepsilon = \boldsymbol{a},$$

и *нулевое преобразование* ω, отображающее всякий вектор $\boldsymbol{a}$ в нуль,

$$\boldsymbol{a}\omega = 0.$$

Сделаем некоторый обзор всех линейных преобразований линейного пространства M_n. Пусть

$$\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n \qquad (4.17)$$

— базис этого пространства; как и раньше, базис (4.17), расположенный в столбец, будем обозначать через $\boldsymbol{e}$. Так как всякий вектор $\boldsymbol{a}$ пространства M_n однозначно представляется в виде линейной комбинации векторов базиса (4.17), то, ввиду (4.16), образ вектора $\boldsymbol{a}$ с теми же коэффициентами выражается через образы векторов (4.17). Другими словами, *всякое линейное преобразование* φ *пространства* M_n *однозначно определяется заданием образов*

$$\boldsymbol{e}_1\varphi, \boldsymbol{e}_2\varphi, \ldots, \boldsymbol{e}_n\varphi$$

всех векторов фиксированного базиса (4.17).

Какова бы не была упорядоченая система из n векторов пространства M_n,

$$\boldsymbol{c}_1, \boldsymbol{c}_2, ,.., \boldsymbol{c}_n, \tag{4.18}$$

существует, притом единственное, такое линейное преобразование φ *этого пространства, что* (4.18) *служит системой образов векторов базиса* (4.17) *при этом преобразовании,*

$$\boldsymbol{e}_i\varphi = \boldsymbol{c}_i, \quad и{=}1, 2.....г. \tag{4.19}$$

Единственность преобразования φ уже была доказана выше и нужно доказать лишь его существование. Определим преобразование φ следующим образом: если $\boldsymbol{a}$ — произвольный вектор пространства и

$$\boldsymbol{a} = \sum_{i=1}^{n} \alpha_i \ \boldsymbol{e}_i$$

- его запись в базисе (4.17), то положим

$$\boldsymbol{a}\varphi = \sum_{i=1}^{n} \alpha_i \ \boldsymbol{c}_i. \tag{4.20}$$

Докажем линейность этого преобразования. Если

$$\boldsymbol{b} = \sum_{i=1}^{n} \beta_i \ \boldsymbol{e}_i$$

- любой другой вектор пространства, то

$$(\boldsymbol{a}+\boldsymbol{b})\varphi = \left[\sum_{i=1}^{n} (\alpha_i + \beta_i)\boldsymbol{e}_i\right]\varphi = \sum_{i=1}^{n} (\alpha_i + \beta_i)\boldsymbol{c}_i =$$

$$= \sum_{i=1}^{n} \alpha_i \boldsymbol{c}_i + \sum_{i=1}^{n} \beta_i \boldsymbol{c}_i = \boldsymbol{a}\varphi + \boldsymbol{b}\varphi.$$

Если же γ - любое число, то

$$(\gamma\boldsymbol{a})\varphi = \left[\sum_{i=1}^{n} (\gamma\alpha_i)\boldsymbol{e}_i\right]\varphi = \sum_{i=1}^{n} (\gamma\alpha_i)\boldsymbol{c}_i = \gamma\sum_{i=1}^{n} \alpha_i \boldsymbol{c}_i = \gamma(\boldsymbol{a}\varphi)$$

Что же касается справедливости равенств (4.19), то она вытекает из определения (4.20) преобразования φ, так как все координаты вектора $\boldsymbol{e}_i$ в базисе (4.17) равны нулю, кроме i-й координаты, которая равна единице.

Нами установлено, следовательно, взаимно-однозначное соответствие между всеми линейными преобразованиями линейного пространства M_n и всеми упорядоченными системами (4.20) *из n векторов этого пространства.*

Всякий вектор $\boldsymbol{c}_i$ - обладает, однако, определенной записью в базисе (4.17),

$$\boldsymbol{c}_i = \sum_{j=1}^{n} \alpha_{ij} \boldsymbol{e}_j, \quad i=1,2,...,n. \qquad (4.21)$$

Из координат вектора $\boldsymbol{c}_i$ в базисе (4.17) можно составить квадратную матрицу

$$A = (\alpha_{ij}), \qquad (4.22)$$

беря в качестве ее i-й строки строку координат вектора $\boldsymbol{c}_i$, $i = 1, 2, ..., n$. Так как система (4.18) была произвольной, то матрица A будет произвольной квадратной матрицей порядка n с действительными элементами.

Мы имеем, таким образом, взаимно однозначное соответствие между всеми линейными преобразованиями пространства M_n *и всеми квадратными матрицами порядка* n*;* это соответствие зависит, конечно, от выбора базиса (4.17).

Будем говорить, что матрица A *задает* линейное преобразование φ в базисе (4.17), или, короче, что A есть *матрица линейного преобразования* φ в базисе (4.17). Если через $\boldsymbol{e}\varphi$ мы обозначим столбец, составленный из образов векторов базиса (4.17), то из (4.19), (4.21) и (4.22) вытекает следующее матричное равенство, которое полностью описывает связи, существующие между линейным преобразованием φ, базисом $\boldsymbol{e}$ и матрицей A, задающей это линейное преобразование в этом базисе:

$$\boldsymbol{e}\varphi = A\boldsymbol{e}. \qquad (4.23)$$

Покажем, как, зная матрицу A линейного преобразования φ в базисе (4.17), по координатам вектора $\boldsymbol{a}$ в этом базисе найти координаты его образа $\boldsymbol{a}\varphi$. Если

$$\boldsymbol{a} = \sum_{i=1}^{n} \alpha_i \boldsymbol{e}_i,$$

то

$$\boldsymbol{a}\varphi = \sum_{i=1}^{n} \alpha_i (\boldsymbol{e}_i\varphi),$$

что равносильно матричному равенству

$$\boldsymbol{a}\varphi = (\alpha_1, \alpha_2, \ldots, \alpha_г)(\boldsymbol{e}\varphi).$$

Используя (4.23) и учитывая то, что ассоциативность умножения матриц легко проверяется и в том случае, когда одна из матриц является столбцом, составленным из векторов, мы получаем:

$$\boldsymbol{a}\varphi = [(\alpha_1, \alpha_2, \ldots, \alpha_г)A]\boldsymbol{e}.$$

Отсюда следует, что *строка координат вектора* $\boldsymbol{a}\varphi$ *равна строке координат вектора* $\boldsymbol{a}$*, умноженной справа на матрицу* A *линейного преобразования* φ*, все в базисе* (4.17).

Пример. Пусть в базисе $\boldsymbol{e}_1$ $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ трехмерного линейного пространства линейное преобразование φ задается матрицей

$$\boldsymbol{A} = \begin{pmatrix} -2 & 1 & 0 \\ 1 & 3 & 2 \\ 0 & -4 & 1 \end{pmatrix}$$

Если

$$\boldsymbol{a} = 5\boldsymbol{e}_1 + \boldsymbol{e}_2 - 2\boldsymbol{e}_3,$$

то

$$(5, 1, -2)\begin{pmatrix} -2 & 1 & 0 \\ 1 & 3 & 2 \\ 0 & -4 & 1 \end{pmatrix} = (-9, 16, 0),$$

т.е.

$$\boldsymbol{a}\varphi = -9\boldsymbol{e}_1 + 16\boldsymbol{e}_2.$$

Связь между матрицами линейного преобразования в разных базисах. Матрица, которая задает линейное преобразование, зависит от выбора базиса. Покажем, какова связь между матрицами, задающими в разных базисах одно и то же линейное преобразование.

Пусть даны базисы $\boldsymbol{e}$ и $\boldsymbol{e}'$ с матрицей перехода T,

$$\boldsymbol{e}' = To, \tag{4.24}$$

и пусть линейное преобразование φ задается в этих базисах соответственно матрицами A и A',

$$\boldsymbol{e}\varphi = A\boldsymbol{e}, \quad \boldsymbol{e}'\varphi = A'\boldsymbol{e}'. \tag{4.25}$$

Второе из равенств (4.25) приводит, ввиду (4.24), к равенству

$$(T\boldsymbol{e})\varphi = A'(T\boldsymbol{e}).$$

Однако

$$(T\boldsymbol{e})\varphi = T(\boldsymbol{e}\varphi).$$

Действительно, если $(\tau_{i1}, \tau_{i2}, ..., \tau_{in})$ — i-я строка матрицы T, то

$$(\tau_{i1}\boldsymbol{e}_1 + \tau_{i2}\boldsymbol{e}_2 + ... + \tau_{in}\boldsymbol{e}_n)\varphi = \tau_{i1}(\boldsymbol{e}_1\varphi) + \tau_{i2}(\boldsymbol{e}_2\varphi) + ... + \tau_{in}(\boldsymbol{e}_n\varphi).$$

Таким образом, ввиду (4.25),

$$(T\boldsymbol{e})\varphi = T(\boldsymbol{e}\varphi) = T(A\boldsymbol{e}) = (TA)\boldsymbol{e},$$

$$A'(T\boldsymbol{e}) = (A'T)\boldsymbol{e},$$

т.е.

$$(TA)\boldsymbol{e} = (A'T)\boldsymbol{e}.$$

Если хотя бы для одного i, $1 \le i \le n$ i-я строка матрицы TA будет отлична от i-й строки матрицы $A'T$, то две различные линейные комбинации векторов $\boldsymbol{e}_1, \boldsymbol{e}_2, .., \boldsymbol{e}_n$ окажутся равными друг другу, что противоречит линейной независимости базиса $\boldsymbol{e}$. Таким образом,

$$TA = A'T,$$

откуда, через невырожденость матрицы перехода T,

$$A' = TA\,T^{-1}, \quad A = T^{-1}A'T. \qquad (4.26)$$

Заметим, что квадратные матрицы B и C называются *подобными,* если они связаны равенством

$$C = Q^{-1}BQ,$$

где Q — некоторая невырожденная матрица. При этом говорят, что матрица C получена из матрицы B *трансформированием* матрицы Q.

Доказанные выше равенства (4.26) можно сформулировать, таким образом, в виде следующей теоремы:

Матрицы, задающие одно и то же самое линейное преобразование в разных базисах, подобны между собой. При этом матрица линейного преобразования φ *в базисе* ***e'*** *получается трансформированием матрицы этого преобразования в базисе* ***e*** *матрицей перехода от базиса* ***e'*** *к базису* ***e***.

Подчеркнем, что если матрица A задает линейное преобразование φ в базисе ***e***, то любая матрица B, подобная матрице A,

$$B = Q^{-1}AQ,$$

также задает преобразование φ в некотором базисе, а именно в базисе, получающийся из базиса ***e*** при помощи матрицы перехода Q^{-1}.

Операции над линейными преобразованиями.

Сопоставляя каждому линейному преобразованию пространства M_n его матрицу в фиксированном базисе, мы получаем, как доказано, взаимно-однозначное соответствие между всеми линейными преобразованиями и всеми квадратными матрицами порядка n. Естественно ожидать, что операциям сложения и умножение матриц, а также умножение матрицы на число, будут отвечать аналогичные операции над линейными преобразованиями.

Пусть в пространстве M_n даны линейные преобразования φ и ψ. Назовем *суммой* этих преобразований преобразованиу φ+ψ, определяемое равенством

$$\boldsymbol{a}(\varphi+\psi)=\boldsymbol{a}\varphi + \boldsymbol{a}\psi; \qquad (4.27)$$

оно переводит, следовательно, любой вектор ***a*** в сумму его образов при преобразованиях φ и ψ.

Преобразование φ+ψ *является линейным.* Действительно, для любых векторов ***a*** и ***b*** и любого числа α

$$(\boldsymbol{a}+\boldsymbol{b})(\varphi+\psi)=(\boldsymbol{a}+\boldsymbol{b})\varphi+(\boldsymbol{a}+\boldsymbol{b})\psi=\boldsymbol{a}\varphi+\boldsymbol{b}\varphi+\boldsymbol{a}\psi+b\psi=\boldsymbol{a}(\varphi+\psi)+\boldsymbol{b}(\varphi+\psi);$$

$$(\alpha\boldsymbol{a})(\varphi+\psi)=(\alpha\boldsymbol{a})\varphi+(\alpha\boldsymbol{a})\psi=\alpha(\boldsymbol{a}\varphi)+\alpha(\boldsymbol{a}\psi)=\alpha(\boldsymbol{a}\varphi+\boldsymbol{a}\psi)=\alpha[\boldsymbol{a}(\varphi+\psi)].$$

С другой стороны, назовем *произведением* линейных преобразований φ и ψ преобразование φψ, которое определяется равенством

$$\boldsymbol{a}(\varphi\psi)=(\boldsymbol{a}\varphi)\psi, \qquad (4.28)$$

т.е., получающееся в результате последовательного выполнения преобразований φ и ψ.

Преобразование φψ *является линейным:*

$(\boldsymbol{a}+\boldsymbol{b})(\varphi\psi)=[(\boldsymbol{a}+\boldsymbol{b})\varphi]\psi=(\boldsymbol{a}\varphi+\boldsymbol{b}\varphi)\psi=(\boldsymbol{a}\varphi)\psi+(\boldsymbol{b}\varphi)\psi=\boldsymbol{a}(\varphi\psi)+\boldsymbol{b}(\varphi\,\psi)$;

$(\alpha\boldsymbol{a})(\varphi\psi)=[(\alpha\boldsymbol{a})\varphi]\psi=[\alpha(\boldsymbol{a}\varphi)]\psi=\alpha[(\boldsymbol{a}\varphi)\psi]=\alpha[\boldsymbol{a}(\varphi\psi)]$.

Назовем, наконец, *произведением линейного преобразования* φ *на число* к преобразование кφ, определяемое равенством

$$\boldsymbol{a}(\kappa\varphi)=\kappa(\boldsymbol{a}\varphi); \qquad (4.29)$$

образа при преобразовании φ всех векторов умножаются, следовательно, на число к.

Преобразование кφ *является линейным:*

$(\boldsymbol{a}+\boldsymbol{b})(\kappa\varphi)=\kappa[(\boldsymbol{a}+\boldsymbol{b})\varphi]=\kappa(\boldsymbol{a}\varphi+\boldsymbol{b}\varphi)=\kappa(\boldsymbol{a}\varphi)+\kappa(\boldsymbol{b}\varphi)=\boldsymbol{a}(\kappa\varphi)+\boldsymbol{b}(\kappa\varphi)$;

$(\alpha\boldsymbol{a})(\kappa\varphi)=\kappa[(\alpha\boldsymbol{a})\varphi]=\kappa[\alpha(\boldsymbol{a}\varphi)]=\alpha[\kappa(\boldsymbol{a}\varphi)]=\alpha[\boldsymbol{a}(\kappa\varphi)]$.

Пусть в базисе $\boldsymbol{e_1,e_2,\dots,e_n}$ преобразования φ и ψ задаются соответственно матрицами $A=(\alpha_{ij})$ и $B=(\beta_{ij})$,

$$\boldsymbol{e}\varphi=A\boldsymbol{e},\quad \boldsymbol{e}\psi=B\boldsymbol{e}.$$

Тогда, ввиду (4.27),

$$\boldsymbol{e}_i(\varphi+\psi)=\boldsymbol{e}_i\varphi+\boldsymbol{e}_i\psi=\sum_{j=1}^{n}\alpha_{ij}\boldsymbol{e}_j+\sum_{j=1}^{n}\beta_{ij}\boldsymbol{e}_j=\sum_{j=1}^{n}(\alpha_{ij}+\beta_{ij})\boldsymbol{e}_j,$$

т.е.

$$\boldsymbol{e}(\varphi+\psi)=(A+B)\boldsymbol{e}.$$

Таким образом, *матрица суммы линейных преобразований в любом базисе равна сумме матриц этих преобразований в том же базисе.*

С другой стороны, ввиду (4.28),

$$\boldsymbol{e}_i(\varphi\psi)=(\boldsymbol{e}_i\varphi)\psi=\left(\sum_{j=1}^{n}\alpha_{ij}\boldsymbol{e}_j\right)\psi=\sum_{j=1}^{n}\alpha_{ij}(\boldsymbol{e}_j\psi)=\sum_{j=1}^{n}\alpha_{ij}\left(\sum_{k=1}^{n}\beta_{ik}\boldsymbol{e}_k\right)=$$

$$=\sum_{k=1}^{n}\left(\sum_{j=1}^{n}\alpha_{ij}\beta_{ik}\right)\boldsymbol{e}_k,$$

т.е.

$$\boldsymbol{e}(\varphi\psi)=(AB)\boldsymbol{e}.$$

Другими словами, *матрица произведения линейных преобразований в любом базисе равна произведению матриц этих преобразований в том же базисе.*

Наконец, ввиду (4.29),

$$\boldsymbol{e}_i(\kappa\varphi)=\kappa(\boldsymbol{e}_i\varphi)=\kappa\sum_{j=1}^{n}\alpha_{ij}\boldsymbol{e}_j=\sum_{j=1}^{n}(\kappa\alpha_{ij})\boldsymbol{e}_j,$$

т.е.

$$\boldsymbol{e}(\kappa\varphi) = (\kappa B)\boldsymbol{e}.$$

Следовательно, *матрица, задающая в некотором базисе произведения линейного преобразования* φ *на число* к, *равна произведению матрицы самого преобразования* φ *в этом базисе на число* к.

Из полученных результатов следует, что операции над линейными преобразованиями обладают теми же свойствами, что и операции над матрицами. Так, сложение линейных преобразований коммутативно и ассоциативно, а умножение ассоциативно, но при $n>1$ не коммутативно. Для линейных преобразований существует однозначное вычитание. Отметим также, что *тождественное преобразование* ε *играет среди линейных преобразований роль единицы, а нулевое преобразование* ω - *роль нуля.* Действительно, в любом базисе преобразование ε задается единичной матрицей, а преобразование ω - нулевой матрицей.

4.4. Линейные подпространства

Подмножество L линейного пространства M называется *линейным подпространством* этого пространства, если оно само является линейным пространством по отношению определенным в M операциям сложения векторов и умножения вектора на число. Так, в трехмерном евклидовом пространстве совокупность векторов, которые выходят из начала координат и лежат на некоторой плоскости (или некоторой прямой), проходящей через начало координат, будет линейным подпространством.

Для того чтобы непустое подмножество L пространства M было его линейным подпространством, достаточно выполнения следующих требований:

1. *Если векторы* $\boldsymbol{a}$ *и* $\boldsymbol{b}$ *принадлежат к L, то в L содержится и вектор* $\boldsymbol{a}+\boldsymbol{b}$.
2. *Если вектор* $\boldsymbol{a}$ *принадлежит к L, то в L содержится и вектор* $\alpha\boldsymbol{a}$ *при любом значении числа* α.

Действительно, ввиду условия 2, множество L содержит нулевой вектор: если вектор $\boldsymbol{a}$ принадлежит к L, то L содержит и вектор $0\cdot\boldsymbol{a} = 0$. Далее, L вместе со всяким своим вектором $\boldsymbol{a}$ содержит, снова ввиду свойства 2, и противоположный ему вектор $-\boldsymbol{a} = (-1)\cdot\boldsymbol{a}$, а потому ввиду свойства 1 к L принадлежит и разность любых двух векторов из L. Что же касается всех остальных требований, которые входят в определение линейного пространства, то они, выполняясь в *M,* будут выполняться и в *L*.

Примерами линейных подпространств пространства *M* могут служить само пространство *M,* а также множество, которое состоит из одного нулевого вектора, — так называемое *нулевое подпространство.* Рассмотрим следующий пример: берем в просторные *M* любую конечную систему векторов

$$\boldsymbol{a}_1,\ \boldsymbol{a}_2,\dots,\boldsymbol{a}_r \tag{4.30}$$

и обозначаем через *L* множество всех тех векторов, которые являются линейными комбинациями векторов (4.30). Докажем, что *L* будет линейным подпространством. В самом деле, если

$$\boldsymbol{b} = \alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2 + \dots + \alpha_r\boldsymbol{a}_r,\quad \boldsymbol{c} = \beta_1\boldsymbol{a}_1 + \beta_2\boldsymbol{a}_2 + \dots + \beta_r\boldsymbol{a}_r$$

то

$$\boldsymbol{b} + \boldsymbol{c} = (\alpha_1 + \beta_1)\,\boldsymbol{a}_1 + (\alpha_2 + \beta_2)\,\boldsymbol{a}_2 + \dots + (\alpha_r + \beta_r)\,\boldsymbol{a}_r,$$

т.е. вектор $\boldsymbol{b} + \boldsymbol{c}$ принадлежит к *L*; к *L* принадлежит и вектор

$$\gamma\boldsymbol{b} = (\gamma\alpha_1)\boldsymbol{a}_1 + (\gamma\alpha_2)\boldsymbol{a}_2 + \dots + (\gamma\alpha_r)\boldsymbol{a}_r,$$

при любом числе γ.

Говорят, что это линейное подпространство *L порождено* системой векторов (4.30); к *L* принадлежат, в частности, сами векторы (4.30).

Впрочем, *всякое линейное подпространство конечномерного линейного пространства порождается конечной системой векторов,* так как если оно не является нулевым, то обладает даже конечным базисом. Размерность линейного подпространства *L* не больше размерности *n* самого просторную M_n, причем равна *n* лишь при $L=M_n$. Размерностью нулевого подпространства следует считать число 0.

Для всякого k, $0<k<n$, *в просторные* M_n *существуют линейные подпространства размерности k* - достаточно взять подпространство, порожденное любой системой из *k* линейно независимых векторов.

Пусть в пространстве *M* даны линейные подпространства L_1 и L_2. Совокупность L_0 векторов, которые принадлежат как к L_1, так и к L_2, будет, как легко проверить, линейным подпространством; оно называется *пересечением* подпространств L_1 и L_2. С другой стороны, линейным подпространством будет и *сумма* $\overline{L}$ подпространств L_1 и L_2, т.е. совокупность всех тех векторов из *M,* которые можно представить в виде суммы двух слагаемых, одного из L_1, другого из L_2. Если размерности подпространств L_1, L_2, L_0 и $\overline{L}$ суть, соответственно, d_1 d_2, d_0 и $\overline{d}$, то имеет место следующая формула:

$$\overline{d} = d_1 + d_2 - d_0, \tag{4.31}$$

т.е. *размерность суммы двух подпространств равна сумме размерностей этих подпространств минус размерность их пересечения.*

Для доказательства берем произвольный базис

$$\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{\boldsymbol{d}_0} \qquad (4.32)$$

подпространства L_0 и дополняем его к базису

$$\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{\boldsymbol{d}_0}, \ \boldsymbol{b}_{\boldsymbol{d}_0+1}, \ldots, \boldsymbol{b}_{\boldsymbol{d}_1} \qquad (4.33)$$

подпространства L_1 и к базису

$$\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{\boldsymbol{d}_0}, \ \boldsymbol{c}_{\boldsymbol{d}_0+1}, \ldots, \boldsymbol{c}_{\boldsymbol{d}_1} \qquad (4.34)$$

подпространства L_2. Легко видеть, используя определение подпространства $\overline{L}$, что это подпространство порождается системой векторов

$$\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{\boldsymbol{d}_0}, \ \boldsymbol{b}_{\boldsymbol{d}_0+1}, \ldots, \boldsymbol{b}_{\boldsymbol{d}_1}, \boldsymbol{c}_{\boldsymbol{d}_0+1}, \ldots, \boldsymbol{c}_{\boldsymbol{d}_1} \qquad (4.35)$$

Формула (4.31) будет, следовательно, доказана, если мы докажем линейную независимость системы (4.35).

Пусть имеет место равенство

$$\alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \ldots + \alpha_{\boldsymbol{d}_0} \boldsymbol{a}_{\boldsymbol{d}_0} + \beta_{\boldsymbol{d}_0+1} \boldsymbol{b}_{\boldsymbol{d}_0+1} + \ldots + \beta_{\boldsymbol{d}_1} \boldsymbol{b}_{\boldsymbol{d}_1} +$$

$$+ \gamma_{\boldsymbol{d}_0+1} \boldsymbol{c}_{\boldsymbol{d}_0+1} + \ldots + \gamma_{\boldsymbol{d}_1} \boldsymbol{c}_{\boldsymbol{d}_1} = 0$$

с некоторыми числовыми коэффициентами. Тогда

$$\boldsymbol{d} = \alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \ldots + \alpha_{\boldsymbol{d}_0} \boldsymbol{a}_{\boldsymbol{d}_0} + \beta_{\boldsymbol{d}_0+1} \boldsymbol{b}_{\boldsymbol{d}_0+1}, \ldots, \beta_{\boldsymbol{d}_1} \boldsymbol{b}_{\boldsymbol{d}_1} =$$

$$= -\gamma_{\boldsymbol{d}_0+1} \boldsymbol{c}_{\boldsymbol{d}_0+1} - \ldots - \gamma_{\boldsymbol{d}_1} \boldsymbol{c}_{\boldsymbol{d}_1}. \qquad (4.36)$$

Левая часть этого равенства содержится в L_1, правая — в L_2, поэтому вектор $\boldsymbol{d}$, который равен как левой, так и правой части этого равенства, принадлежит к L_0 и, следовательно, линейно выражается через базис (4.32). Правая часть равенства (4.32) показывает, однако, что вектор $\boldsymbol{d}$ линейно выражается и через векторы $\boldsymbol{c}_{\boldsymbol{d}_0+1}, \ldots, \boldsymbol{c}_{\boldsymbol{d}_1}$. Отсюда, ввиду линейной независимости системы (4.34), вытекает, что все коэффициенты $\gamma_{\boldsymbol{d}_0+1}, \ldots, \gamma_{\boldsymbol{d}_1}$ равны нулю, т.е. $\boldsymbol{d}=0$, а тогда, ввиду линейной независимости системы (4.33), все коэффициенты

$$\alpha_1, \ldots, \alpha_{\boldsymbol{d}_0}, \beta_{\boldsymbol{d}_0+1}, \ldots, \beta_{\boldsymbol{d}_1}$$

также равны нулю. Этим доказана линейная независимость системы (4.35).

Область значений и ядро линейного преобразования. Пусть в линейном пространстве M_n задано линейное преобразование φ. Если L — любое линейное подпространство пространства M_n, то совокупность $L\varphi$ образов всех векторов из L при преобразовании φ также будет линейным подпространством, как вытекает из определений линейного подпространства и линейного преобразования. В частности, линейным подпространством будет и совокупность $L_n\varphi$ образов всех векторов пространства M_n; она называется *областью значений* преобразования φ.

Найдем размерность области значений. Для этого заметим, что, так как все матрицы, которые задают преобразование φ в разных базисх, подобны между собой, то все они имеют тот самый ранг. Это число можно назвать, следовательно, *рангом* линейного преобразования φ.

Размерность области значений линейного преобразования φ *равна рангу этого преобразования.*

В самом деле, пусть φ задается в базисе $\boldsymbol{e}_l, \boldsymbol{e}_2, ..., \boldsymbol{e}_n$ матрицей A. Подпространство $M_n\varphi$ порождается векторами

$$\boldsymbol{E}_1\varphi, \boldsymbol{e}_2\varphi, ..., \boldsymbol{e}_n\varphi \tag{4.37}$$

и поэтому базисом подпространства $M_n\varphi$ будет служить, в частности, любая максимальная линейно независимая подсистема системы (4.37). Однако максимальное число линейно независимых векторов в системе (4.37) равно максимальному числу линейно независимых строк матрицы A, т.е. равно рангу этой матрицы. Теорема доказана.

Мы знаем, что при линейном преобразовании φ нулевой вектор переходит у самого себя. Совокупность $N(\varphi)$ всех векторов пространства M_n, которые отображаются при φ в нулевой вектор, будет, следовательно, непустой и является, очевидно, линейным подпространством. Это подпространство называется *ядром* преобразования φ, а его размерность — *дефектом* этого преобразования.

Для любого линейного преобразования φ *пространства* M_n *сумма ранга и дефекта этого преобразования равна размерности* n *всего пространства.*

Действительно, если r — ранг преобразования φ, то подпространство $M_n\varphi$ обладает базисом из r векторов

$$\boldsymbol{a}_1, \boldsymbol{a}_2,...,\boldsymbol{a}_r. \tag{4.38}$$

В пространстве M_n можно выбрать такие векторы

$$\boldsymbol{b}_1, \boldsymbol{b}_2,...,\boldsymbol{b}_r, \tag{4.39}$$

что

$$\boldsymbol{b}_i\varphi = \boldsymbol{a}_i\ ,\ i=1,\ 2,\ ..\ .,\ r;$$

выбор векторов (4.39) не является однозначным. Если бы некоторая нетривиальная линейная комбинация векторов (4.39) отображалась преобразованием φ в нуль, в частности, если бы векторы (4.39) были линейно зависимыми, то векторы (4.38) оказались бы сами линейно зависимыми против предположения. Поэтому линейное подпространство L, порожденное векторами (4.39), имеет размерность r, а его пересечение с подпространством $N(\varphi)$ равно нулю.

С другой стороны, сумма подпространств L и $N(\varphi)$ совпадает со всем пространством M_n. Действительно, если $\boldsymbol{c}$ - любой вектор пространства, то вектор $\boldsymbol{d}=\boldsymbol{c}\varphi$ принадлежит к подпространству $M_n\varphi$. Тогда в подпространстве L найдется такой вектор $\boldsymbol{b}$, что

$$\boldsymbol{b}\varphi=\boldsymbol{d}$$

— вектор $\boldsymbol{b}$ записывается через систему (4.39) с теми же коэффициентами, с которыми вектор $\boldsymbol{d}$ записывается через базис (4.38). Отсюда

$$\boldsymbol{c} = \boldsymbol{b} + (\boldsymbol{c}-\boldsymbol{b}),$$

причем вектор $\boldsymbol{c} — \boldsymbol{b}$ содержится в подпространстве $N(\varphi)$, так как

$$(\boldsymbol{c} — \boldsymbol{b})\varphi =\boldsymbol{c}\varphi — \boldsymbol{b}\varphi=\boldsymbol{d}—\boldsymbol{d}=0.$$

Из полученных результатов и доказанной выше формулы (4.31) вытекает утверждение теоремы.

Невырожденные линейные преобразования. Линейное преобразование φ линейного пространства M_n называется *невырожденым,* если оно удовлетворяет каждому из следующих условий, равносильность которых вытекает из доказанных выше теорем:

1. Ранг преобразования φ равен n.
2. Областью значений преобразования φ служит все пространство M_n.
3. Дефект преобразования φ равен нулю.

Для невырожденых линейных преобразований можно указать также много других определений, равносильных указанным выше, в частности определения 4 - 6.

4. Различные векторы пространства M_n имеют при преобразовании φ различные образы.

Действительно, если преобразование φ имеет свойство 4, то ядро этого преобразования состоит лишь из нулевого вектора, т.е. выполняется и свойство 3. Если же векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ таковы, что $\boldsymbol{a}\neq\boldsymbol{b}$, но $\boldsymbol{a}\varphi=\boldsymbol{b}\varphi$, то $(\boldsymbol{a} — \boldsymbol{b})\neq 0$, но $(\boldsymbol{a} — \boldsymbol{b})\varphi=0$, т.е. свойство 3 не выполняется.

Из 2 и 4 вытекает

5. Преобразование φ есть взаимно-однозначным отображением пространства M_n на все это пространство.

Из 5 следует, что для невырожденного линейного преобразования φ существует *обратное преобразование* φ^{-1}, переводящее всякий вектор $\boldsymbol{a}\varphi$ в вектор $\boldsymbol{a}$,

$$(\boldsymbol{a}\varphi)\varphi^{-1}=\boldsymbol{a}.$$

Преобразование φ^{-1} *будет линейным,* так как

$$(\boldsymbol{a}\varphi + \boldsymbol{b}\varphi)\varphi^{-1}=[(\boldsymbol{a} + \boldsymbol{b})\varphi]\,\varphi^{-1}= \boldsymbol{a} + \boldsymbol{b},$$

$$[\alpha(\boldsymbol{a}\,\varphi)]\varphi^{-1} =[(\alpha\boldsymbol{a})\varphi]\varphi^{-1} = \alpha\boldsymbol{a}.$$

Из определения преобразования φ^{-1} вытекает, что

$$\varphi\varphi^{-1} = \varphi^{-1}\varphi = \boldsymbol{e}; \qquad (4.40)$$

равенства (4.40) могут сами рассматриваться как определение обратного преобразования. Отсюда и из последних результатов предыдущего раздела следует, что *если невырожденное линейное преобразование* φ *задается в некотором базисе матрицей A, невырожденной ввиду свойство* 1, *то преобразование* φ^{-1} *задается в этом базисе матрицей* A^{-1}.

Мы приходим, таким образом, к следующему определению невырожденного линейного преобразования:

6. Для преобразования φ существует обратное линейное преобразование φ^{-1}.

4.5. Характеристические корни и собственные значения

Пусть $A=(\alpha_{ij})$ — квадратная матрица порядка n с действительными элементами. Пусть, с другой стороны, λ - некоторое неизвестное. Тогда матрица $A—\lambda E$, где E-единичная матрица порядка n, называется *характеристической матрицей* матрицы A. Так как в матрице λE по главной диагонали стоят λ, а все другие элементы равны нулю, то

$$A—\lambda E=\begin{pmatrix} \alpha_{11}-\lambda & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22}-\lambda & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn}-\lambda \end{pmatrix}$$

Определитель матрицы $A—\lambda E$ будет многочленом от λ, притом степени n. В самом деле, произведение элементов, которые стоят на главной диагонали, будет многочленом от λ со старшим членом $(—1)^n\lambda^n$, все же остальные члены определителя не содержат по меньшей мере двух из числа элементов, которые стоят на главной диагонали, и поэтому их степень относительно λ не превосходит $n—2$. Коэффициенты этого многочлена можно было бы легко найти. Так, коэффициент при λ^{-1} равен

$$(—1)^{n—1}(\alpha_{12}+\alpha_{22} + \dots + \alpha_{nn}),$$

а свободный член совпадает с определителем матрицы A.

Многочлен n-й степени $|A—\lambda E|$ называется *характеристическим многочленом* матрицы A, а его корни, которые могут быть как действительными, так и комплексными, называются *характеристическими корнями* этой матрицы.

Подобные матрицы обладают одинаковыми характеристическими многочленами и, следовательно, одинаковыми характеристическими корнями.

Пусть, в самом деле,

$$B=Q^{-1}AQ.$$

Тогда, учитывая, что матрица λE перестановочна с матрицей Q, а $|Q^{-1}|=|Q|^{-1}$, получаем:

$$|B-\lambda E|=|Q^{-1}AQ-\lambda E|=|Q^{-1}(A-\lambda E)Q| =$$
$$=|Q|^{-1}\cdot|A-\lambda E|\cdot|Q|=|A-\lambda E|,$$

что и требовалось доказать.

Из этого результата вытекает, ввиду доказанной в п. 4.4 теоремы о связи между матрицами, задающими линейное преобразование в разных базисах, что *хотя линейное преобразование* φ *может задаваться в разных базисах разными матрицами, однако все эти матрицы имеют один и тот же набор характеристических корней.* Поэтому эти корни можно называть *характеристическими корнями самого преобразования* φ. Весь набор этих характеристических корней, причем каждый корень берется с той кратностью, какую он имеет в характеристическому многочлене, называется *спектром* линейного преобразования φ.

Характеристические корни играют при изучении линейных преобразований очень большую роль. Одно из применений характеристических корней мы сейчас укажем.

Пусть в действительном линейном пространстве M_n задано линейное преобразование φ. Если вектор $\boldsymbol{b}$, отличный от нуля, переводится преобразованием φ в вектор, пропорциональный самому $\boldsymbol{b}$,

$$\boldsymbol{b}\varphi=\lambda_0\boldsymbol{b}, \qquad (4.41)$$

где λ_0 - некоторое действительное число, то вектор $\boldsymbol{b}$ называется *собственным вектором* преобразования φ, а число λ_0 — *собственным значением* этого преобразования, причем говорят, что собственный вектор $\boldsymbol{b}$ *относится* к собственному значению λ_0.

Заметим, что так как $\boldsymbol{b}\neq 0$, то число λ_0, которое удовлетворяет условию (4.41), определяется для вектора $\boldsymbol{b}$ однозначно. Подчеркнем, далее, что нулевой вектор не считается собственным вектором преобразования φ, хотя он удовлетворяет условию (4.41), притом для любого λ_0.

Вращение евклидовой плоскости вокруг начала координат на угол, который не является кратным π, является примером линейного преобразования, которое не имеет собственных векторов. Примером другого крайнего случая является растяжение плоскости, при котором все векторы, которые выходят из начала координат, растягиваются, скажем, в пять раз. Это будет линейное преобразование, причем все ненулевые векторы плоскости будут для него собственными; все они относятся к собственному значению 5.

Действительные характеристические корни линейного преобразования φ, если они существуют, и только они служат собственными значениями этого преобразования.

Пусть, в самом деле, преобразование φ имеет в базисе $\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n$ матрицу $A=(\alpha_{ij})$ и пусть вектор

$$\boldsymbol{b}=\sum_{i=1}^{n}\beta_i \boldsymbol{e}_i$$

является собственным вектором преобразования φ,

$$\boldsymbol{b}\varphi=\lambda_0\boldsymbol{b}. \tag{4.42}$$

Как доказано в п. 4.4,

$$\boldsymbol{b}\varphi=[(\beta_1, \beta_2, ..., \beta_n)A]\boldsymbol{e}. \tag{4.43}$$

Равенства (4.42) и (4.43) приводят к системе равенств

$$\begin{gathered}\beta_1\alpha_{11}+\beta_2\alpha_{21}+...+\beta_n\alpha_{n1}=\lambda_0\beta_1,\\ \beta_1\alpha_{12}+\beta_2\alpha_{22}+...+\beta_n\alpha_{n2}=\lambda_0\beta_2,\\ \dots\dots\dots\dots\dots\dots\\ \beta_1\alpha_{1n}+\beta_2\alpha_{2n}+...+\beta_n\alpha_{nn}=\lambda_0\beta_n.\end{gathered} \tag{4.44}$$

Так как $\boldsymbol{b}\neq 0$, то не все числа $\beta_1, \beta_2, ..., \beta_n$ равны нулю. Таким образом, ввиду (4.44), система линейных однородных уравнений

$$\begin{gathered}(\alpha_{11}-\lambda_0)x_1+\alpha_{21}x_2+...+\alpha_{n1}x_n=0,\\ \alpha_{12}x_1+(\alpha_{22}-\lambda_0)x_2+...+\alpha_{n2}x_n=0,\\ \dots\dots\dots\dots\dots\dots\\ \alpha_{1n}x_1+\alpha_{2n}x_2+...+(\alpha_{nn}-\lambda_0)x_n=0.\end{gathered} \tag{4.45}$$

имеет ненулевое решение, а потому ее определитель равен нулю,

$$\begin{vmatrix}\alpha_{11}-\lambda_0\ \alpha_{21}\ ...\ \alpha_{n1}\\ \alpha_{12}\ \alpha_{22}-\lambda_0\ ...\ \alpha_{n2}\\ \dots\dots\dots\dots\dots\\ \alpha_{1n}\ \alpha_{2n}\ ...\ \alpha_{nn}-\lambda_0\end{vmatrix}=0 \tag{4.46}$$

Транспонируя, получаем

$$|A-\lambda_0 E|=0, \tag{4.47}$$

т.е. собственное значение λ_0 на самом деле оказалось характеристическим корнем матрицы A и, следовательно, линейного преобразования φ, притом действительным.

Обратно, пусть λ_0 будет любым действительным характеристическим корнем преобразования φ и, следовательно, матрицы A. Тогда имеет место равенство (4.47), а поэтому и получающееся из него транспонированием равенство (4.47). Отсюда следует, что система линейных однородных уравнений (4.45) обладает ненулевым решением, притом даже действительным, так как все коэффициенты этой системы действительны. Если это решение обозначим через

$$(\beta_1, \beta_2, ..., \beta_r), \tag{4.48}$$

то имеют место равенства (4.44). Обозначим через $\boldsymbol{b}$ вектор пространства M_n, имеющий в базисе $\boldsymbol{e_1}, \boldsymbol{e_2}, ..., \boldsymbol{e_n}$ строку координат (4.48); ясно, что $\boldsymbol{b}\neq 0$. Тогда справедливо равенство (4.43), а из (4.44) и (4.43) следует (4.42). Вектор $\boldsymbol{b}$ оказался, таким образом, собственным вектором преобразования φ, относящимся к собственному значению λ_0. Теорема доказана.

Заметим, что если бы мы рассматривали комплексное линейное пространство, то требование действительности характеристического корня было бы лишним, т.е. была бы доказанная теорема: *характеристические корни линейного преобразования комплексного линейного пространства и только они служат собственными значениями этого преобразования.* Отсюда следует, что *в комплексном линейном пространстве всякое линейное преобразование обладает собственными векторами.*

Возвращаясь к рассмотренному нами действительному случаю, отметим, что совокупность собственных векторов линейного преобразования φ, относящихся к собственному значению λ_0, совпадает с совокупностью ненулевых действительных решений системы линейных однородных уравнений (4.45). Отсюда следует, что *совокупность собственных векторов линейного преобразования* φ, *относящихся к собственному значению* λ_0, *будет, после добавления к ней нулевого вектора, линейным подпространством пространства* M_n. В самом деле, *совокупность (действительных) решений любой системы линейных однородных уравнений от п неизвестных будет линейным подпространством пространства* M_n.

Линейные преобразования с простым спектром. Во многих случаях оказывается необходимым знать, может ли данное линейное преобразование φ иметь в некотором базисе диагональную матрицу.

На самом деле далеко не всякое линейное преобразование может быть задано диагональной матрицей. Необходимые и достаточные условия для этого указаны в книге автора по дискретно-непрерывной математике ”Матрицы”, а сейчас приведем одно достаточное условие.

Докажемм сначала следующие вспомогательные результаты:

Линейное преобразование φ *тогда и только тогда задается в базисе* $\boldsymbol{e_1}, \boldsymbol{e_2}, ..., \boldsymbol{e_n}$ *диагональной матрицей, если все векторы этого базиса являются собственными векторами преобразования* φ.

Действительно, равенство

$$\boldsymbol{e}_i\varphi = \lambda_i \boldsymbol{e}_i$$

равносильно потому, что в i-й строке матрицы, задающей преобразование φ в указанном базисе, равны нулю все элементы, которые стоят вне главной диагонали, а на главной диагонали (т.е. на i-м месте) стоит число λ_i.

Собственные векторы $\boldsymbol{b}_1, \boldsymbol{b}_2, ..., \boldsymbol{b}_k$ *линейного преобразования* φ, *относящиеся к различным собственным значениям, составляют линейно независимую систему.*

Будем доказывать это утверждение индукцией по k, так как при k=1 оно справедливое — один собственный вектор, будучи отличным от нуля, составляет линейно независимую систему.

Пусть

$$\boldsymbol{b}_i\varphi=\lambda_i\boldsymbol{b}_i, \quad i=1, 2, ..., k,$$

и

$$\lambda_i \neq \lambda_j \text{ при } i \neq j.$$

Если существует линейная зависимость

$$\alpha_1\boldsymbol{b}_1 + \alpha_2\boldsymbol{b}_2 + ... + \alpha_k\boldsymbol{b}_k = 0, \qquad (4.49)$$

где, например, $\alpha_1 \neq 0$, то, применяя к обеим частям равенства (4.49) преобразование φ, получим

$$\alpha_1\lambda_1\boldsymbol{b}_1 + \alpha_2\lambda_2\boldsymbol{b}_2 + ... + \alpha_k\lambda_k\boldsymbol{b}_k = 0.$$

Вычитая отсюда равенство (4.56), умноженное на λ_k, получаем

$$\alpha_1(\lambda_1-\lambda_k)\boldsymbol{b}_1 + \alpha_2(\lambda_2-\lambda_k)\boldsymbol{b}_2 + ... + \alpha_{k-1}(\lambda_{k-1}-\lambda_k)\boldsymbol{b}_{k-1} = 0.$$

что дает нетривиальную линейную зависимость между векторами $\boldsymbol{b}_1, \boldsymbol{b}_2, ..., \boldsymbol{b}_{k-1}$, так как $\alpha_1(\lambda_1-\lambda_k) \neq 0$.

Говорят, что линейное преобразование φ действительного линейного пространства M_n имеет *простой спектр,* если все его характеристические корни действительны и различны. Преобразование φ имеет, следовательно, n различных собственных значений, а поэтому, по доказанной теореме, в пространстве M_n существует базис, составленный из собственных векторов этого преобразования. Таким образом, *всякое линейное преобразование с простым спектром может быть задано диагональной матрицей.*

Переходя от линейного преобразования к матрицам, его задающим, мы получаем следующий результат:

Всякая матрица, все характеристические корни которой действительны и различны, подобна диагональной матрице или, как говорят, *такая матрица приводится к диагональному виду.*

4.6. Примеры решения задач

Операции в линейном пространстве

Прежде чем перейти к рассмотрению некоторых методов решения задач, напомним ряд положений линейного пространства.

Раньше пространство было определено как множество, которое наделено определенной структурой. Были рассмотрены метрические пространства, структура которых определялась тем, что каждой паре элементов приписывалось действительное число, которое удовлетворяет определенным свойствам и называется метрическим. Однако введение метрики далеко не исчерпывает всех структурных свойств различных пространств. В частности, если множество X состоит из действительных или комплексных чисел, то к важным структурным свойствам относится возможность получения одних элементов множества из других путем добавления этих элементов или умножение элемента на скаляр. Множества, которые обладают этими свойствами, относятся к классу линейных пространств. Линейные пространства должны удовлетворять следующим условиям (законам композиции):

1) Закон внутренней композиции: каждой паре элементов $x,y\in X$ однозначно определен третий элемент $z\in X$, который называется их суммой и обозначается $x+y$, причем

$x+y=y+x$ (коммутативность);

$x+(y+v)=(x+y)+v$ (ассоциативность);

в X существует такой элемент 0, что $x+0=x$ для всех $z\in X$, (существование нуля);

для каждого $z\in X$, существует такой элемент $-x$, что $x+(-x)=0$ (существование противоположного элемента);

2) Закон внешней композиции: для любого числа α и любого элемента $z\in X$, определен элемент $\alpha x\in X$, причем

$$(\alpha+\beta)x=\alpha x+\beta x;$$

$$\alpha(x+y)=\alpha x+\alpha y.$$

Условия 1 и 2 называются *условиями аддитивности* и *однородности* линейного пространства.

Множества, элементы которых допускают выполнения операций сложения и умножения на скаляр, весьма разнообразны. Рассмотрим несколько примеров подобных множеств.

Множество вещественных чисел R с обычным определением операций сложения и умножения.

Множество векторов, т.е. направленных отрезков на плоскости или в пространстве.

Множество всех функций $C_{[a,\,в]}$, непрерывных на заданном сегменте [*a, b*], если для любых $x(t)$, $y(t) \in C_{[a,в]}$ под $x(t)+y(t)$ понимать сумму значений $x(t)$ и $y(t)$, взятых при одних и тех же значениях t, а под $\alpha x(t)$ понимать новую функцию, получаемую из $x(t)$ умножением всех ее значений на α. Нулевой функцией будет функция $x(t)=0$, тождественно равная нулю на всем интервале [*a, b*].

Множество всех многочленов степени, которая не превышает натурального числа *n*.

Линейное пространство называется *действительным (комплексным),* если числа (скаляры) в его определении берутся из поля действительных (комплексных) чисел. Так, обычные трехмерные векторы образуют действительное линейное пространство. Внутренним законом этого пространства является геометрическое сложение векторов $\boldsymbol{x}+\boldsymbol{y}$, а внешним законом - умножение вектора на действительное число λ, т.е. $\lambda\boldsymbol{x}$.

Любое поле *K* можно интерпретировать как векторное пространство над самим собой $(X=K)$ со сложением в качестве внутреннего закона и умножением в качестве внешнего закона.

Линейное пространство называют также *векторным пространством,* независимо от природы элементов множества *X,* на котором оно определено.

Однако в дальнейшем мы сосредоточим свое внимание на линейных пространствах, элементами которых являются упорядоченные последовательности *n* действительных чисел, которые мы будем называть *векторами.*

Отметим**,** что под понятие вектора можно подвести и функцию $x(t)$, если рассматривать ее как множество значений $x(t)$ при различных t, а также многочлен, который может быть задан множеством коэффициентов при разных степенях переменной. Тоэму термин «вектор» часто используется для обозначения элементов произвольного линейного пространства.

В *n*-мерном векторном пространстве внутренней операцией является сумма векторов

$$(\boldsymbol{x_1}, \boldsymbol{x_2} ..., \boldsymbol{x_n})+ (\boldsymbol{y_1}, \boldsymbol{y_2} ..., \boldsymbol{y_n})=(\boldsymbol{x_1}+ \boldsymbol{y_1}, \boldsymbol{x_2}+ \boldsymbol{y_2},..., \boldsymbol{x_n}+ \boldsymbol{y_n}).$$

Нейтральным элементом относительно сложения является нулевой вектор 0 = (0, 0, ..., 0), а обратный к $\boldsymbol{x}= (\boldsymbol{x_1}, \boldsymbol{x_2} ..., \boldsymbol{x_n})$ — вектор $-\boldsymbol{x}=(-\boldsymbol{x_1}, -\boldsymbol{x_2},...., -\boldsymbol{x_n})$. Внешняя операция - произведение скаляра на вектор определяется как

$$\lambda\boldsymbol{x}= (\lambda\boldsymbol{x_1}, \lambda\boldsymbol{x_2} ..., \lambda\boldsymbol{x_n}),$$

в результате которой получается вектор той же размерности, что и исходный.

В общем случае законы композиции на множестве объектов линейного пространства могут быть заданы любым способом, лишь бы они удовлетворяли определениям, которые приведены выше. Пусть, например, на трехэлементном множестве $X = \{a, b, c\}$ внутренний закон композиции задан таблицей:

$+$	a	b	c
a	b	c	a
b	c	a	b
c	a	b	c

Относительно этого закона (он никак не связан с арифметическим понятием суммы чисел) множество $X = \{a, b, c\}$ образует абелеву группу с нейтральным элементом c. Внешний закон зададим над полем вычетов по модулю 3 следующими таблицами:

$\cdot$	a	b	c
0	c	c	c
1	a	b	c
2	b	a	c

;

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

;

$\cdot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

.

Первая таблица определяет внешнюю операцию линейного пространства, а другие две - внутренние операции поля K вычетов по модулю 3. Легко проверить, что заданные в такой образом законы композиции удовлетворяют всем требованиям, которые приведены выше, и определяют линейное пространство.

Действия над пространствами

Пусть M — линейное пространство над полем K и L — подмножество из M. Рассмотрим на L внутренний и внешний законы композиции, индуцированные соответственно внутренним и внешним законами композиции на M. Если эти законы превращают L в линейное пространство над K, то L называется *линейным подпространством* линейного пространства M. Следовательно, подпространство L обязательно содержит нейтральный элемент 0 относительно первого закона композиции, а также всякие линейные комбинации $\lambda_1\boldsymbol{x}_1+\lambda_2\boldsymbol{x}_2$ для любой пары $\boldsymbol{x}_1$ и $\boldsymbol{x}_2$ из L и любых λ_1 и λ_2 из поля K, над которым определено пространство M.

Пусть $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2 ..., \boldsymbol{x}_m\}$ — некоторая совокупность векторов из M. Конечные линейные комбинации этих векторов

$$\lambda_1\boldsymbol{x}_1+\lambda_2\boldsymbol{x}_2+\ldots+\lambda_m\boldsymbol{x}_m$$

с коэффициентами $\lambda_i \in K$ образуют *минимальное подпространство* M' пространства M. Говорят, что пространство M' *рождено* множеством X, а множество X есть *система образующих пространства* M'. Например, пространство, порожденное вектором $\boldsymbol{u}$, состоит из всех элементов вида $\alpha\boldsymbol{u}$.

Размерность любого подпространства n-мерного пространства M не превышает числа n. Если в L векторы $\boldsymbol{e}_1$ $\boldsymbol{e}_2$, ..., $\boldsymbol{e}_m$ образуют базис, причем $m<n$, то его можно дополнить независимой совокупностью векторов $\boldsymbol{e}_{m+1}, ..., \boldsymbol{e}_n$ так, что он будет базисом в M.

Пересечение подпространств L_1 и L_2 есть подпространство $L_1 \cap L_2$, векторы которого принадлежат одновременно и L_l и L_2. Во всех случаях пересечение содержит нейтральный элемент 0 относительно первого закона композиции (сложения векторов), т.е. $\{0\} \subset L_1 \cap L_2$.

Сумма подпространств (алгебраическая сумма) L_1+L_2 есть подпространство, элементами которого являются все векторы $\boldsymbol{x}+\boldsymbol{y}$, где $\boldsymbol{x} \in L_1$ и $\boldsymbol{y} \in L_2$.

Прямое произведение линейных пространств $V= M\times T$ есть множество всех упорядоченных пар векторов $(\boldsymbol{x},\boldsymbol{y}) \in V$, где $\boldsymbol{x} \in M$ и $\boldsymbol{y} \in T$. Оно представляет собой новое пространство, элементами которого являются векторы $\boldsymbol{z}=(\boldsymbol{x}, \boldsymbol{y})$, причем операции над этими векторами определяются следующим образом:

$$\boldsymbol{z}'+\boldsymbol{z}''=(\boldsymbol{x}'+\boldsymbol{x}'', \boldsymbol{y}'+\boldsymbol{y}'') \text{ и } \lambda\boldsymbol{z}=(\lambda\boldsymbol{x}, \lambda\boldsymbol{y}), \text{ где } \lambda \in K.$$

Для конечномерных пространств $\dim V = \dim M + \dim T$.

Пусть $\boldsymbol{x}= (\boldsymbol{x}_1 ,\boldsymbol{x}_2, ..., \boldsymbol{x}_m)$ — m-мерные векторы из M и $\boldsymbol{y}=(\boldsymbol{y}_1, \boldsymbol{y}_2 \ldots, \boldsymbol{y}_n)$ — n-мерные векторы из T. Тогда элементами прямого произведения $V=M\times T$ будут $(m+n)$-мерные векторы вида

$$\boldsymbol{z}= (\boldsymbol{x}, \boldsymbol{y})=(\boldsymbol{x}_1\, \boldsymbol{x}_2, ..., \boldsymbol{x}_m; \boldsymbol{y}_1\, \boldsymbol{y}_2, ..., \boldsymbol{y}_n).$$

Векторы из V можно представить как $\boldsymbol{z}= (\boldsymbol{x}, \boldsymbol{y})= (\boldsymbol{x}, 0) + (0, \boldsymbol{y})$. Это значит, что прямое произведение $M\times T$ есть сумма некоторых пространств M' и T', элементами которых есть соответственно векторы $(\boldsymbol{x}, 0)$ и $(0, \boldsymbol{y})$. В то же время M' и T' являются подпространствами V, причем все отличные от нуля векторы у них различные, и только при

$x = y = 0$ векторы $(x, 0)$ и $(0, y)$ совпадают и равны нейтральному вектору 0. Поэтому $M' \bigcap T' = \{0\}$ и $M' + T' = V$.

Если подпространства линейного пространства M удовлетворяют этим условиям, то говорят, что M разложимо в *прямую сумму* L'_1 и L'_2 и записывают $L'_1 \oplus L'_2 = M$.

В общем случае линейное пространство M есть прямая сумма подпространств $L'_1, L'_2, \ldots, L'_k$, т.е. $M = L'_1 \oplus L'_2 \oplus \ldots \oplus L'_k$, если

$$M = L'_1 + L'_2 + \ldots + L'_k \text{ и } L'_1 \bigcap (L'_1 + L'_2 + \ldots + L'_{u-1} + L'_{i+1} + \ldots + L'_k) = \{0\}$$

для всех $i = 1, 2, \ldots, k$. Это значит, что любой вектор x из M выражается единственных образом как сумма $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2 + \ldots + \boldsymbol{x}_k$, где $x_i \in L'_i$.

Прямая сумма ассоциативна и коммутативна. Можно также утверждать, что пространство M порождается множеством всех векторов подпространств $L'_1, L'_2, \ldots, L'_k$.

Если размерности подпространств $\dim L_i = n_i$, то объединение базисов $L'_1, L'_2, \ldots, L'_k$ образует базис пространства M и его размерность

$$\dim M = \sum_{i=1}^{k} n_i \ .$$

Если $M = L'_1 \oplus L'_2$, то L'_1 и L'_2 называются *алгебраически дополнительными,* причем

$$\dim L'_1 + \dim L'_2 = \dim M.$$

Пусть размерности произвольных подпространств L_1 и L_2 линейного пространства M равны n_1 и n_2, а размерность пересечения $L_1 \cap L_2$ равна s. Рассмотрим пространство M_k, порожденнное подпространствами L_1 и L_2, т.е. $M_k = L_1 \oplus L_2$, размерность которого обозначим через k. Тогда имеет место тождество: $k + s = n_1 + n_2$. В частном случае, если $L_1 \bigcap L_2 = \{0\}$, т.е. размерность $L_1 \cap L_2$ равна нулю, то $k = n_1 + n_2$.

Рассмотрим еще несколько примеров с использованием следующих свойств сложения и умножения вектора на число:

1. $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$.
2. $(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z})$.
3. Существует *нулевой* вектор **0** такой, что $\boldsymbol{x} + \boldsymbol{0} = \boldsymbol{x}$.
4. Для каждого вектора $\boldsymbol{x}$ существует *противоположный* вектор $\boldsymbol{y} = -\boldsymbol{x}$ такой, что $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{0}$.
5. $1 \cdot \boldsymbol{x} = \boldsymbol{x}$.
6. $\lambda(\mu \boldsymbol{x}) = (\lambda\mu)\boldsymbol{x}$.
7. $(\lambda + \mu)\boldsymbol{x} = \lambda \boldsymbol{x} + \mu \boldsymbol{x}$.
8. $\lambda(\boldsymbol{x} + \boldsymbol{y}) = \lambda \boldsymbol{x} + \lambda \boldsymbol{y}$.

1-а) Совокупность векторов, которые лежат на одной прямой, образует линейное пространство, так как сложение и умножение таких векторов на действительное число приводит нас снова к векторам, которые лежат на этой прямой, и свойства сложения и умножения (их 8) легко проверяются. Обозначим такое линейное пространство через L_1.

1-б*)* Совокупность векторов, которые лежат в одной плоскости, также оказывается замкнутой относительно сложения и умножения на действительное число; свойства сложения и умножения для них выполняются, и поэтому эта совокупность образует линейное пространство, которое мы обозначим через L_2.

1-в*)* Совокупность всех векторов пространства также является линейным пространством. Обозначим его через L_3.

1-г) Совокупность векторов, которые лежат в плоскости *XOY,* начала которых совпадают с началом координат, а концы лежат в первом квадранте, не образует линейного пространства, так как оказывается незамкнутой относительно умножения на число: при $\lambda<0$ вектор $\lambda\boldsymbol{x}$ не принадлежит первому квадранту.

1-д*)* Рассмотрим множество, элементом которого является упорядоченная совокупность *n* действительных чисел: $\boldsymbol{x}=\{x_1,x_2,\dots. x_n\}$. Определим сложение элементов $\boldsymbol{x}$ и $\boldsymbol{y}=\{y_1,y_2,\dots. y_n\}$ и умножение элемента $\boldsymbol{x}$ на действительное число λ с помощью равенств

$$\boldsymbol{x}+\boldsymbol{y}=\{x_1+y_1,\ x_2+y_2,\ \dots,\ x_n+y_n\},$$

$$\lambda\boldsymbol{x}=\{\lambda\boldsymbol{x}_1,\ \lambda\boldsymbol{x}_2,\ \dots,\ \lambda\boldsymbol{x}_n\}.$$

Такое множество элементов образует линейное пространство, так как определенные в нем операции сложения и умножения на число обладают, как легко видеть, всеми восемью указанными выше свойствами этих операций. Например, нулевым вектором в этом пространстве будет вектор $\mathbf{0}=\{0,0,\dots,0\}$, а вектором $-\boldsymbol{x}$ — вектор $\{-x_1,-x_2,\dots,-x_n\}$. Будем обозначать это пространство через L_n.

1-е*)* Совокупность всех многочленов степени не выше *n*

$$P(t)=a_0+a_1t+\dots+a_nt^n,$$

для которых обычным образом определены сложения и умножения на действительное число, как легко проверить, также образует линейное пространство.

1-ж*)* Множество непрерывных на отрезке $[a,\ b]$ функций $\varphi(t)$ также образует линейное пространство, если для этих функций естественным образом определить операции сложения и умноженич на число. Это пространство мы будем обозначать $C_{[a,\ b]}$.

Индивидуальные тестовые задачи

1. Покажите, что множество действительных чисел образует метрическое пространство, если задать метрику одним из следующих способов:

а) $\rho(x,y)=|x-y|$;

б) $\rho(x,y)=d/(x-y)$, где $d=|x-y|$;

б) $\rho(x,y)=\begin{cases} 0\ при\ x=y, \\ 1\ при\ x\neq y. \end{cases}$

2. Множество слов длины n, состоящих из n символов (букв, цифр, знаков и т.п.) конечного алфавита, можно рассматривать как метрическое пространство, если ввести на нем соответствующим образом метрику. Например, в качестве расстояния $\rho(x, y)$ между двумя словами можно принять количество позиций, в которых слова x и y содержат различные символы.

а) Покажите, что введенное таким образом расстояние удовлетворяет аксиомам метрического пространства.

б) Найдите расстояния между русскими словами x = (лекция), y = (секрет), z = (секция).

в) Найдите расстояния между двоичными кортежами:

x = (100101101),

y = (001101001),

z= (101001010).

г) Проверьте на данных примерах свойства метрики.

3. Пусть M — множество, которое состоит из двух элементов a и b. Покажите, что система подмножеств, которая содержит, наряду с M и пустым множеством, одноэлементное множество $\{b\}$, является топологией в M. Какие подмножества данного топологического пространства являются замнутыми?

4. Постройте все топологии в пространстве M, состоящее из трех, четырех и пяти элементов.

5. Покажите, что законы композиции определяют линейное пространство.

6. На приведенных ниже множествах обычным образом введены внутренний закон (сложение) и внешний закон (умножение на число). Кокие из этих множеств образуют линейные пространства:

а) многочлены степени n;

б) многочлены степени, меньшей или равной n;

в) многочлены степени, большей или равной n;

г) все векторы плоскости;

д) все векторы плоскости, не параллельные оси абсцисс;

е) все векторы плоскости, которые выходят из начала координат и расположенные в первом квадранте;

ж) квадратные матрицы n-го порядка.

7. Является ли линейно-независимой совокупность трехмерных векторов:

а) $\boldsymbol{x}_1 = (5, 1, 4)$; $\boldsymbol{x}_2 = (0, 3, 2)$; $\boldsymbol{x}_3 = (2, 1, 2)$;

б) $\boldsymbol{x}_1 = (2, 3, 1)$; $\boldsymbol{x}_2 = (3, 4, 2)$; $\boldsymbol{x}_3 = (1, 5, 4)$?

8. Дано множество X, которое состоит из шести векторов:

$\boldsymbol{x}_1 = (1, 1,0, 0)$; $\boldsymbol{x}_2 = (1, 0, 1, 0)$; $\boldsymbol{x}_3 = (1, 0, 0, 1)$; $\boldsymbol{x}_4 =(0, 1, 1,0)$; $\boldsymbol{x}_5 = (0, 1, 0, 1)$; $\boldsymbol{x}_6 = (0, 0, 1, 1)$. Найдите максимальное линейно-независимое подмножество X' этого множества и выразите другие векторы как линейные комбинации векторов из X'. Имеет ли задача другие решения?

9. Выяснить, образует ли линейное пространство

а) совокупность всех векторов пространства L_2 (см. пример 1-б)), за исключением векторов, параллельных некоторой заданной прямой;

б) совокупность всех векторов пространства L_2, концы которых лежат на заданной прямой;

в) совокупность всех векторов пространства L_3 (см. пример 1-в)), концы которых не лежат на данной прямой.

10. Выяснить, образует ли линейное пространство совокупность векторов пространства L_n (см. пример 1-д)), для которых

а) $\boldsymbol{x}_1 + \boldsymbol{x}_2 + \ldots + \boldsymbol{x}_n = 0$;

б) $\boldsymbol{x}_1 + \boldsymbol{x}_2 + \ldots + \boldsymbol{x}_n = 1$;

в) $\boldsymbol{x}_1 = \boldsymbol{x}_3$;

г) $\boldsymbol{x}_2 = \boldsymbol{x}_4 = \ldots$;

д) первая компонента - целе число;

е) $\boldsymbol{x}_1$ или $\boldsymbol{x}_3$ равен нулю;

11. Выяснить, образует ли линейное пространство совокупность многочленов, степень которых равна n (ср. с примером 1-е))?

12. Пусть R^+ — совокупность положительных действительных чисел. Под «сложением» двух чисел будем понимать их обычное умножение, а под «умножением» элемента $p \in R^+$ на действительное число λ будем понимать обычное возведение числа p в степень λ. Образует ли множество R^+ с таким образом введенными на нем операциями линейное пространство? Чему равен «противоположный» элемент для $p \in R^+$. Какой элемент служит «нулем» этого просторную?

13. Доказать, что сумма и пересечения двух линейных подпространств линейного просторную L сами являются линейными подпространствами L.

14. Перечислить все типы подпространств пространства L_3.

15. Какие из совокупности векторов задачи 10 образуют подпространства L_n?

16. Пусть $\boldsymbol{a}$ и $\boldsymbol{b}$ - два линейно независимых вектора пространства L_2 (см. пример 1-б).

1) Определить, при каком значении $\boldsymbol{a}$ следующие пары векторов линейно зависимы (коллинеарны):

а) $\alpha\boldsymbol{a}+2\boldsymbol{b}$, $\boldsymbol{a} — \boldsymbol{b}$;

б) $(\alpha+1)\boldsymbol{a} + \boldsymbol{b}$, $2\boldsymbol{b}$;

в) $\alpha\boldsymbol{a}+\boldsymbol{b}$, $\boldsymbol{a}+\alpha\boldsymbol{b}$.

2) Найти α и β, если

а) $3\boldsymbol{a}+5\boldsymbol{b}= \alpha\boldsymbol{a}+(2\beta +1)\boldsymbol{b}$;

б) $(2\alpha — \beta — 1)\boldsymbol{a} — (3\alpha + \beta + 10)\boldsymbol{b}=0$.

2. Пусть $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ — три линейно независимых вектора пространства L_3 (см. пример 1-б).

а) При каком значении векторы

$$\boldsymbol{x}= \alpha\boldsymbol{a} +4\boldsymbol{b}+2\boldsymbol{c},\ \boldsymbol{y} =\boldsymbol{a}+\alpha\boldsymbol{b} — \boldsymbol{c}$$

линейно зависимы (коллинеарны)?

б) При каком значении α векторы

$$\boldsymbol{x}=\alpha\boldsymbol{a}+\boldsymbol{b}+3\boldsymbol{c},\ \boldsymbol{y}=\alpha\boldsymbol{a}—2\boldsymbol{b} + \boldsymbol{c},\ \ \boldsymbol{z}=\boldsymbol{a} — \boldsymbol{b}+\boldsymbol{c}$$

линейно зависимы (компланарны)?

17. Доказать, что в пространстве многочленов степени, не большей n (см. пример 1-е), многочлены

$$P_0(t) =1, P_1(t) =t,..... P_n(t) = t^n$$

линейно независимы.

18. Доказать, что в пространстве $C_{[a,\ b]}$ (см. пример 1-ж) существует любое число линейно независимых векторов.

19. Доказать линейную зависимость векторов $\boldsymbol{a}_1=\{0,\ 1,\ 1\}$, $\boldsymbol{a}_2=\{1, 1, 2\}$, $\boldsymbol{a}_3=\{1, 2, 3\}$ пространства L_3.

20. Доказать, что совокупность векторов будет линейно зависимой, если она содержит

а) два равных вектора,

б) два коллинеарных вектора.

21. Доказать, что если векторы $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ линейно независимы, то и векторы $\boldsymbol{a}_1+\boldsymbol{a}_2$, $\boldsymbol{a}_3+ \boldsymbol{a}_4$, $\boldsymbol{a}_3+ \boldsymbol{a}_1$ линейно независимы.

22. Доказать, что векторы $\boldsymbol{a}_1 =\{l,\ 1,\ 1\}$, $\boldsymbol{a}_2 =\{1,\ 1, 2\}$ и $\boldsymbol{a}_3=\{1, 2, 3\}$ образуют базис пространства L_3 и записать в этом базисе вектор $\boldsymbol{x}=\{6, 9, 14\}$.

23. Какова размерность пространства $C_{[a,\ b]}$ (см. пример 1-же).

24. Найти размерность и базис пространства, рассмотренного в задаче 12.

25. Доказать, что если размерность подпространства L'_n совпадает с размерностью пространства L_n, то $L_n \equiv L'_n$,

26. Доказать, что сумма размерностей двух линейных подпространств L' и L'' пространства - L_n равна размерности суммы этих подпространств, сложенной с размерностью их пересечения.

27. Доказать, что если размерность суммы двух линейных подпространств L' и L'' на единицу больше размерности их пересечения, то сумма совпадает с одним из этих подпространств, а пересечение - с другим.

28. Доказать, что если два подпространств L' и L'' пространства L имеют общим лишь нулевой вектор, то сумма их розмерностей не превосходит размерности L.

29. Что представляет собой пересечение и сумма двух двумерных подпространств пространства L_3?

Линейным подпространством L' пространства L_n, *определяемым векторами* $\boldsymbol{a}_1$, $\boldsymbol{a}_2$,..., $\boldsymbol{a}_m$, называется подпространство наименьшей размерности, содержащее эти векторы.

30. Найти размерность s суммы и размерность d пересечения линейных подпространств L' и L'' пространства L_4, определяемых векторами

L': $\boldsymbol{a}_1 = \{1, 1, 1, 1\}$, $\boldsymbol{a}_2=\{1, —1, 1, -1\}$, $\boldsymbol{a}_3=\{1, 3, 1, 3\}$;

L'': $\boldsymbol{b}_1= \{l, 2, 0, 2\}$, $\boldsymbol{b}_2=\{1, 2, 1, 2\}$, $\boldsymbol{b}_3= \{3, 1, 3, 1\}$.

31. Найти базисы суммы и пересечения линейных подпространств L' и L'' пространства L_4, определяемых векторами

L': $\boldsymbol{a}_1=\{1, 2, 1, — 2\}$, $\boldsymbol{a}_2 = \{2, 3, 1, 0\}$, $\boldsymbol{a}_3 = \{1, 2, 2, - 3\}$;

L'': $\boldsymbol{b}_1=\{l, 1, 1, 1\}$, $\boldsymbol{b}_2=\{1, 0, 1, -1\}$; $\boldsymbol{b}_3=\{1, 3, 0, -4\}$.

32. Найти какой-нибудь базис и определить размерность следующих подпространств пространства L_n:

а) совокупность n-мерных векторов, у которых первая и вторая координаты равны;

б) совокупность n-мерных векторов, у которых координаты с четными номерами равны между собой;

в) совокупности n-мерных векторов вида $\{\alpha, \beta, \alpha, \beta, ...\}$;

г) совокупности n-мерных векторов $\{x_1, ..., x_n\}$, для которых

$$x_1+ x_2+, ...,+ x_n=0.$$

5. Аффинное пространство

5.1. Основные понятия

Определение. *Аффинным пространством* над полем F называется пара $(\mathfrak{A}, V)$, где $\mathfrak{A}$ -- непустое множество (точек), V -- линейное пространство над полем F, такая, что

1) любой паре точек $P,\ Q \in \mathfrak{A}$ сопоставлен вектор $\overrightarrow{PQ} \in V$,

2) для каждой точки $P \in \mathfrak{A}$ и каждого вектора $\mathbf{x} \in V$ существует единственная точка Q такая, что $\overrightarrow{PQ} = \mathbf{x}$,

3) для любых трех точек $P,\ Q,\ S \in \mathfrak{A}$ справедливо равенство $\overrightarrow{PQ} + \overrightarrow{QS} = \overrightarrow{PS}$.

Замечание. 1) $\overrightarrow{PP} = \mathbf{0}$, так как $\overrightarrow{PP} + \overrightarrow{PP} = \overrightarrow{PP}$.

2) $\overrightarrow{PQ} = -\overrightarrow{QP}$, так как $\overrightarrow{PQ} + \overrightarrow{QP} = \overrightarrow{PP} = \mathbf{0}$.

3) Если $Q = P + \mathbf{x}$, то $(P + \mathbf{x}) + \mathbf{y} = P + (\mathbf{x} + \mathbf{y})$.

Определение. Аффинное пространство называется n-мерным, если линейное пространство V n-мерно. Под системой координат понимается пара $(O, (\mathbf{e}_1, \ldots, \mathbf{e}_n))$, где $O \in \mathfrak{A}$, а

$(\mathbf{e}_1, \dots, \mathbf{e}_n)$ - базис в V. *Координаты* произвольной точки P - это координаты вектора $\overrightarrow{OP}$ в базисе $(\mathbf{e}_1, \dots, \mathbf{e}_n)$.

Теорема. Пусть $(O, (\mathbf{e}_1, \dots, \mathbf{e}_n))$ и $(\tilde{O}, (\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n))$ - две системы координат n-мерного аффинного пространства. Тогда столбцы координат X и $\tilde{X}$ произвольной точки P относительно выбранных систем координат связаны равенством $X = C\tilde{X} + A$, т.е. $x^i = c^i_j \tilde{x}^j + a^i$, где C - матрица перехода от $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ к $(\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n)$ ($\tilde{\mathbf{e}}_i = c^j_i \mathbf{e}_j$), A - столбец координат нового начала $\tilde{O}$ относительно старой системы координат.

Замечание. $1)$ Если $(\mathbf{e}_1, \dots, \mathbf{e}_n) = (\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n)$, то $X = \tilde{X} + A$.

$2)$ Если $0 = \tilde{O}$, то $X = C\tilde{X}$.

Определение. *Аффинно линейная функция* -- это такая функция $f : \mathfrak{A} \to F$, что существует линейная функция $Df : V \to F$, удовлетворяющая условию: для любых точек $P, Q \in \mathfrak{A}$ справедливо равенство $f(Q) = f(P) + Df(\overrightarrow{PQ})$. Функция Df называется *дифференциалом* функции f.

Пусть $(O, (\mathbf{e}_1, \ldots, \mathbf{e}_n))$ - система координат в аффинном пространстве $(\mathfrak{A}, V)$, и $f : \mathfrak{A} \to F$ - линейная функция такая, что $f(O) = b$ и $Df(\mathbf{e}_i) = a_i$. Тогда $f(Q) = f(O) + Df(\overrightarrow{OQ}) = b + x^i a_i$, где $(x^1, \ldots, x^n)$ - координаты точки Q.

Определение. *Плоскостью* Π в аффинном пространстве $(\mathfrak{A}, V)$ называется множество точек вида $\Pi = P + U = \{Q = P + \mathbf{x}, \mathbf{x} \in U\}$, где $P \in \mathfrak{A}$ и U -- подпространство в V.

Лемма. Если P и $P' \in \Pi$, то $P + U = P' + U$.

Определение. *Размерность* $\dim \Pi$ плоскости Π -- это размерность $\dim U$ линейного пространства U. Если $\dim \Pi = n - 1$, то плоскость Π называется *гиперплоскостью*.

Замечание. Пара (Π, U) является аффинным пространством, так как для любых двух точек $P, Q \in \Pi$ имеем $\overrightarrow{PQ} \in U$ и все аксиомы выполнены.

Если $U = \mathrm{Lin}(\mathbf{a}_1, \ldots, \mathbf{a}_k)$, где $\mathbf{a}_i \in V$, то каждая точка $Q \in \Pi$ плоскости $\Pi = P + U$ имеет вид $Q = P + t^i \mathbf{a}_i$, где $t^i \in F$. Если X и B -- это столбцы координат точек Q и P соответственно относительно выбранной системы координат $(O, (\mathbf{e}_1, \ldots, \mathbf{e}_n))$, а $\mathbf{a}_i = (a_i^1, \ldots, a_i^n)$ -- координаты вектора a_i в базисе $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$, то мы получаем параметрическое задание плоскости: $x^j = b^j + t^i a_i^j$.

Определение. Точки $A_0, A_1, \ldots, A_k$ находятся *в общем положении* (*аффинно независимы*), если система векторов $(\overrightarrow{A_0 A_1}, \ldots, \overrightarrow{A_0 A_k})$ линейно независима.

Замечание. Если точки $A_0, A_1, \ldots, A_k$ находятся в общем положении, то система $(\overrightarrow{A_i A_0}, \ldots, \overrightarrow{A_i A_{i-1}}, \overrightarrow{A_i A_{i+1}}, \ldots, \overrightarrow{A_i A_k})$ линейно независима, так как она выражается через систему $(\overrightarrow{A_0 A_1}, \ldots, \overrightarrow{A_0 A_k})$.

На самом деле, эти системы эквивалентны, так как $\overrightarrow{A_iA_j} = \overrightarrow{A_0A_j} - \overrightarrow{A_0A_i}$ и $\overrightarrow{A_0A_j} = \overrightarrow{A_iA_j} - \overrightarrow{A_iA_0}$.

Следовательно,

$$\operatorname{rank}(\overrightarrow{A_0A_1}, \ldots, \overrightarrow{A_0A_k}) = \operatorname{rank}(\overrightarrow{A_iA_0}, \ldots, \overrightarrow{A_iA_{i-1}}\overrightarrow{A_iA_{i+1}}, \ldots, \overrightarrow{A_iA_k}).$$

Теорема. Для любых точек $A_0, A_1, \ldots, A_k$ общего положения существует единственная k-мерная плоскость, содержащая эти точки.

Теорема. Пусть $(O, (\mathbf{e}_1, \ldots, \mathbf{e}_n))$ система координат в аффинном пространстве $(\mathfrak{A}, V)$. Тогда

1) множество всех точек Q, координаты которых удовлетворяют совместной линейной системе

$$\begin{cases} a_i^1 x^i = b^1 \\ \ldots \\ a_i^m x^i = b^m, \end{cases} \qquad (1)$$

образуют плоскость в аффинном пространстве размерности $n - \operatorname{rank}(a_j^i)$;

2) для любой k-мерной плоскости Π существует линейная система (1) ранга $n-k$ такая, что множество точек, координаты которых удовлетворяют этой системе, есть плоскость Π.

Замечание. Если плоскость задается системой (1), то направляющее подпространство соответствует однородной системе.

Определение. Две плоскости Π_1 и Π_2 *пересекаются*, если есть общая точка, т.е. $\Pi_1 \cap \Pi_2 \neq \emptyset$.

Теорема. $1)$ Две плоскости $\Pi_1 = P_1 + U_1$ и $\Pi_2 = P_2 + U_2$ имеют общую точку тогда и только тогда, когда вектор $\overrightarrow{P_1P_2} \in U_1 + U_2$.

$2)$ Если $\Pi_1 \cap \Pi_2 \neq \emptyset$, то $\Pi_1 \cap \Pi_2$ - плоскость с направляющим подпространством $U_1 \cap U_2$.

Определение. Плоскости $\Pi_i = P_i + U_i$, $i = 1, 2$, *параллельны* $\Pi_1 \parallel \Pi_2$, если $U_1 \subset U_2$.

Теорема. $1)$ Если $\Pi_1 \parallel \Pi_2$, то либо $\Pi_1 \cap \Pi_2 = \emptyset$, либо $\Pi_1 \subset \Pi_2$.

$2)$ Пусть $\dim\Pi_1 = \dim\Pi_2$ и эти плоскости заданы линейными системами вида (1) в некоторой системе координат. Тогда необходимым и достаточным условием параллельности плоскостей

Π_i является эквивалентность соответствующих однородных систем вида (2).

Определение. Две плоскости *скрещиваются*, если у них нет общих точек и они не параллельны.

Теорема. 1) Наименьшей плоскостью Π, содержащей плоскости $\Pi_i = P_i + U_i$, $i = 1, 2$, является плоскость $\Pi = P_1 + U$, где $U = U_1 + U_2 + \mathrm{Lin}(\overrightarrow{P_1 P_2})$.

2) $\dim\Pi = \dim(U_1 + U_2)$, если $\Pi_1 \cap \Pi_2 \neq \emptyset$, и $\dim\Pi = \dim(U_1 + U_2) + 1$, если $\Pi_1 \cap \Pi_2 = \emptyset$.

Определение. *Евклидовым аффинным* пространством называется вещественное аффинное пространство $(\mathfrak{A}, E)$, где E - евклидово пространство.

Расстоянием $\rho(P, Q)$ между любыми двумя точками $P, Q \in \mathfrak{A}$ называется $|\overrightarrow{PQ}|$. *Расстоянием* $\rho(M, X)$ от точки $M \in \mathfrak{A}$ до множества $X \subset \mathfrak{A}$ называется $\inf_{N \in X} \rho(M, N)$.

Лемма. Для любых трех точек $P, Q, S \in \mathfrak{A}$ имеет место

1) $\rho(P, Q) = \rho(Q, P)$,

2) $\rho(P,S) \leqslant \rho(P,Q)+\rho(Q,S)$,

3) $|\overrightarrow{PS}|^2 = |\overrightarrow{PQ}|^2 + |\overrightarrow{QS}|^2$, если $\overrightarrow{PQ} \perp \overrightarrow{QS}$.

Теорема. 1) Для любой точки $M \in \mathfrak{A}$ и любой плоскости $\Pi = P + U$ в евклидовом аффинном пространстве $(\mathfrak{A}, E)$ существует единственная точка $Q \in \Pi$ такая, что $\overrightarrow{MQ} \in U^{\perp}$.

2) $|\overrightarrow{MQ}| = \rho(M, \Pi)$, причем для любой точки $Q' \in \Pi$ и $Q' \neq Q$ справедливо неравенство $|\overrightarrow{MQ'}| > |\overrightarrow{MQ}|$.

Определение. Пусть даны две плоскости $\Pi_i = P_i + U_i$, $i = 1,\,2$. *Расстоянием* между плоскостями $\rho(\Pi_1, \Pi_2)$ называется $\inf_{N_i \in \Pi_i} \rho(N_1, N_2)$.

Найдем расстояние между двумя плоскостями $\Pi_i = P_i + U_i$, $i = 1,\,2$. Пусть $U = U_1 + U_2$. Тогда $\overrightarrow{P_1P_2} = \mathbf{u} + \mathbf{z}$, где

$\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2 \in U$, $\mathbf{u}_i \in U_i$ и $\mathbf{z} \in U^{\perp}$. Рассмотрим две точки $M_1 = P_1 + \mathbf{u}_1 \in \Pi_1$ и $M_2 = P_2 - \mathbf{u}_2 \in \Pi_2$. Имеем

$$\overrightarrow{M_1M_2} = \overrightarrow{M_1P_1} + \overrightarrow{P_1P_2} + \overrightarrow{P_2M_2} = -\mathbf{u}_1 + \overrightarrow{P_1P_2} - \mathbf{u}_2 = \mathbf{z} \in (U_1 + \mathbf{u}_2)^{\perp}.$$

Следовательно, $\overrightarrow{M_1M_2}$ - общий перпендикуляр.

Напомним, что матрицей Грамма системы векторов $(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ называется матрица

$$G(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \begin{pmatrix} (\mathbf{a}_1, \mathbf{a}_1) & \ldots & (\mathbf{a}_1, \mathbf{a}_n) \\ \vdots & \ddots & \vdots \\ (\mathbf{a}_n, \mathbf{a}_1) & \ldots & (\mathbf{a}_n, \mathbf{a}_n) \end{pmatrix}.$$

Определитель этой матрицы называется определителем Грамма. Отметим, что определитель Грамма не меняется при процессе ортогонализации.

Теорема. Для любой точки M и k-мерной плоскости $\Pi = P + U$, где $U = \mathrm{Lin}(\mathbf{a}_1, \ldots, \mathbf{a}_k)$,

$$\rho^2(M, \Pi) = \frac{\det G(\mathbf{a}_1, \ldots, \mathbf{a}_k, \overrightarrow{PM})}{\det G(\mathbf{a}_1, \ldots, \mathbf{a}_k)}.$$

Определение. Пусть $O \in \mathfrak{A}$ -- произвольная точка в аффинном пространстве $(\mathfrak{A}, V)$, и $(\mathbf{a}_1, \ldots, \mathbf{a}_k)$ -- линейно независимая система векторов из V. *Параллелепипедом* называется следующее множество $\{Q = O + t^i\mathbf{a}_i | 0 \leqslant t^i \leqslant 1\}$.

Рассмотрим теперь аффинное евклидово пространство.

Объем V параллелепипеда, построенного на векторах $(\mathbf{a}_1, \ldots, \mathbf{a}_k)$, равен $V = \sqrt{\det G(\mathbf{a}_1, \ldots, \mathbf{a}_k)}$, а высота h, опущенная на основание $(\mathbf{a}_1, \ldots, \mathbf{a}_{k-1})$, -- $h = \dfrac{\sqrt{\det G(\mathbf{a}_1, \ldots, \mathbf{a}_k)}}{\sqrt{\det G(\mathbf{a}_1, \ldots, \mathbf{a}_{k-1})}}$.

5.2. Аффинные преобразования и движения

Определение. *Аффинным* отображение из одного аффинного пространства $(\mathfrak{A}_1, V_1)$ в другое $(\mathfrak{A}_2, V_2)$ называется такое отображение $\varphi\colon \mathfrak{A}_1 \to \mathfrak{A}_2$, для которого существует линейное отображение $D\varphi\colon V_1 \to V_2$ (*дифференциал отображения* φ), удовлетворяющее условию: для любых двух точек $P, Q \in \mathfrak{A}_1$ справедливо равенство $\varphi(Q) = \varphi(P) + D\varphi(\overrightarrow{PQ})$ или $\overrightarrow{\varphi(P)\varphi(Q)} = D\varphi(\overrightarrow{PQ})$.

Теорема. Пусть $O_i \in \mathfrak{A}_i$ -- произвольные точки, и $\mathcal{A}\colon V_1 \to V_2$ -- линейное отображение. Тогда существует единственное аффинное отображение $\varphi\colon \mathfrak{A}_1 \to \mathfrak{A}_2$ такое, что $\varphi(O_1) = O_2$ и $D\varphi = \mathcal{A}$. При этом, отображение φ биективно тогда и только тогда, когда отображение $\mathcal{A}$ биективно.

Замечание. Таким образом, для любой точки $P = O_1 + \mathbf{x} \in \mathfrak{A}_1$ получаем $\varphi(P) = O_2 + \mathcal{A}(\mathbf{x})$

Определение. Аффинное отображение φ называется *аффинным преобразованием*, если $(\mathfrak{A}_1, V_1) = (\mathfrak{A}_2, V_2)$ и φ является биекцией.

Замечание. Все аффинные преобразования обозначаются $\mathrm{Aff}\,(\mathfrak{A})$.

Пусть даны два аффинных преобразования φ и ψ. Тогда

$$(\varphi\psi)(Q) = \varphi(\psi(Q)) = \varphi(\psi(P) + D\psi(\overrightarrow{PQ})) = \varphi(\psi(P)) + (D\varphi)\big((D\psi)(\overrightarrow{PQ})\big) = (\varphi\psi)(P) + (D\varphi D\psi)(\overrightarrow{PQ}) = (\varphi\psi)(P) + (D\varphi\psi)(\overrightarrow{PQ})$$

.

Следствие. 1) Аффинное отображение $\varphi\colon \mathfrak{A} \to \mathfrak{A}$ является аффинным преобразованием тогда и только тогда, когда $D\varphi$ есть невырожденный линейный оператор.

2) Если φ - аффинное преобразование, то φ^{-1} также является аффинным преобразованием с дифференциалом $(D\varphi)^{-1}$.

Теорема. Пусть $(A_0, A_1, \ldots, A_n)$ и $(B_0, B_1, \ldots, B_n)$ -- две системы точек общего положения в n-мерном аффинном пространстве. Тогда существует единственное аффинное преобразование φ такое, что $\varphi(A_i) = B_i$.

Теорема. Аффинное преобразование переводит любую k-мерную плоскость в k-мерную плоскость. В частности, прямые переходят в прямые, причем φ сохраняет отношение, в котором точка делит отрезок.

Теорема. Пусть $(O, (\mathbf{e}_1, \ldots, \mathbf{e}_n))$ - система координат аффинного пространства $(\mathfrak{A}, V)$, а φ - аффинное преобразование. Тогда столбец X координат произвольной точки P и столбец Y координат образа $\varphi(P)$ связаны соотношением $Y = AX + B$, где A - матрица оператора $D\varphi$ в базисе $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$, а B - столбец координат точки $\varphi(O)$.

Определение. *Параллельный перенос* (*сдвиг*) - это такое аффинное преобразование, у которого оператор $D\varphi$ является тождественным линейным оператором.

Пусть φ - сдвиг. Тогда $Y = X + B$ или

$$\varphi(P) = \varphi(O) + D\varphi(\overrightarrow{OP}) = \varphi(O) + \overrightarrow{OP} = P + \overrightarrow{O\varphi(O)} = P + \mathbf{a},$$

где $\overrightarrow{O\varphi(O)} = \mathbf{a}$. Таким образом, $\varphi = \tau_{\mathbf{a}}$ - сдвиг на вектор $\mathbf{a}$. Отметим, что сдвиги образуют группу: $\tau_{\mathbf{a}}\tau_{\mathbf{b}} = \tau_{\mathbf{a}+\mathbf{b}}$ и $(\tau_{\mathbf{a}})^{-1} = \tau_{-\mathbf{a}}$.

Теорема. Пусть φ - аффинное преобразование, P - любая точка. Тогда существует единственное разложение $\varphi = \tau_{\mathbf{a}}\psi$ ($\varphi = \psi\tau_{\mathbf{b}}$), где ψ - аффинное преобразование с неподвижной точкой P ($\psi(P) = P$). При этом $\mathbf{a} = \overrightarrow{P\varphi(P)}$ ($\mathbf{b} = \overrightarrow{\varphi^{-1}(P)P}$).

Определение. Пусть $(\mathfrak{A}, E)$ - евклидово аффинное пространство. Аффинное преобразование называется *движением*, если $D\varphi$ является ортогональным оператором.

Замечание. Движения образуют группу: произведение движений - движение, если φ - движение, то φ^{-1} тоже движение ($D(\varphi^{-1}) = (D\varphi)^{-1}$).

Пусть φ -- движение. Тогда

$$\rho(\varphi(P),\varphi(Q)) = |\overrightarrow{\varphi(P)\varphi(Q)}| = |D\varphi(\overrightarrow{PQ})| = |\overrightarrow{PQ}| = \rho(P,Q).$$

По теореме, любое преобразование можно представить в виде $\varphi = \tau_{\mathbf{a}}\psi$, где $\psi(P) = P$. Тогда $\psi(P+\mathbf{x}) = P + D\psi(\mathbf{x})$ и $D\varphi = D\tau_{\mathbf{a}}D\psi$, где $D\tau_{\mathbf{a}}$ -- тождественный оператор.

Если φ - движение, то оператор $\mathcal{A} = D\psi$ является ортогональным. Иначе, по теореме, $\mathcal{A} = \mathcal{BC}$, где $\mathcal{B}$ - ортогональный оператор, а $\mathcal{C}$ - положительно определенный самосопряженный. Положим $\psi_1(P+\mathbf{x}) = P + \mathcal{B}(\mathbf{x})$ и $\psi_2(P+\mathbf{x}) = P + \mathcal{C}(\mathbf{x})$. Тогда

$$\psi_1\psi_2(P+\mathbf{x}) = \psi_1(P+\mathcal{C}(\mathbf{x})) = P + \mathcal{B}(\mathcal{C}(\mathbf{x})) = \psi,$$

т.е. $\psi = \psi_1\psi_2$, где ψ_1 - движение с неподвижной точкой P, а ψ_2 - растяжение пространства во взаимно ортогональных направлениях с разными коэффициентами.

Определение. *Определитель* $\det\varphi$ аффинного преобразования φ - это определитель $\det D\varphi$.

Замечание. Определитель движения равен ± 1.

Пусть $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ - ортонормированный базис в V. Тогда $\varphi(O + \mathbf{e}_i) = \varphi(O) + D\varphi(\mathbf{e}_i) = \varphi(O) + \mathbf{a}_i$, где $\mathbf{a}_i = D\varphi(\mathbf{e}_i)$. Оператор $A = D\varphi$ - невырожден, то $(\mathbf{a}_1, \ldots, \mathbf{a}_n)$- базис в пространстве в V. Если A- матрица оператора Aв базисе $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$, то A-- матрица перехода от $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ к $(\mathbf{a}_1, \ldots, \mathbf{a}_n)$. Получаем

$$\begin{pmatrix} (\mathbf{a}_1, \mathbf{a}_1) & \ldots & (\mathbf{a}_1, \mathbf{a}_n) \\ \vdots & \ddots & \vdots \\ (\mathbf{a}_n, \mathbf{a}_1) & \ldots & (\mathbf{a}_n, \mathbf{a}_n) \end{pmatrix} = A^\top \begin{pmatrix} (\mathbf{e}_1, \mathbf{e}_1) & \ldots & (\mathbf{e}_1, \mathbf{e}_n) \\ \vdots & \ddots & \vdots \\ (\mathbf{e}_n, \mathbf{e}_1) & \ldots & (\mathbf{e}_n, \mathbf{e}_n) \end{pmatrix} A = A^\top A.$$

Таким образом, $|\det \varphi| = |\det A|$ - число, показывающее во сколько увеличатся объемы.

Определение. Пусть V- n-мерное вещественное пространство, и $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$, $(\mathbf{e}'_1, \ldots, \mathbf{e}'_n)$ - два базиса в V. Базис $(\mathbf{e}'_1, \ldots, \mathbf{e}'_n)$ *ориентирован* как базис $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$, если для матрицы перехода Cвыполнено неравенство $\det C > 0$. В этом случае будем писать $(\mathbf{e}_1, \ldots, \mathbf{e}_n) \sim (\mathbf{e}'_1, \ldots, \mathbf{e}'_n)$.

Замечание. Отношение $\sim$ является отношением эквивалентности, так как

1) $(\mathbf{e}_1, \ldots, \mathbf{e}_n) \sim (\mathbf{e}_1, \ldots, \mathbf{e}_n)$;

2) если $(\mathbf{e}'_1, \ldots, \mathbf{e}'_n) \sim (\mathbf{e}_1, \ldots, \mathbf{e}_n)$, то $(\mathbf{e}_1, \ldots, \mathbf{e}_n) \sim (\mathbf{e}'_1, \ldots, \mathbf{e}'_n)$;

3) если $(\mathbf{e}''_1, \ldots, \mathbf{e}''_n) \sim (\mathbf{e}'_1, \ldots, \mathbf{e}'_n) \sim (\mathbf{e}_1, \ldots, \mathbf{e}_n)$, то $(\mathbf{e}''_1, \ldots, \mathbf{e}''_n) \sim (\mathbf{e}_1, \ldots, \mathbf{e}_n)$.

Пусть дан базис $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$. Рассмотрим другой базис $(\mathbf{e}'_1, \ldots, \mathbf{e}'_n) = (\mathbf{e}_1, \ldots, \mathbf{e}_n)C$, где $\det C < 0$. Докажем, что любой другой базис $(\mathbf{e}''_1, \ldots, \mathbf{e}''_n)$ будет ориентирован так же как $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ или $(\mathbf{e}'_1, \ldots, \mathbf{e}'_n)$. Пусть $(\mathbf{e}''_1, \ldots, \mathbf{e}''_n) = (\mathbf{e}'_1, \ldots, \mathbf{e}'_n)D$.

Если $\det D > 0$, то $(\mathbf{e}''_1, \ldots, \mathbf{e}''_n)$ ориентирован как $(\mathbf{e}'_1, \ldots, \mathbf{e}'_n)$.

Если $\det D < 0$, то $\det CD > 0$ и $(\mathbf{e}''_1, \ldots, \mathbf{e}''_n)$ ориентирован как $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$.

Задать ориентацию в V (в $\mathfrak{A}$) означает выбрать один из двух классов ориентации, он задает положительную ориентацию, а другой - отрицательную.

Пусть φ - аффинное преобразование в $(\mathfrak{A}, V)$ с дифференциалом $D\varphi = \mathcal{A}$, где V - n-мерное вещественное пространство. Рассмотрим произвольный базис $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ в V, тогда $(\mathcal{A}(\mathbf{e}_1), \ldots, \mathcal{A}(\mathbf{e}_n))$ - тоже базис в V. Если A -- матрица оператора $\mathcal{A}$ в базисе $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$, то A -- матрица перехода от базиса $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ к базису $(\mathcal{A}(\mathbf{e}_1), \ldots, \mathcal{A}(\mathbf{e}_n))$. Базисы $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ и $(\mathcal{A}(\mathbf{e}_1), \ldots, \mathcal{A}(\mathbf{e}_n))$ одинаково ориентированы тогда и только тогда, когда $\det\varphi = \det\mathcal{A} = \det A > 0$.

Если $\det\varphi = \det\mathcal{A} = \det A > 0$, то преобразование φ *сохраняет ориентацию*. Если $\det\varphi = \det\mathcal{A} = \det A < 0$, то преобразование φ *меняет ориентацию*.

Пусть базисы $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$, $(\mathbf{e}'_1, \ldots, \mathbf{e}'_n)$ противоположно ориентированы. И пусть $(\mathbf{e}_1(t), \ldots, \mathbf{e}_n(t))$ - такой базис, что $\mathbf{e}_i(0) = \mathbf{e}_i$ и $\mathbf{e}_i(1) = \mathbf{e}'_i$. Обозначим через $C(t)$ матрицу перехода от $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ к $(\mathbf{e}_1(t), \ldots, \mathbf{e}_n(t))$. Тогда

$C(0) = C$ и $C(1) = C$, причем $\det C < 0$. Так как $\det C(t)$ - непрерывная функция, то существует точка t_0 такая, что $\det C(t_0) = 0$. Тогда $(\mathbf{e}_1(t_0), \dots, \mathbf{e}_n(t_0))$ не является базисом.

Лемма. Пусть $\mathcal{A} : E \to E$ - ортогональный оператор в n-мерном евклидовом пространстве E, а $U = \{\mathbf{x} \in E \mid \mathcal{A}(\mathbf{x}) = \mathbf{x}\}$. Тогда $U^{\perp}$ инвариантно относительно $\mathcal{A}$, и оператор $\mathcal{A} - \mathcal{E}$ сюръективен на $U^{\perp}$.

Теорема. Любое движение $\varphi : \mathfrak{A} \to \mathfrak{A}$ аффинно евклидова пространства $(\mathfrak{A}, E)$ может быть разложено в произведение $\varphi = \tau_{\mathbf{a}} \psi$, где движение ψ имеет некоторую неподвижную точку, а $\tau_{\mathbf{a}}$ - такой сдвиг на вектор $\mathbf{a}$, что $D\varphi(\mathbf{a}) = \mathbf{a}$.

Определение. Движение называется *собственным*, если $\det \varphi = 1$. Движение называется *несобственным*, если $\det \varphi = -1$.

Рассмотрим движение φ в пространстве $\mathbb{R}^n$, где $n = 1, 2, 3$.

(I) Пусть $n = 1$, т.е. имеем движение на прямой. Следовательно, $D\varphi = \pm \mathcal{E}$

1) Если φ - собственное движение, то φ - это сдвиг вдоль прямой.

2) Если φ - несобственное движение, т.е. $D\varphi = -\mathcal{E}$, то, из теоремы, $U = \{0\}$, $\tau_{\mathbf{0}}$ - сдвиг на нулевой вектор и $\varphi = \psi$ имеет неподвижную точку P. Таким образом, $\varphi(P + \mathbf{x}) = P + D\varphi(\mathbf{x}) = P - \mathbf{x}$ - симметрия относительно точки.

(II) Пусть $n = 2$, т.е. движение на плоскости.

1) Если φ - собственное движение и $D\varphi = \mathcal{E}$, то φ - сдвиг.

2) Если φ - собственное движение и $D\varphi \neq \mathcal{E}$, то выберем канонический ортонормированный базис $(\mathbf{e}_1, \mathbf{e}_2)$, в котором матрица оператора $D\varphi$ имеет вид

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}, \qquad \varphi \neq 0.$$

Из теоремы, $U = \{0\}$, $\tau_{\mathbf{0}}$ - сдвиг на нулевой вектор и $\varphi = \psi$ имеет неподвижную точку P. Таким образом, $\varphi(P + \mathbf{x}) = P + D\varphi(\mathbf{x})$ - поворот на некоторый угол вокруг неподвижной точки.

3) Если φ - несобственное движение, то выберем канонический ортонормированный базис $(\mathbf{e}_1, \mathbf{e}_2)$, в котором матрица оператора $D\varphi$ имеет вид

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Из теоремы, $U = \mathrm{Lin}(\mathbf{e}_1)$, $\varphi = \tau_{\mathbf{a}}\psi$, где $\tau_{\mathbf{a}}$ - сдвиг на вектор $\mathbf{a} = \alpha\mathbf{e}_1$ и ψ имеет неподвижную точку P. Таким образом, для любого вектора $\mathbf{x} = x^1\mathbf{e}_1 + x^2\mathbf{e}_2$ имеем $\varphi(P + \mathbf{x}) = \psi(P + \mathbf{x}) + \mathbf{a} = P + D\varphi(\mathbf{x}) + \mathbf{a} = P + x^1\mathbf{e}_1 - x^2\mathbf{e}_2 + \alpha\mathbf{e}_1$

- отражение относительно некоторой прямой со сдвигом вдоль этой прямой (*скользящая симметрия*).

(III) Пусть $n = 3$, т.е. движение в пространстве.

1) Если φ -- собственное движение и $D\varphi = \mathcal{E}$, то φ -- сдвиг.

2) Если φ -- собственное движение и $D\varphi \neq \mathcal{E}$, то выберем канонический ортонормированный базис $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, в котором матрица оператора $D\varphi$ имеет вид

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}, \qquad \varphi \neq 0.$$

Из теоремы, $U = \mathrm{Lin}(\mathbf{e}_1)$, $\varphi = \tau_{\mathbf{a}}\psi$, где $\tau_{\mathbf{a}}$ - сдвиг на вектор $\mathbf{a} = \alpha\mathbf{e}_1$ и ψ имеет неподвижную точку P. Таким образом, для любого вектора $\mathbf{x} = x^1\mathbf{e}_1 + x^2\mathbf{e}_2 + x^3\mathbf{e}_3$ имеем $\varphi(P + \mathbf{x}) = \psi(P + \mathbf{x}) + \mathbf{a} = P + D\varphi(\mathbf{x}) + \mathbf{a}$, где прямая $P + x^1\mathbf{e}_1$ состоит из неподвижных точек, - поворот вокруг некоторой прямой с последующим сдвигом вдоль этой прямой (*винтовое движение*).

3) Если φ - несобственное движение с неподвижной точкой P, то выберем канонический ортонормированный базис $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, в котором матрица оператора $D\varphi$ имеет вид

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Пусть $\varphi = \varphi_1\varphi_2$, где φ_1 и φ_2 имеют неподвижную точку и оператор $D\varphi_1$ в базисе $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ имеет вид

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix},$$

а оператор $D\varphi_2$ -

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Для любого вектора $\mathbf{x} = x^1\mathbf{e}_1 + x^2\mathbf{e}_2 + x^3\mathbf{e}_3$ имеем $\varphi_2(P+\mathbf{x}) = P + D\varphi(\mathbf{x}) = P - x^1\mathbf{e}_1 + x^2\mathbf{e}_2 + x^3\mathbf{e}_3$

- зеркальная симметрия относительно плоскости (для φ_2 $U = \mathrm{Lin}(\mathbf{e}_1, \mathbf{e}_2)$). Таким образом, φ- зеркальная симметрия относительно плоскости и поворот относительно прямой ортогональной плоскости.

4) Если φ - несобственное движение без неподвижных точек, то, из теоремы, $\varphi = \tau_{\mathbf{a}}\psi$, где ψ имеет неподвижную точку, следовательно $U \neq \{0\}$. Получаем, что $D\varphi$ имеет собственное значение 1. Тогда можно выбрать канонический ортонормированный базис $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, в котором матрица оператора $D\varphi$ имеет вид

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Имеем, $U = \mathrm{Lin}(\mathbf{e}_1, \mathbf{e}_2)$ и для любого вектора $\mathbf{x} = x^1\mathbf{e}_1 + x^2\mathbf{e}_2 + x^3\mathbf{e}_3$

$$\varphi(P + \mathbf{x}) = P + D\varphi(\mathbf{x}) + \mathbf{a} = P + x^1\mathbf{e}_1 + x^2\mathbf{e}_2 - x^3\mathbf{e}_3 + \mathbf{a}$$

-- зеркальная симметрия относительно плоскости с последующим сдвигом вдоль этой плоскости (*скользящая симметрия*).

5.3. Аффинные квадратичные функции и квадрики

Пусть $(\mathfrak{A}, V)$ -- аффинное пространство, где V -- линейное пространство над полем F.

Определение. *Аффинной квадратичной функцией* называется отображение $\varphi : \mathfrak{A} \to F$, для которого

1) существует точка $O \in \mathfrak{A}$,

2) существует квадратичная функция $q : V \to F$,

3) существует линейная функция $f : V \to F$, такие, что для любой точки $P \in \mathfrak{A}$ справедливо равенство $\varphi(P) = q(\overrightarrow{OP}) + f(\overrightarrow{OP}) + \varphi(O)$

.

Пусть задана система координат $\left(O, (\mathbf{e}_1, \ldots, \mathbf{e}_n)\right)$, и $(x^1, \ldots, x^n)$ - координаты точки P, т.е. координаты вектора $\overrightarrow{OP}$ в базисе $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$, $A = (a_{ij})_{n \times n}$ - матрица квадратичной функции q в базисе $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ и $f(\mathbf{e}_i) = b_i$. Получаем

$$\varphi(P) = a_{ij} x^i x^j + b_i x^i + c,$$

где $c = \varphi(O)$.

Перейдем к системе координат с новым центром $\left(O', (\mathbf{e}_1, \ldots, \mathbf{e}_n)\right)$, и пусть $(x_0^1, \ldots, x_0^n)$ - координаты O' в системе $\left(O, (\mathbf{e}_1, \ldots, \mathbf{e}_n)\right)$. Тогда $\overrightarrow{OP} = \overrightarrow{OO'} + \overrightarrow{O'P}$, т.е. $x^i = x_0^i + \widetilde{x}^i$. Получаем

$$\varphi(P) = a_{ij}(x_0^i + \widetilde{x}^i)(x_0^j + \widetilde{x}^j) + b_i(x_0^i + \widetilde{x}^i) + c = a_{ij}\widetilde{x}^i\widetilde{x}^j + 2a_{ij}\widetilde{x}^i x_0^j + a_{ij}x_0^i x_0^j + b_i\widetilde{x}^i + b_i x_0^i + c =$$

$$= a_{ij}\widetilde{x}^i\widetilde{x}^j + \widetilde{b}_i\widetilde{x}^i + \widetilde{c},$$

где $\tilde{b}_i = b_i + 2a_{ij}x_0^j$, $\tilde{c} = c + b_i x_0^i + a_{ij}x_0^i x_0^j = \varphi(O')$.

Получаем, что чисто квадратичная функция q не зависит от выбора точки O и $\varphi(P) = q(\overrightarrow{O'P}) + f'(\overrightarrow{O'P}) + \varphi(O')$.

Определение. Точка O' называется *центральной* для аффинной квадратичной функции φ, если для любого вектора $\mathbf{x} \in V$ справедливо равенство $\varphi(O' + \mathbf{x}) = \varphi(O' - \mathbf{x})$. *Центр* аффинной квадратичной функции - это множество всех ее центральных точек.

Теорема. Если центр аффинной квадратичной функции φ на n-мерном аффинном пространстве не пуст, то он является плоскостью размерности $n - \mathrm{rank} q$. В случае, когда чисто квадратичная часть q невырождена, для φ имеется ровно одна центральная точка.

Замечание. 1) Пусть дана квадратичная функция $\varphi(x^1, \ldots, x^n)$ от координат точки P. Чтобы найти центр нужно решить систему уравнений $\left.\frac{\partial\varphi}{\partial x^i}\right|_{(x_0^1, \ldots, x_0^n)} = 0$.

2) Если O' -- центральная точка, то $\varphi(O' \pm \mathbf{x}) = q(\mathbf{x}) + \varphi(O')$.

3) Пусть O_1 и O_2 - две центральные точки. Тогда $\varphi(O_2) = \varphi(O_1 + \overrightarrow{O_1O_2}) = q(\overrightarrow{O_1O_2}) + \varphi(O_1)$ и $\varphi(O_1) = \varphi(O_2 - \overrightarrow{O_1O_2}) = q(\overrightarrow{O_1O_2}) + \varphi(O_2)$.

Следовательно, $\varphi(O_1) = \varphi(O_2)$ и для любого вектора $\mathbf{a} \in \mathrm{Lin}(\overrightarrow{O_1O_2})$ верны равенства $q(\mathbf{a}) = 0$ и $\varphi(O_1 + \mathbf{a}) = \varphi(O_1)$. Получаем, что прямая, проходящая через O_1 и O_2, состоит из центральных точек.

Определение. Аффинно квадратичная функция называется *центральной*, если ее центр не пуст.

Теорема. Для всякой аффинной квадратичной функции φ в n-мерном аффинном пространстве существует каноническая система координат $\left(O, (\mathbf{e}_1, \ldots, \mathbf{e}_n)\right)$, в которой

1) $\varphi(P) = (x^1)^2 + \ldots + (x^k)^2 - (x^{k+1})^2 - \ldots - (x^{k+l})^2 + c$, если φ является центральной,

2) $\varphi(P) = (x^1)^2 + \ldots + (x^k)^2 - (x^{k+1})^2 - \ldots - (x^{k+l})^2 + x^{k+l+1}$, если φ не является центральной. Здесь $(x^1, \ldots, x^n)$ -- координаты точки P. Канонический вид аффинной квадратичной функции не зависит от выбора канонического базиса.

Теорема. Для всякой аффинной квадратичной функции φ в n-мерном евклидовом пространстве $(\mathfrak{A}, E)$ существует каноническая прямоугольная система координат $(O, (\mathbf{e}_1, \ldots, \mathbf{e}_n))$, в которой

1) $\varphi(P) = \sum_{i=1}^{k+l} \lambda_i (x^i)^2 + c$, если φ является центральной,

2) $\varphi(P) = \sum_{i=1}^{k+l} \lambda_i (x^i)^2 + \mu x^{k+l+1}$, если φ не является центральной. Здесь $\lambda_1, \ldots, \lambda_k > 0$ и $\lambda_{k+1}, \ldots, \lambda_{k+l} < 0$.

Определение. Пусть φ - аффинно квадратичная функция. *Квадрика* - это множество всех таких точек $P \in \mathfrak{A}$, что $\varphi(P) = 0$. Квадрика называется *центральной*, если она задается центральной квадратичной функцией.

Теорема. Для всякой квадрики в вещественном n-мерном аффинном пространстве существует подходящая система координат, в которой квадрика имеет уравнение одного из следующего типов:

$(I)\ a)$ $(x^1)^2 + \ldots + (x^k)^2 - (x^{k+1})^2 - \ldots - (x^{k+l})^2 = 1$, $k, l \geqslant 0$, $k + l \leqslant n$, или

$b)$ $(x^1)^2 + \ldots + (x^k)^2 - (x^{k+1})^2 - \ldots - (x^{k+l})^2 = 0$, $k \geqslant l$, если квадрика является центральной;

(II) $(x^1)^2 + \ldots + (x^k)^2 - (x^{k+1})^2 - \ldots - (x^{k+l})^2 + x^{k+l+1} = 0$, $k \geqslant l$, $k + l < n$, если квадрика не является центральной.

Классификация.

$(I)a)$

- если $k = n$, то эллипсоид,

- если $l = n$, то мнимый эллипсоид,

- если $k, l > 0$ и $k + l = n$, то гиперболоиды,

- если $k + l < n$, то цилиндры;

$(I)b)$

- если $k + l = n$, то конусы,

- если $k+l<n$, то конические цилиндры;

(II)

- если $k+l+1=n$, то параболоиды ($l=0$ - эллиптические, иначе - гиперболические),

- если $k+l+1<n$, то параболический цилиндр.

Теорема. Для всякой квадрики в вещественном n-мерном евклидовом аффинном пространстве существует каноническая прямоугольная система координат, в которой квадрика задается уравнением одного из следующего типов:

(I) $a)$

$$\lambda_1(x^1)^2+\ldots+\lambda_k(x^k)^2-\lambda_{k+1}(x^{k+1})^2-\ldots-\lambda_{k+l}(x^{k+l})^2=1,$$

$k,\, l\geqslant 0$, $k+l\leqslant n$, или

$b)$ $\lambda_1(x^1)^2+\ldots+\lambda_k(x^k)^2-\lambda_{k+1}(x^{k+1})^2-\ldots-\lambda_{k+l}(x^{k+l})^2=0$,

$k\geqslant l$, если квадрика является центральной;

(II)

$$\lambda_1(x^1)^2+\ldots+\lambda_k(x^k)^2-\lambda_{k+1}(x^{k+1})^2-\ldots-\lambda_{k+l}(x^{k+l})^2+x^{k+l+1}=0,$$

$k\geqslant l$, $k+l<n$, если квадрика не является центральной. Здесь все

5.4. Проективное пространство

Пусть $(\mathfrak{A}, V)$ - $(n+1)$-мерное аффинное пространство, и $O \in \mathfrak{A}$ - произвольная точка.

Определение. n-*мерным проективным пространством* $P(V)$ называется множество всех прямых в аффинном пространстве $(\mathfrak{A}, V)$, проходящих через точку O, т.е. точка в $P(V)$ - это прямая в $(\mathfrak{A}, V)$, проходящая через O.

Пусть $X \in P(V)$ - точка в $P(V)$. Тогда $X = O + \mathrm{Lin}(\mathbf{x})$, $\mathbf{x} \in V$, в $(\mathfrak{A}, V)$. Будем писать $X = \tilde{\mathbf{x}}$, причем $\tilde{\mathbf{x}} = \widetilde{\lambda \mathbf{x}}$ для $\lambda \neq 0$, и $\tilde{\mathbf{x}} = \tilde{\mathbf{y}}$ тогда и только тогда, когда $\mathbf{x} = \lambda \mathbf{y}$ для некоторого $\lambda \neq 0$.

Пусть $(\mathbf{e}_0, \ldots, \mathbf{e}_n)$ - базис в V, и $\mathbf{x} \neq \mathbf{0}$ - произвольный вектор в V. Тогда $\mathbf{x} = \xi^i \mathbf{e}_i$ и набор $(\xi^0 : \xi^1 : \ldots : \xi^n)$ - это *однородные координаты* точки $X = \tilde{\mathbf{x}}$. Пусть $(\mathbf{e}'_0, \ldots, \mathbf{e}'_n) = (\mathbf{e}_0, \ldots, \mathbf{e}_n) C$ другой базис в V, и $\mathbf{x} = \zeta^j \mathbf{e}'_j$. Тогда

$$\begin{cases} \lambda\xi^0 = c^0_j\zeta^j \\ \dots \\ \lambda\xi^n = c^n_j\zeta^j. \end{cases}$$

Определение. Пусть $k \geqslant 0$ и $U \subset V$ - $(k+1)$-мерное подпространство. Тогда все точки $X = O + \mathrm{Lin}(\mathbf{x}) = \tilde{\mathbf{x}}$ для всех $\mathbf{x} \in U$ составляют k*-мерную плоскость в* $P(V)$.

Теорема. Пусть плоскости Π_1 и Π_2 в $P(V)$ имеют размерности k и l, причем $k + l \geqslant n = \dim P(V)$. Тогда $\Pi_1 \cap \Pi_2 \neq \emptyset$.

Пусть $\mathcal{A}\colon V \to V$ - невырожденный линейный оператор на V. Определим оператор $\tilde{\mathcal{A}}\colon P(V) \to P(V)$ положив $\tilde{\mathcal{A}}(\tilde{\mathbf{x}}) = \widetilde{\mathcal{A}(\mathbf{x})}$ для любой точки $\tilde{\mathbf{x}} \in P(V)$, где $V \ni \mathbf{x} \neq \mathbf{0}$.

Корректность.

1) $\mathcal{A}(\mathbf{x}) \neq \mathbf{0}$, так как $\mathrm{Ker}\mathcal{A} = \{\mathbf{0}\}$.

2) Если $\tilde{\mathbf{x}} = \widetilde{\lambda\mathbf{x}}$, то $\tilde{\mathcal{A}}(\widetilde{\lambda\mathbf{x}}) = \widetilde{\mathcal{A}(\lambda\mathbf{x})} = \widetilde{\lambda\mathcal{A}(\mathbf{x})} = \widetilde{\mathcal{A}(\mathbf{x})}$.

Определение. *Проективным преобразованием* в $P(V)$ называется преобразование $\widetilde{\mathcal{A}}$, заданное с помощью некоторого невырожденного линейного оператора $\mathcal{A}$ на V по формуле $\widetilde{\mathcal{A}}(\tilde{\mathbf{x}}) = \widetilde{\mathcal{A}(\mathbf{x})}$ для любой точки $\tilde{\mathbf{x}} \in P(V)$. Если $\mathcal{A} = \mathcal{E}$, то $\widetilde{\mathcal{E}}$ -- тождественное преобразование в $P(V)$.

Свойства операторов на $P(V)$.

1) Рассмотрим два оператора $\mathcal{A}: V \to V$ и $\lambda\mathcal{A}: V \to V$ в V.

Тогда $\widetilde{\mathcal{A}} = \widetilde{\lambda\mathcal{A}}$, так как $\widetilde{\lambda\mathcal{A}}(\tilde{\mathbf{x}}) = \widetilde{\lambda\mathcal{A}(\mathbf{x})} = \widetilde{\mathcal{A}(\mathbf{x})} = \widetilde{\mathcal{A}}(\tilde{\mathbf{x}})$.

2) Рассмотрим два оператора $\mathcal{A}: V \to V$ и $\mathcal{B}: V \to V$ в V.

Тогда $\widetilde{\mathcal{A}}\widetilde{\mathcal{B}} = \overline{\mathcal{A}\mathcal{B}}$, так как $(\widetilde{\mathcal{A}}\widetilde{\mathcal{B}})(\tilde{\mathbf{x}}) = \widetilde{\mathcal{A}}(\widetilde{\mathcal{B}}(\tilde{\mathbf{x}})) = \widetilde{\mathcal{A}}(\widetilde{\mathcal{B}(\mathbf{x})}) = \widetilde{\mathcal{A}(\mathcal{B}(\mathbf{x}))} = (\widetilde{\mathcal{A}\mathcal{B})(\mathbf{x}}) = (\widetilde{\mathcal{A}\mathcal{B}})(\tilde{\mathbf{x}})$.

3) Пусть для оператора $\mathcal{A}: V \to V$ существует $\mathcal{A}^{-1}: V \to V$.

Тогда для оператора $\widetilde{\mathcal{A}}: P(V) \to P(V)$ существует $\widetilde{\mathcal{A}}^{-1}: P(V) \to P(V)$, равный $\widetilde{\mathcal{A}}^{-1} = \overline{\mathcal{A}^{-1}}$, так как $\widetilde{\mathcal{A}}\overline{\mathcal{A}^{-1}} = \widetilde{E}$ и $\overline{\mathcal{A}^{-1}}\widetilde{\mathcal{A}} = \widetilde{E}$.

4) Пусть $0 \neq \mathbf{x} = \xi^i \mathbf{e}_i \in V$, и $A = (a^i_j)_{(n+1)\times(n+1)}$ - матрица оператора в базисе $(\mathbf{e}_0, \ldots, \mathbf{e}_n)$. Тогда $\mathbf{y} = \mathcal{A}(\mathbf{x}) = \eta^i \mathbf{e}_i$ и однородные координаты $\widetilde{\mathbf{y}}$ и $\widetilde{\mathbf{x}}$ связаны формулой

$$\rho \begin{pmatrix} \eta^0 \\ \vdots \\ \eta^n \end{pmatrix} = A \begin{pmatrix} \xi^0 \\ \vdots \\ \xi^n \end{pmatrix}.$$

Определение. Точки $A_0, A_1, \ldots, A_{n+1} \in P(V)$ *находятся в общем положении*, если точки $(A_0, A_1, \ldots, A_{i-1}, \widehat{A_i}, A_{i+1}, \ldots, A_{n+1})$ не лежат в одной $(n-1)$-мерной плоскости в $P(V)$.

Теорема. Пусть $(A_0, A_1, \ldots, A_{n+1})$ и $(B_0, B_1, \ldots, B_{n+1})$ - две системы точек общего положения в n-мерном проективном пространстве. Тогда существует единственное проективное преобразование $\widetilde{\mathcal{A}}\colon P(V) \to P(V)$ такое, что $\widetilde{\mathcal{A}}(A_i) = B_i$.

Пусть A_1, A_2, A_3, A_4 - различные точки на проективной прямой, т.е. $A_i = \widetilde{\mathbf{a}_i}$ и $\mathbf{a}_i \in U$, где $\dim U = 2$. Любые два вектора линейно независимы, пусть это $\mathbf{a}_1$ и $\mathbf{a}_2$. Тогда $\mathbf{a}_3 = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2$ и $\mathbf{a}_4 = \beta_1 \mathbf{a}_1 + \beta_2 \mathbf{a}_2$. Так как точки A_1, A_2, A_3, A_4 различны, то все α_i и $\beta_j \neq 0$.

Определение. *Двойным отношением* (A_1, A_2, A_3, A_4) точек называется $\frac{\alpha_2 \beta_1}{\alpha_1 \beta_2}$.

Корректность. Если $\mathbf{a}_3$ заменить на $\lambda \mathbf{a}_3$, то α_1 и α_2 перейдут в $\lambda\alpha_1$ и $\lambda\alpha_2$ соответственно, и ничего не изменится. Аналогично для $\mathbf{a}_4$.

Если $\mathbf{a}_1$ заменить на $\lambda \mathbf{a}_1$, то α_1 и β_1 перейдут в α_1/λ и β_1/λ соответственно, и ничего не изменится. Аналогично для $\mathbf{a}_2$.

Теорема. Под действием произвольного проективного преобразования $\widetilde{A}$ любая k-мерная плоскость переходит в k-мерную плоскость. В частности, прямые переходят в прямые, а для различных четырех точек A_1, A_2, A_3, A_4, лежащих на одной прямой справедливо равенство

$$(\widetilde{A}(A_1), \widetilde{A}(A_2), \widetilde{A}(A_3), \widetilde{A}(A_4)) = (A_1, A_2, A_3, A_4).$$

Пусть $(\mathfrak{A}, V)$ - аффинное пространство, и $O \in \mathfrak{A}$ - произвольная точка. Выберем в $\mathfrak{A}$ систему координат $(O, (\mathbf{e}_0, \ldots, \mathbf{e}_n))$. Пусть $O_0 = O + \mathbf{e}_0$, $U_0 = \mathrm{Lin}(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ и $\Pi_0 = O_0 + U_0$ - гиперплоскость в $\mathfrak{A}$. Эта плоскость называется *стандартной аффинной картой*. Гиперплоскость $P(U_0)$ в $P(V)$ называется *бесконечно удаленной гиперплоскостью* по отношению к аффинной карте Π_0.

Определение. *Изображением* точки $P(V) \ni \tilde{\mathbf{x}} = O + \mathrm{Lin}(\mathbf{x})$ на аффинной карте называется точка пересечения в пространстве $\mathfrak{A}$ прямой $O + \mathrm{Lin}(\mathbf{x})$ с гиперплоскостью Π_0.

Теорема. Точка $\tilde{\mathbf{x}}$ проективного пространства $P(V)$ имеет изображение на аффинной карте Π_0 тогда и только тогда, когда ее однородная координата $\xi^0 \neq 0$ ($\tilde{\mathbf{x}}$ имеет координаты $(\xi^0 : \ldots \xi^n)$). Если $\xi_0 \neq 0$, то координаты изображения точки $\tilde{\mathbf{x}}$ на аффинной карте суть $\left(\frac{\xi^1}{\xi^0}, \ldots, \frac{\xi^n}{\xi^0}\right)$.

$$x^1 = \frac{\xi^1}{\xi^0}, \ldots, x^n = \frac{\xi^n}{\xi^0}$$

Определение. Числа называются *неоднородными координатами* точки $\tilde{\mathbf{x}}$.

Пусть $A_i = \tilde{\mathbf{a}}_i$, $i = 1, 2, 3, 4$, - различные точки на проективной прямой, и x_i - их неоднородные координаты. Тогда вектора $\mathbf{a}_i$ имеют координаты $(1, x_i)$, и из равенств $(1, x_3) = \alpha_1(1, x_1) + \alpha_2(1, x_2)$, $(1, x_4) = \beta_1(1, x_1) + \beta_2(1, x_2)$ следует, что $\alpha_1 = \frac{x_3 - x_2}{x_1 - x_2}$, $\alpha_2 = \frac{x_3 - x_1}{x_2 - x_1}$, $\beta_1 = \frac{x_4 - x_2}{x_1 - x_2}$, $\beta_2 = \frac{x_4 - x_1}{x_2 - x_1}$. Таким образом, $(A_1, A_2, A_3, A_4) = \frac{x_3 - x_1}{x_3 - x_2} : \frac{x_4 - x_1}{x_4 - x_2}$.

Пусть $\tilde{\mathcal{A}}$ проективное преобразование, и $\tilde{\mathbf{y}} = \tilde{\mathcal{A}}(\tilde{\mathbf{x}}) = \widetilde{\mathcal{A}(\mathbf{x})}$.

Пусть $(\xi^0 : \ldots : \xi^n)$ и $(\eta^0 : \ldots : \eta^n)$ - однородные координаты для $\widetilde{\mathbf{x}}$ и $\widetilde{\mathbf{y}}$ соответственно. Обозначим $x^i = \dfrac{\xi^i}{\xi^0}$ и $y^j = \dfrac{\eta^j}{\eta^0}$. Тогда

$$y^j = \frac{\eta^j}{\eta^0} = \frac{a_i^j \xi^i}{a_i^0 \xi^i} = \frac{a_0^j + a_i^j x^i}{a_0^0 + a_i^0 x^i}.$$

Пусть $\Pi = P(U)$, где $\dim U = k+1$. Так как $\widetilde{\mathbf{x}} \in \Pi$ тогда и только тогда, когда $\mathbf{x} \in U$, то U, а значит и Π. задается однородной системой ранга $n-k$ относительно однородных координат $(\xi^0 : \ldots : \xi^n)$.

Определение. *Квадрикой* в проективном пространстве $P(V)$ называется множество всех точек, однородные координаты $(\xi^0 : \ldots : \xi^n)$ которых в некоторой системе координат удовлетворяют уравнению вида $\sum_{i,j=0}^{n} a_{ij}\xi^i\xi^j = 0$.

Теорема. Для всякой квадрики в n-мерном проективном пространстве существует каноническая система координат, в которой квадрика задается уравнением вида $(\xi^0)^2 + \ldots + (\xi^k)^2 - (\xi^{k+1})^2 - \ldots - (\xi^{k+l})^2 = 0$, где $k + l \leqslant n$ и $k + 1 \geqslant l$.

(I) Если $k = n$ и $l = 0$, то нет точек.

(II) Если $k < n$ и $l = 0$, то плоскость.

$\lambda_i > 0$

.

6. Евклидовы пространства и выпуклые множества

6.1. Определение евклидова пространства. Ортонормированные базисы

Понятие n-мерного линейного пространства далеко не в полной мере обобщает понятие плоскости или трехмерного евклидового пространства — в n-мерном случае при n>3 не определены ни длина вектора, ни угол между векторами, и потому невозможно развитие той богатой геометрической теории, которая хорошо знакомая читателю для n=2 и n=3. Оказывается, что это положение может быть исправлено следующим образом.

Из курса аналитической геометрии известно, что и в плоскости, и в трехмерном пространстве можно ввести понятие скалярного умножения векторов. Оно определяется при помощи длин векторов и угла между ними, но, как оказывается, и длина вектора, и угол между векторами в свою очередь могут быть выражены через скалярные произведения. Мы определим поэтому в любом n-мерном линейном пространстве понятие скалярного умножения, причем определим аксиоматически, с помощью некоторых свойств, которыми, как хорошо известно, скалярное умножение векторов плоскости или трехмерного пространства обладает. При этом, учитывая те непосредственные цели, ради которых этот раздел включен в курс дискретной математики, вводить определения длины вектора и угла между векторами мы не станем. Читателя, который интересуется построением геометрии в n-мерных пространствах, мы отсылаем к специальной литературе.

Отметим, что всюду в этом разделе, кроме конца этого раздела, рассматриваются *действительные* линейные пространства.

Будем говорить, что в *n*-мерном действительном линейном пространстве M_n определено *скалярное умножение,* если всякой паре векторов ***a, b*** поставлено в соответствие действительное число, которое обозначается символом ***(a, b)*** и названное *скалярным произведением* векторов ***a*** и ***b***, причем выполняются следующие условия (здесь ***a, b, c*** - любые векторы пространства M_n, α - любое действительное число):

I. $(\boldsymbol{a}, \boldsymbol{b})=(\boldsymbol{b}, \boldsymbol{a})$.

II. $(\boldsymbol{a}+\boldsymbol{b}, \boldsymbol{c})= (\boldsymbol{a}, \boldsymbol{c})+(\boldsymbol{b}, \boldsymbol{c})$.

III. $(\alpha\boldsymbol{a}, \boldsymbol{b})= \alpha(\boldsymbol{a}, \boldsymbol{b})$.

IV. Если α≠0, то скалярный квадрат вектора ***a*** строго положителен,

$$(\boldsymbol{a}, \boldsymbol{a})>0.$$

Отметим, что из III при α=0 следует равенство

$$(0, \boldsymbol{b})=0, \tag{1}$$

т.е. *скалярное произведение нулевого вектора на любой вектор* ***b*** *равно нулю;* равен нулю, в частности, скалярный квадрат нулевого вектора.

Из II и III вытекает следующая формула для скалярного произведения линейных комбинаций двух систем векторов:

$$\left(\sum_{i=1}^{k}\alpha_i\boldsymbol{a}_i, \sum_{j=1}^{l}\beta_j\boldsymbol{b}_j\right)=\sum_{i=1}^{k}\sum_{j=1}^{l}\alpha_i\beta_j(\boldsymbol{a}_i, \boldsymbol{b}_j) \tag{2}$$

Если в *n*-мерном линейном пространстве определено скалярное умножение, то это пространство называется *n*-мерным *евклидовым пространством.*

При любом n в n-мерном линейном пространстве M_n можно определить скалярное умножение, т.е. можно превратить это пространство в евклидово.

В самом деле, возьмем в пространстве M_n любой базис $\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n$. Если

$$\boldsymbol{a}=\sum_{i=1}^{n}\alpha_i\boldsymbol{e}_i \ , \qquad \boldsymbol{b}=\sum_{i=1}^{n}\beta_i\boldsymbol{e}_i \ ,$$

то положим

$$(\boldsymbol{a}, \boldsymbol{b}) = \sum_{i=1}^{n}\alpha_i\beta_i \ . \tag{3}$$

Легко проверяется, что условия 1 — IV будут выполнены, т.е. равенство (1) определяет в просторные M_n скалярное умножение. Мы видим, что в *n*-мерном линейном пространстве скалярное умножение можно задать, вообще говоря, многими различными способами — определение (3) зависит от выбора базиса, а мы пока не знаем, нельзя ли ввести скалярное умножение и каким-либо принципиально иным способом. Нашей ближайшей целью является обозрение всех возможных способов превращения n-мерного линейного пространства в евклидово пространство и установление того, что в некотором смысле для всякого n существует одно-единственное n-мерное евклидово пространство.

Пусть дано произвольное n-мерное евклидово пространство E_n, т.е. в n-мерном линейном пространстве произвольным способом введено скалярное умножение. Векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ называются *ортогональными,* если их скалярное произведение равно нулю,

$$(\boldsymbol{a}, \boldsymbol{b}) = 0.$$

Из (1) следует, что нулевой вектор ортогонален к любому вектору; могут существовать, однако, и ненулевые ортогональные векторы.

Система векторов называется *ортогональной системой,* если все векторы этой системы попарно ортогональны между собой.

Всякая ортогональная система ненулевых векторов линейно независима.

Пусть, в самом деле, в E_n данная система векторов $\boldsymbol{a_1}, \boldsymbol{a_2}, ..., \boldsymbol{a_k}$, причем $\boldsymbol{a}_i \neq 0$, $i = 1, 2, ..., k$, и

$$(\boldsymbol{a}_i, \boldsymbol{a}_j) = 0 \quad \text{при} \quad i \neq j. \tag{4}$$

Если

$$\alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2 + ... + \alpha_k\boldsymbol{a}_k = 0,$$

то, скалярно умножая обе части этого равенства на вектор $\boldsymbol{a}_i$ $1 \leq i \leq k$, получаем ввиду (1), (2) и (4):

$$0 = (0, \boldsymbol{a}_i) = (\alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2 + ... + \alpha_k a_k, \boldsymbol{a}_i) =$$
$$= \alpha_1(\boldsymbol{a}_1, \boldsymbol{a}_i) + \alpha_2 (\boldsymbol{a}_2, \boldsymbol{a}_i) + ... + \alpha_k(\boldsymbol{a}_k, \boldsymbol{a}_i) = \alpha_i(\boldsymbol{a}_i, \boldsymbol{a}_i).$$

Отсюда, так как $(\boldsymbol{a}_i, \boldsymbol{a}_i) > 0$ по IV, вытекает $\alpha_i = 0$, $i = 1, 2, ..., k$, что и требовалось доказать.

Опишем процесс ортогонализации, т.е. некоторый способ перехода от любой линейно независимой системы из k векторов

$$\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_k \tag{5}$$

евклидова просттранства E_n к ортогональной системе, которая также состоит из k ненулевых векторов; эти векторы будут обозначены через $\boldsymbol{b}_1, \boldsymbol{b}_2, ..., \boldsymbol{b}_k$.

Положим $\boldsymbol{b}_1 = \boldsymbol{a}_1$, т.е. первый вектор системы (5) войдет и в строящуюся нами ортогональную систему. Положим, далее,

$$\boldsymbol{b}_2 = \alpha_1\boldsymbol{b}_1 + \boldsymbol{a}_2.$$

Так как $\boldsymbol{b}_1 = \boldsymbol{a}_1$, а векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ линейно независимы, то вектор $\boldsymbol{b}_2$ отличен от нуля при любом числе α_1. Подберем это число из условия, которое вектор $\boldsymbol{b}_2$ должен быть ортогонален к вектору $\boldsymbol{b}_1$:

$$0 = (\boldsymbol{b}_1, \boldsymbol{b}_2) = (\boldsymbol{b}_1, \alpha_1\boldsymbol{b}_1 + \boldsymbol{a}_2) = \alpha_1(\boldsymbol{b}_1, \boldsymbol{b}_1) + (\boldsymbol{b}_1, \boldsymbol{a}_2),$$

откуда, ввиду IV,

$$\alpha_1 = -(\boldsymbol{b}_1, \boldsymbol{a}_2)/(\boldsymbol{b}_1, \boldsymbol{b}_1)$$

Пусть уже построена ортогональная система ненулевых векторов $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_l$; дополнительно предположим, что для всякого i, $1 \le i \le l$ вектор $\boldsymbol{b}_i$ является линейной комбинацией векторов $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_i$ Это предположение будет выполняться тогда и для вектора $\boldsymbol{b}_{l+1}$, если он будет избран в виде

$$\boldsymbol{b}_{l+1} = \alpha_1 \boldsymbol{b}_1 + \alpha_2 \boldsymbol{b}_2 + \ldots + \alpha_l \boldsymbol{b}_l + \boldsymbol{a}_{l+1},$$

Вектор $\boldsymbol{b}_{l+1}$ будет при этом отличен от нуля, так как система (5) линейно независима, а вектор $\boldsymbol{a}_{l+1}$ не входит в записи векторов $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_l$. Коэффициенты α_i, $i = 1, 2, \ldots, l$, подберем из условия, что вектор $\boldsymbol{b}_{l+1}$ должен быть ортогональным ко всем векторам $\boldsymbol{b}_i$, $i = 1, 2 \ldots, l$:

$$0 = (\boldsymbol{b}_i, \boldsymbol{b}_{l+1}) = (\boldsymbol{b}_i, \alpha_1 \boldsymbol{b}_1 + \alpha_2 \boldsymbol{b}_2 + \ldots + \alpha_l \boldsymbol{b}_l, + \boldsymbol{a}_{l+1}) =$$

$$= \alpha_1 (\boldsymbol{b}_i, \boldsymbol{b}_1) + \alpha_2 (\boldsymbol{b}_i, \boldsymbol{b}_2) + \ldots + \alpha_l (\boldsymbol{b}_i, \boldsymbol{b}_l) + (\boldsymbol{b}_i, \boldsymbol{a}_{l+1});$$

отсюда, так как векторы $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_l$ ортогональны между собой,

$$\alpha_i (\boldsymbol{b}_i, \boldsymbol{b}_i) + (\boldsymbol{b}_i, \boldsymbol{a}_{l+1}) = 0,$$

т.е.

$$\alpha_i = -(\boldsymbol{b}_i, \boldsymbol{a}_{l+1})/(\boldsymbol{b}_l, \boldsymbol{b}_i), \quad i = 1, 2, \ldots, l$$

Продолжая этот процесс, мы построим искомую ортогональную систему $\boldsymbol{b}_1\ \boldsymbol{b}_2, \ldots, \boldsymbol{b}_k$.

Применяя процесс ортогонализации к произвольному базису пространства E_n, мы получим ортогональную систему из n ненулевых векторов, т.е., так как эта система по доказанному линейно независима, *ортогональный базис.* При этом, используя замечание, сделанное в связи с первым шагом процесса ортогонализации, а также учитывая то, что всякий ненулевой вектор можно включить в некоторый базис пространства, можно сформулировать следующее утверждение:

Всякое евклидово пространство обладает ортогональными базисами, причем любой ненулевой вектор этого пространства входит в состав некоторого ортогонального базиса.

В дальнейшем важную роль будет играть один специальный вид ортогональных базисов; базисы этого вида отвечают прямоугольным декартовим системам координат, которые используются в аналитической геометрии.

Назовем вектор $\boldsymbol{b}$ *нормированным,* если его скалярный квадрат равен единице,

$$(\boldsymbol{b}, \boldsymbol{b}) = 1.$$

Если $\boldsymbol{a} \neq 0$, откуда $(\boldsymbol{a}, \boldsymbol{a}) > 0$, то *нормированием* вектора $\boldsymbol{a}$ называется переход к вектору

$$\boldsymbol{b} = \frac{1}{\sqrt{(a, a)}} a \ .$$

Вектор $\boldsymbol{b}$ будет нормированным, так как

$$(\boldsymbol{b},\boldsymbol{b})=\left(\frac{1}{\sqrt{(a,a)}}a,\ \frac{1}{\sqrt{(a,a)}}a\right)=\left(\frac{1}{\sqrt{(a,a)}}a\right)^2(\boldsymbol{a},\boldsymbol{a})=1.$$

Базис $\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n$ евклидова пространства E_n называется *ортонормированным,* если он ортогонален, а все его векторы нормированы, т.е.

$$(\boldsymbol{e}_i, \boldsymbol{e}_j) = 0 \text{ при } i\neq j,$$
$$(\boldsymbol{e}_j, \boldsymbol{e}_i)=1, \quad i=1, 2, ... n. \tag{6}$$

Всякое евклидово пространство обладает ортонормированными базисами.

Для доказательства достаточно взять любой ортогональный базис и нормировать все его векторы. Базис останется при этом ортогональным, так как при любых α и β из $(\boldsymbol{a}, \boldsymbol{b}) = 0$ следует

$$(\alpha\boldsymbol{a}, \beta\boldsymbol{b}) = \alpha\beta(\boldsymbol{a}, \boldsymbol{b}) = 0.$$

Базис $\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n$ *евклидова пространства* E_n *тогда и только тогда будет ортонормированным, если скалярное произведение любых двух векторов пространства равно сумме произведений соответствующих координат этих векторов в указанном базисе, т.е. с*

$$\boldsymbol{a}=\sum_{i=1}^{n}\alpha_i\boldsymbol{e}_i\ , \qquad \boldsymbol{b}=\sum_{j=1}^{n}\beta_j e_j\ , \tag{7}$$

следует

$$(\boldsymbol{a}, \boldsymbol{b}) = \sum_{i=1}^{n}\alpha_i\beta_i\ . \tag{8}$$

Действительно, если для нашего базиса выполняются равенства (6), то

$$(\boldsymbol{a}, \boldsymbol{b}) = \left(\sum_{i=1}^{n}\alpha_i e_i\ , \sum_{j=1}^{n}\beta_j e_j\right) = \sum_{i,j=1}^{n}\alpha_i\beta_j(e_i, e_j) = \sum_{i=1}^{n}\alpha_i\beta_i$$

Обратно, если наш базис таков, что для любых векторов $\boldsymbol{a}$ и $\boldsymbol{b}$, записанных в этом базисе в виде (7), справедливо равенство (8), то, беря в качестве $\boldsymbol{a}$ и $\boldsymbol{b}$ любые два вектора этого базиса $\boldsymbol{e}_i$ и $\boldsymbol{e}_j$, различные или одинаковые, мы из (8) выведем равенства (6).

Сопоставляя полученный сейчас результат с изложенным ранее доказательством существования n-мерных евклидовых пространств для любого n, можно высказать следующее утверждение: *если в n-мерном линейном пространстве M_n выбран произвольный базис, то в M_n можно так задать скалярное умножение, что в полученном* *евклидовом пространстве выбранный базис будет одним из ортонормированных базисов.*

Изоморфизм евклидовых пространств. Евклидовы пространства E и E' называются *изоморфными,* если между векторами этих пространств можно установить такое взаимно-однозначное соответствие, что выполняются следующие требования:

1) это соответствие является изоморфным соответствием между E и E', рассматриваемыми как линейные пространства;

2) при этом соответствии сохраняется скалярное произведение; другими словами, если образами векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ из E служат соответственно векторы $\boldsymbol{a}'$ и $\boldsymbol{b}'$ мз E', то

$$(\boldsymbol{a}, \boldsymbol{b})=(\boldsymbol{a}', \boldsymbol{b}'). \qquad (9)$$

Из условия 1) следует, что *изоморфные евклидовы пространства имеют одну и ту же размерность.*

Докажем обратное утверждение:

Любые евклидовы пространства E и E', которые имеют одну и ту же размерность n, изоморфны между собой.

В самом деле, выберем в пространствах E и E' ортонормированные базисы

$$\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n \qquad (10)$$

и, соответственно,

$$\boldsymbol{e}'_1, \boldsymbol{e}'_2, \ldots, \boldsymbol{e}'_n \qquad (11)$$

Ставя в соответствие всякому вектору

$$\boldsymbol{a}=\sum_{i=1}^{n}\alpha_i \boldsymbol{e}_i ,$$

из E вектор

$$\boldsymbol{a}'=\sum_{i=1}^{n}\alpha_i e'_i ,$$

из E', который имеет в базисе (11) те же координаты, что и вектор $\boldsymbol{a}$ в базисе (10), мы получим изоморфное соответствие между линейными пространствами E и E'. Покажем, что выполняется и равенство (9): если

$$\boldsymbol{b}=\sum_{i=1}^{n}\beta_i e_i , \quad \boldsymbol{b}'=\sum_{i=1}^{n}\beta_i e'_i ,$$

то, в силу (8) - учесть ортонормированность базисов (8) и (9) -

$$(\boldsymbol{a}, \boldsymbol{b})= \sum_{i=1}^{n}\alpha_i\beta_i = (\boldsymbol{a}', \boldsymbol{b}').$$

Естественно изоморфные евклидовы пространства не считать разными. Поэтому для всякого n существует единственное n-мерное евклидово пространство в том же смысле, в каком для всякого n существует единственное n-мерное действительное линейное пространство.

На случай комплексных линейных пространств понятия и результаты этого раздела переносятся следующим образом. Комплексное линейное пространство называется *унитарным пространством,* если в нем задано скалярное умножение, причем $(\boldsymbol{a}, \boldsymbol{b})$ будет, вообще говоря, комплексным числом; при этом должны выполняться аксиомы II — IV (в формулировке последней аксиомы следует подчеркнуть, что скалярный квадрат ненулевого вектора действителен и строго положителен), а аксиома I заменяется аксиомой

I' $$(a, b) = \overline{(b, a)}\ ,$$

где черточка обозначает, как обычно, переход к сопряженному комплексному числу.

Скалярное умножение уже не будет, следовательно, коммутативным. Тем не менее, равенство, симметричное аксиоме II, остается справедливым,

II' $$(a, b + c) = (a, b) + (a, c),$$

так как

$$(a, b+c)=\overline{(b+c, a)} = \overline{(b,a)+(c,a)} = \overline{(b,a)} + \overline{(c,a)} = (a,b)+(a,c).$$

С другой стороны,

III' $$(a, \alpha b)=\bar{\alpha}\,(a, b),$$

так как

$$(a, \alpha b) = \overline{(\alpha b, a)} = \overline{\alpha(b,a)} = \bar{\alpha}\ \overline{(b,a)} = \bar{\alpha}\ (a, b).$$

Понятие ортогональности и ортонормированной системы векторов переносятся на случай унитарных пространств без всяких изменений.

Как и выше, доказывается существование ортонормированных базисов во всяком конечномерном унитарном пространстве. При этом, однако, если $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ — ортонормированный базис и векторы $\boldsymbol{a}, \boldsymbol{b}$ имеют в этом базисе запись (4.63), то

$$(\boldsymbol{a}, \boldsymbol{b})= \sum_{i=1}^{n} \alpha_i \overline{\beta_i}\ .$$

6.2. Ортогональные матрицы, ортогональные преобразования

Пусть дано действительное линейное преобразование n неизвестных:

$$x_i=\sum_{k=1}^{n} q_{ik}y_k\ ,\ u=1.2.....n; \qquad (12)$$

матрицу этого преобразования обозначим через Q. Это преобразование переводит сумму квадратов неизвестных $x_1, x_2, ..., x_n$, т.е. квадратичную форму $x^2{}_1+\ x^2{}_2+\ ...+\ x^2{}_n$, являющуюся нормальным видом положительно определенных квадратичных форм, в некоторую квадратичную форму от неизвестных $y_1, y_2, ..., y_n$. Случайно эта новая квадратичная форма сама может оказаться суммой квадратов неизвестных $y_1, y_2, ..\ .,y_n$, т.е. может иметь место равенство

$$x^2{}_1+ x^2{}_2+ ...+ x^2{}_n =y^2{}_1+y^2{}_2+...+y^2{}_n \qquad (13)$$

тождественное после замены неизвестных $x_1, x_2, ..., x_n$ их выражениями (12). Линейное преобразование неизвестных (12), обладающее этим свойством, т.е., как говорят, оставляющее сумму квадратов неизвестных инвариантной, называется *ортогональным преобразованием неизвестных,* а его матрица Q — *ортогональной матрицей.*

Существует много других определений ортогонального преобразования и ортогональной матрицы, эквивалентных приведенным выше. Укажем некоторые из них, которые необходимы для дальнейшего.

Мы знаем закон, по которому преобразуется матрица квадратичной формы при выполнении линейного преобразования неизвестных. Применяя его к нашему случаю и учитывая, что матрицей квадратичной формы, являющейся суммой квадратов всех неизвестных, служит единичная матрица E, мы получим, что равенство (13) равносильно матричному равенству

$$Q'E\ Q =E,$$

т.е.

$$Q' = E. \qquad (14)$$

Отсюда

$$Q' = Q^{-1} \qquad (15)$$

а потому справедливо и равенство

$$QQ' = E. \qquad (16)$$

Таким образом, ввиду (13), *ортогональную матрицу Q можно определить как такую матрицу, для которой транспонированная матрица Q' равна обратной матрице Q^{-1}.*

Каждое из равенств (14) и (16) также может быть принято в качестве определения ортогональной матрицы.

Так как столбцы матрицы Q' являются строками матрицы Q, то из (16) следует следующее утверждение: *квадратная матрица Q тогда и только тогда будет ортогональной, если сумма квадратов всех элементов любой ее строки равна единице, а сумма произведений соответствующих элементов любых двух ее различных строк равна нулю.* Из (14) следует аналогичное утверждение для столбцов матрицы Q.

Переходя в равенстве (14) к определителям, мы получим, ввиду того, что $|Q'| = |Q|$, равенство

$$|Q|^2 = 1.$$

Отсюда следует, что *определитель ортогональной матрицы равен* ± 1. Таким образом, *всякое ортогональное преобразование неизвестных является невырожденным*. При этом утверждать обратное нельзя; отметим также, что далеко не всякая матрица с определителем, равным ± 1, будет ортогональной.

Матрица, обратная к ортогональной, сама будет ортогональной. Действительно, переходя в (15) к транспонированным матрицам, мы получим:

$$(Q^{-1})' = (Q')' = Q=(Q^{-1})^{-1}.$$

С другой стороны, *произведение ортогональных матриц само ортогонально*. Действительно, если матрицы Q и R ортогональные, то, используя (15) а также равенство $X'=Y'Q'$ и аналогичное равенство, которое справедливо для обратной матрицы, мы получим:

$$(Q/R)'= R''= R^{-1}Q^{-1} =(Q/R)^{-1}.$$

В позже будет использовано следующее утверждение:

Матрица перехода от ортонормированного базиса евклидова пространства к любому другому его ортонормированному базису является ортогональной.

Пусть, в самом деле, в пространстве E_n задано два ортонормированных базиса $\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n$ и $\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n$ с матрицей перехода $Q=(q_{ij})$,

$$\boldsymbol{e}' = Q\boldsymbol{e}.$$

Так как базис $\boldsymbol{e}$ ортонормированный, то скалярное произведение любых двух векторов, в частности любых двух векторов из базиса $\boldsymbol{e}'$, равно сумме произведений соответствующих координат этих векторов в базисе $\boldsymbol{e}$. Так как, однако, и базис $\boldsymbol{e}'$ ортонормированный, то скалярный квадрат каждого вектора из $\boldsymbol{e}'$ равен единице, а скалярное произведение любых двух различных векторов из $\boldsymbol{e}'$ равно нулю. Отсюда для строк координат векторов базиса $\boldsymbol{e}'$ в базисе $\boldsymbol{e}$, т.е. для строк матрицы Q, вытекают те утверждения, которые, как выведено выше из равенства (16), характерны для ортогональной матрицы.

Ортогональные преобразования евклидова пространства. Сейчас уместно изучить один специальный тип линейных преобразований евклидовых пространств.

Линейное преобразование φ евклидова пространства E_n называется *ортогональным преобразованием этого евклидова пространства,* если оно сохраняет скалярный квадрат всякого вектора, т.е. для любого вектора $\boldsymbol{a}$

$$(\boldsymbol{a}\varphi, \boldsymbol{a}\varphi) = (\boldsymbol{a}, \boldsymbol{a}). \qquad (17)$$

Отсюда выводится следующее более общее утверждение, которое, разумеется, также может быть принято в качестве определения ортогонального преобразования:

Ортогональное преобразование φ евклидова пространства сохраняет скалярное произведение любых двух векторов $\boldsymbol{a}, \boldsymbol{b}$,

$$(\boldsymbol{a}\varphi, \boldsymbol{b}\varphi) = (\boldsymbol{a}, \boldsymbol{b}). \qquad (18)$$

Действительно, ввиду (18)

$$((\boldsymbol{a}+\boldsymbol{b})\varphi, (\boldsymbol{a}+\boldsymbol{b})\varphi) = (\boldsymbol{a}+\boldsymbol{b}, \boldsymbol{a}+\boldsymbol{b}).$$

Однако

$$((\boldsymbol{a}+\boldsymbol{b})\varphi, (\boldsymbol{a}+\boldsymbol{b})\varphi) = (\boldsymbol{a}\varphi+\boldsymbol{b}\varphi, \boldsymbol{a}\varphi+\boldsymbol{b}\varphi) =$$
$$= (\boldsymbol{a}\varphi, \boldsymbol{a}\varphi) + (\boldsymbol{a}\varphi, \boldsymbol{b}\varphi) + (\boldsymbol{b}\varphi, \boldsymbol{a}\varphi) + (\boldsymbol{b}\varphi, \boldsymbol{b}\varphi), (\boldsymbol{a}+\boldsymbol{b}, \boldsymbol{a}+\boldsymbol{b}) =$$
$$=(\boldsymbol{a}, \boldsymbol{a})+(\boldsymbol{a}, \boldsymbol{b}) + (\boldsymbol{b}, \boldsymbol{a}) + (\boldsymbol{b}, \boldsymbol{b}).$$

Отсюда, используя (17) как для $\boldsymbol{a}$, так и для $\boldsymbol{b}$, и учитывая коммутативность скалярного умножения, получаем

$$2(\boldsymbol{a}\varphi, \boldsymbol{b}\varphi) = 2(\boldsymbol{a}, \boldsymbol{b}),$$

а поэтому имеет место и (18).

При ортогональном преобразовании евклидова пространства образы всех векторов любого ортонормированного базиса сами составляют ортонормированный базис. Обратно, если линейное преобразование евклидова пространства переводит хотя бы один ортонормированный базис снова в ортонормированный базис, то это преобразование ортогонально.

В самом деле, пусть φ — ортогональное преобразование пространства E_n, а $\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n$ — произвольный ортонормированный базис этого пространства. Ввиду (17) из равенств

$$(\boldsymbol{e}_i, \boldsymbol{e}_i)=1, \quad i=1, 2, ..., n,$$
$$(\boldsymbol{e}_i, \boldsymbol{e}_j) = 0 \quad \text{при } i \neq j$$

вытекают равенства

$$(\boldsymbol{e}_i\varphi, \boldsymbol{e}_i\varphi)=1, \ i=1, 2, ..., n,$$
$$(\boldsymbol{e}_i\varphi, \boldsymbol{e}_j\varphi) = 0 \quad \text{при } i \neq j,$$

т.е. система векторов $\boldsymbol{e}_1\varphi, \boldsymbol{e}_2\varphi, ..., \boldsymbol{e}_n\varphi$ оказывается ортогональной и нормированной, а поэтому она будет ортонормированным базисом пространства E_n.

Обратно, пусть линейное преобразование φ пространства E_n переводит ортонормированный базис $\boldsymbol{e_1}, \boldsymbol{e_2}, ..., \boldsymbol{e}_n$ снова в ортонормированный базис, т.е. система векторов $\boldsymbol{e}_1\varphi, \boldsymbol{e}_2\varphi, ..., \boldsymbol{e}_n\varphi$ является ортонормированным базисом пространства E_n. Если

$$\boldsymbol{a} = \sum_{i=1}^{n} \alpha_i \boldsymbol{e}_i \ ,$$

— произвольный вектор пространства E_n, то

$$\boldsymbol{a}\varphi = \sum_{i=1}^{n} \alpha_i (e_i \varphi) \ ,$$

т.е. вектор $\boldsymbol{a}\varphi$ имеет в базисе $\boldsymbol{e}\varphi$ те же координаты, что и вектор $\boldsymbol{a}$ в базисе $\boldsymbol{e}$. Эти оба базиса являются, однако, ортонормированными, а поэтому скалярный квадрат любого вектора равен сумме квадратов его координат в каждом из этих базисов. Таким образом,

$$(\boldsymbol{a}, \boldsymbol{a}) = (\boldsymbol{a}\varphi, \boldsymbol{a}\varphi) = \sum_{i=1}^{n} \alpha_i^{\ 2} \ ,$$

т.е. равенство (17) действительно выполняется.

Ортогональное преобразование евклидова пространства в любом ортонормированном базисе задается ортогональной матрицей. Обратно, если линейное преобразование евклидова пространства хотя бы в одном ортонормированном базисе задается ортогональной матрицей, то это преобразование ортогонально.

Действительно, если преобразование φ ортогонально, а базис $\boldsymbol{e_1}, \boldsymbol{e_2}, ..., \boldsymbol{e}_n$ ортонормированный, то и система векторов $\boldsymbol{e}_1\varphi, \boldsymbol{e}_2\varphi, ..., \boldsymbol{e}_n\varphi$ будет ортонормированным базисом. Матрица A преобразования φ в базисе $\boldsymbol{e}$,

$$\boldsymbol{e}\varphi = A\boldsymbol{e}, \tag{19}$$

будет, следовательно, матрицей перехода от ортонормированного базиса $\boldsymbol{e}$ к ортонормированному базису $\boldsymbol{e}\varphi$, т.е., как доказано выше, будет ортогональной.

Обратно, пусть линейное преобразование φ задается в ортонормированном базисе $\boldsymbol{e_1}, \boldsymbol{e_2}, ..., \boldsymbol{e}_n$ ортогональной матрицей A; имеет место, следовательно, равенство (19). Так как базис $\boldsymbol{e}$ ортонормированный, то скалярное произведение любых векторов, в частности любых векторов из системы $\boldsymbol{e}_1\varphi, \boldsymbol{e}_2\varphi, ..., \boldsymbol{e}_n\varphi$, равно сумме произведений соответствующих координат этих векторов в базисе $\boldsymbol{e}$. Поэтому, так как матрица A ортогональна,

$(\boldsymbol{e}_i\varphi, \boldsymbol{e}_i\varphi)=1$, $i=1, 2, \ldots, n$,

$(\boldsymbol{e}_i\varphi, \boldsymbol{e}_j\varphi) = 0$ при $i\neq j$,

т.е. система $\boldsymbol{e}\varphi$ самая оказывается ортонормированным базисом пространства E_n. Отсюда вытекает ортогональность преобразования φ. Как известно из курса аналитической геометрии, среди всех аффинных преобразований плоскости, оставляющих на месте начало координат, вращения (соединенные, может быть, с зеркальными отражениями) являются единственными, сохраняющими скалярное произведение векторов. Таким образом, ортогональные преобразования n-мерного евклидова пространства можно рассматривать как «вращение» этого пространства. К числу ортогональных преобразований евклидова пространства принадлежит тождественное преобразование. С другой стороны, установленная нами связь между ортогональными преобразованиями и ортогональными матрицами, а также связь, которая имеет место между операциями над линейными преобразованиями и над матрицами позволяют из известных свойств ортогональных матриц вывести следующие свойства ортогональных преобразований евклидова пространства:

Всякое ортогональное преобразование является невырожденным и его обратное преобразование также ортогонально.

Произведение любых ортогональных преобразований ортогонально.

6.3. Симметрические преобразования

Линейное преобразование φ n-мерного евклидова пространства называется *симметрическим* (или *самоспряженным),* если для любых векторов $\boldsymbol{a}, \boldsymbol{b}$ этого пространства имеет место равенство

$$(\varphi\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{a}, \boldsymbol{b}\varphi), \tag{20}$$

т.е. символ симметрического преобразования можно при скалярном умножении переносить с одного множителя на другой.

Примерами симметрических преобразований служат тождественное преобразование $\boldsymbol{e}$ и нулевое преобразование ω. Более общим примером является линейное преобразование, при котором всякий вектор увеличивается на фиксированное число α,

$$\boldsymbol{a}\varphi = \alpha\boldsymbol{a}.$$

Действительно, в этом случае

$$(\boldsymbol{a}\varphi, \boldsymbol{b}) = (\alpha\boldsymbol{a}, \boldsymbol{b}) = \alpha(\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{a}, \alpha\boldsymbol{b}) = (\boldsymbol{a}, \boldsymbol{b}\varphi).$$

Роль симметрических преобразований весьма велика и нам необходимо изучить их достаточно детально.

Симметрическре преобразование евклидива пространства в любому ортонормированном базисе задается симметрической матрицей. Обратно, если линейное преобразование евклидова *пространства хотя бы в одном ортонормированном базисе задается симметрической матрицей, то это преобразование симметрическое.*

Действительно, пусть симметрическое преобразование φ задается в ортонормированном базисе $\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n$ матрицей $A=(\alpha_{ij})$. Учитывая то, что в ортонормированном базисе скалярное произведение двух векторов равно сумме произведений соответствующих координат этих векторов, мы получаем:

$$(\boldsymbol{e}_i\varphi, \boldsymbol{e}_j) = \left(\sum_{\boldsymbol{k}=1}^{\boldsymbol{n}} \alpha_{\boldsymbol{ik}}\boldsymbol{e}_{\boldsymbol{k}}, \boldsymbol{e}_{\boldsymbol{j}}\right) = \alpha_{\boldsymbol{ij}},$$

$$(\boldsymbol{e}_i, \boldsymbol{e}_j\varphi) = \boldsymbol{e}_{\boldsymbol{i}}\left(\sum_{\boldsymbol{k}=1}^{\boldsymbol{n}} \alpha_{\boldsymbol{jk}}\boldsymbol{e}_{\boldsymbol{k}}\right) = \alpha_{\boldsymbol{ij}},$$

т.е., ввиду (20),

$$\alpha_{ij} = \alpha_{ji}$$

для всех i и j. Матрица A оказалась, таким образом, симметрической.

Обратно, пусть линейное преобразование φ задается в ортонормированном базисе $\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n$ симметрической матрицей $A=(\alpha_{ij})$,

$$\alpha_{ij} = \alpha_{ji} \text{ для всех } i \text{ и } j. \tag{21}$$

Если

$$\boldsymbol{b} = \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \beta_{\boldsymbol{i}}\boldsymbol{e}_{\boldsymbol{i}}, \qquad \boldsymbol{c} = \sum_{\boldsymbol{j}=1}^{\boldsymbol{n}} \gamma_{\boldsymbol{j}}\boldsymbol{e}_{\boldsymbol{j}},$$

— любые векторы пространства, то

$$\boldsymbol{b}\varphi = \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \beta_{\boldsymbol{i}}(\boldsymbol{e}_{\boldsymbol{i}}\varphi) = \sum_{\boldsymbol{j}=1}^{\boldsymbol{g}}\left(\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \beta_{\boldsymbol{i}}\alpha_{\boldsymbol{ij}}\right)\boldsymbol{e}_{\boldsymbol{j}},$$

$$\boldsymbol{c}\varphi = \sum_{\boldsymbol{j}=1}^{\boldsymbol{n}} \gamma_{\boldsymbol{j}}(\boldsymbol{e}_{\boldsymbol{j}}\varphi) = \sum_{\boldsymbol{i}=1}^{\boldsymbol{g}}\left(\sum_{\boldsymbol{j}=1}^{\boldsymbol{n}} \gamma_{\boldsymbol{j}}\alpha_{\boldsymbol{ji}}\right)\boldsymbol{e}_{\boldsymbol{i}}.$$

Используя ортонормированность базиса $\boldsymbol{e}$, получаем

$$(\boldsymbol{b}\varphi, \boldsymbol{c}) = \sum_{j,i=1}^{n} \beta_i\alpha_{ij}\gamma_j,$$

$$(\boldsymbol{b}, \boldsymbol{c}\varphi) = \sum_{i,j=1}^{n} \beta_i\gamma_j\alpha_{ji}.$$

Ввиду (21) правые части последних равенств совпадают, а поэтому

$$(\boldsymbol{b}\varphi,\boldsymbol{c}) = (\boldsymbol{b},\boldsymbol{c}\varphi)$$

что и требовалось доказать.

Из полученного результата вытекает следующее свойство симметрических преобразований:

Сумма симметрических преобразований, а также произведение симметрического преобразования на число, являются симметрическими преобразованиями.

Докажем следующую теорему:

Все характеристические корни симметрического преобразования действительны.

Так как характеристические корни любого линейного преобразования совпадают с характеристическими корнями матрицы этого преобразования в любом базисе, а симметрическое преобразование задается в ортонормированных базисах симметрическими матрицами, то достаточно доказать следующее утверждение:

Все характеристические корни симметрической матрицы действительны.

В самом деле, пусть λ_0 будет характеристический корень (быть может, комплексный) симметрической матрицы $A = (\alpha_{ij})$,

$$|A - \lambda_0 E| = 0.$$

Тогда система линейных однородных уравнений с комплексными коэффициентами

$$\sum_{j=1}^{n} \alpha_{ij} x_j = \lambda_0 x_i, \qquad i=1, 2, ... n,$$

имеет равный нулю определитель, т.е. обладает ненулевым решением $\beta_1, \beta_2, ..., \beta_n$, вообще говоря, комплексным; таким образом,

$$\sum_{j=1}^{n} \alpha_{ij} \beta_j = \lambda_0 \beta_i, \quad i=1, 2, ... n. \tag{22}$$

Умножая обе части каждой i-го из равенств (22) на число $\overline{\beta}_i$, сопряженное с числом β_i, и складывая отдельно левые и правые части всех получающихся равенств, мы приходим к равенству

$$\sum_{i,j=1}^{n} \alpha_{ij} \beta_j \overline{\beta}_i = \lambda_0 \sum_{i=1}^{n} \beta_i \overline{\beta}_i. \tag{23}$$

Коэффициент при λ_0 в (23) является отличным от нуля действительным числом, будучи суммой неотрицательных действительных чисел, хотя бы одно из которых строго положительно. Действительность числа λ_0 будет поэтому доказана, если мы докажем действительность левой части равенства (23), для чего достаточно показать, что это комплексное число совпадает со своим сопряженным. Здесь будет использована симметричность (действительной) матрицы A.

$$\overline{\sum_{i,j=1}^{n} \alpha_{i,j}\beta_j\overline{\beta}_i} = \sum_{i,j=1}^{n} \overline{\alpha_{ij}\beta_j\overline{\beta}_i} = \sum_{i,j=1}^{n} \alpha_{ij}\overline{\beta}_j\beta_i = \sum_{i,j=1}^{n} \alpha_{ji}\overline{\beta}_j\beta_i =$$

$$= \sum_{i,j=1}^{n} \alpha_{ij}\overline{\beta}_i\beta_j = \sum_{i,j=1}^{n} \alpha_{ij}\beta_j\overline{\beta}_i.$$

Заметим, что предпоследнее равенство получено простой переменной обозначений для индексов суммирования: вместо i поставлено j, вместо j поставлено i. Теорема, следовательно, доказана.

Линейное преобразование φ евклидова пространства E_n тогда и только тогда будет симметрическим, если в пространстве E_n существует ортонормированный базис, составленный из собственных векторов этого преобразования.

В одну сторону это утверждение почти очевидно: если в E_n существует ортонормированный базис $\boldsymbol{e_1}, \boldsymbol{e_2}, ..., \boldsymbol{e_n}$, причем

$$\boldsymbol{e}_i\varphi = \lambda_i\boldsymbol{e}_i, \quad i = 1, 2, ..., n,$$

то в базисе $\boldsymbol{e}$ преобразование φ задается диагональной матрицей

$$\begin{pmatrix} \lambda_1 & & & & 0 \\ & \lambda_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & \lambda_n \end{pmatrix}.$$

Диагональная матрица является, однако, симметрической, а поэтому преобразование φ задается в ортонормированном базисе $\boldsymbol{e}$ симметрической матрицей, т.е. будет симметрическим.

Основное обратное утверждение теоремы мы будем доказывать индукцией по размерности n пространства E_n. В самом деле, при $n = 1$ всякое линейное преобразование φ пространства E_1 непременно переводит люьой вектор в вектор, ему пропорциональный. Отсюда следует, что всякий ненулевой вектор $\boldsymbol{a}$ будет собственным вектором для φ (как, впрочем, следует и то, что всякое линейное преобразование пространства E_1 будет симметрическим). Нормируя вектор $\boldsymbol{a}$, мы получим искомый ортонормированный базис пространства E_1.

Пусть утверждение теоремы уже доказано для $(n—1)$-мерного евклидова пространства и пусть в пространстве E_n задано симметрическое преобразование φ. Из доказанной выше теоремы вытекает существование для φ действительного характеристического корня λ_0. Это число будет, следовательно, собственным значением для преобразования φ. Если $\boldsymbol{a}$ — собственный вектор преобразования φ, который относится к этому собственному значению, то и всякий ненулевой вектор, пропорциональный вектору $\boldsymbol{a}$, будет для φ собственным вектором, который относится к тому же собственному значению λ_0, так как

$$(\alpha\boldsymbol{a})\varphi=\alpha(\boldsymbol{a}\varphi)=\alpha(\lambda_0\boldsymbol{a})=\lambda_0(\alpha\boldsymbol{a}).$$

В частности, нормируя вектор $\boldsymbol{a}$, мы получим такой вектор $\boldsymbol{e}_1$, что

$$\boldsymbol{e}_1\varphi=\lambda_0\boldsymbol{e}_1,$$
$$(\boldsymbol{e}_1, \boldsymbol{e}_1)=1.$$

Доказано, ненулевой вектор $\boldsymbol{e}_1$ можно включить в ортогональный базис

$$\boldsymbol{e}_1, \boldsymbol{e}'_2, \ldots, \boldsymbol{e}'_n, \qquad (24)$$

пространства E_n. Те векторы, первая координата которых в базисе (24) равна нулю, т.е. векторы вида $\alpha\boldsymbol{e}'_2+\ldots+\alpha\boldsymbol{e}'_n$, составляют, очевидно, $(n—1)$-мерное линейное подпространство пространства E_n, которое мы обозначим через L. Это будет даже $(n—1)$-мерное евклидово пространство, так как скалярное произведение, будучи определенным для всех векторов из E_n, определено, в частности, для векторов из L, причем обладает всеми необходимыми свойствами. Подпространство L состоит из всех тех векторов пространства E_n, которые ортогональны к вектору $\boldsymbol{e}_1$. Действительно, если

$$\boldsymbol{a}=\alpha_1\boldsymbol{e}_1+\alpha'\boldsymbol{e}'_2+\ldots+\alpha'\boldsymbol{e}'_n,$$

то, ввиду ортогональнисть базиса (24) и нормированности вектора $\boldsymbol{e}_1$

$$(\boldsymbol{e}_1, \boldsymbol{a})=\alpha_1(\boldsymbol{e}_1, \boldsymbol{e}_1)+\alpha'_2(\boldsymbol{e}_1, \boldsymbol{e}'_2)+\ldots+\alpha'_n(\boldsymbol{e}_1, \boldsymbol{e}'_n)=\alpha_1,$$

т.е. $(\boldsymbol{e}_1, \boldsymbol{a})=0$ тогда и только тогда, если $\alpha_1=0$.

Если вектор $\boldsymbol{a}$ принадлежит к подпространству L, т.е. $(\boldsymbol{e}_1, \boldsymbol{a})=0$, то и вектор $\boldsymbol{a}\varphi$ содержится в L. Действительно, ввиду симметричности преобразования φ,

$$(\boldsymbol{e}_1, \boldsymbol{a}\varphi)=(\boldsymbol{e}_1\varphi, \boldsymbol{a}_1)=(\lambda_0\boldsymbol{e}_1, \boldsymbol{a})=\lambda_0(\boldsymbol{e}_1, \boldsymbol{a})=\lambda_0\cdot 0=0,$$

т.е. вектор $\boldsymbol{a}\varphi$ ортогонален к $\boldsymbol{e}_1$ и поэтому содержится в L. Это свойство подпространства L, называемое его *инвариантностью* относительно преобразования φ, позволяет считать φ, рассматриваемое лишь в применении к векторам из L, линейным преобразованием этого $(n—1)$-мерного евклидова пространства. Оно будет даже симметрическим преобразованием пространства L, так как равенство (20), выполняясь для любых векторов из E_n, будет выполняться, в частности, для векторов, которые лежат в L.

В силу индуктивного предположения в пространстве L существует ортонормированный базис, который состоит из собственных векторов преобразования φ; обозначим его через $e_2, ..., e_n$. Все эти векторы ортогональны к вектору e_1, а поэтому $e_1, e_2, ..., e_n$ будет искомым ортонормированным базисом пространства E_n, состоящей из собственных векторов преобразования φ. Теорема доказана.

6.4. Приведение квадратичной формы к главным осям. Пары форм

Квадратичные формы. Пусть $\mathbf{x}=(x_1, x_2,..., x_n)$ - вектор в n-мерном линейном просторанстве с действительными компонентами. Выражение вида

$$Q(\mathbf{x})=\sum_i \sum_j a_{ij} x_i x_j$$

называется *квадратичной формой.* В этом выражении без ограничения общности можем считать $a_{ij}=a_{ji}$.

Если это не так, то, выполняя преобразование

$$a_{ij}\, x_i\, x_j+ a_{ji}\, x_j\, x_i=(a_{ij}+a_{ji})\, x_i\, x_j=1/2(a_{ij}+a_{ji})\, x_i\, x_j+1/2(a_{ij}+a_{ji})\, x_j\, x_i,$$

можно привести квадратичную форму к виду, когда это условие выполняется. Поэтому числа a_{ij} можно рассматривать как элементы квадратной матрицы A, которая является *вещественной и симметрической.*

Квадратическую форму удобно записывать в векторно-матричных обозначениях. Для этого перепишем выражение для $Q(x)$ в виде

$$Q(\mathbf{x})=\sum_i x_i \sum_j a_{ij} x_j = \sum_i x_i z_i = \mathbf{x'z},$$

где величины

$$z_i=\sum_j a_{ij} x_j\ , \qquad i=1, 2, \ldots, n$$

есть компонентами вектора $\mathbf{z}$, связанного с вектором $\boldsymbol{x}$ линейным преобразованием:

$$\mathbf{z}=A\mathbf{x}.$$

Таким образом, квадратичная форма $Q(\mathbf{x})$ может быть записана в виде

$$Q(\mathbf{x})= \mathbf{x}'A\mathbf{x}.$$

Квадратичную форму $Q(\mathbf{x})$ и матрицу A называют *положительно определенными,* если выполняется условие $Q(\mathbf{x})>0$. При выполнении условия $Q(\mathbf{x})<0$ говорят об *отрицательно определенной* квадратичной форме $Q(\mathbf{x})$ и матрице A. Для нахождения условий положительной (отрицательной) определенности матрицы A введем в рассмотрение линейное преобразование $\mathbf{x}=V\mathbf{y}$, где V — невырожденная матрица. Тогда

$$Q(\mathbf{x})= \mathbf{x}'A\mathbf{x}=\mathbf{y}'V'AV\mathbf{y}= Q(\mathbf{y}),$$

причем матрицей квадратичной формы $Q(\mathbf{y})$ является $V'AV$. Возьмем в качестве V диагонализирующую матрицу. Тогда $V'AV= \Lambda$, так что

$$Q(\mathbf{x})=Q(\mathbf{y})= \mathbf{y}'\Lambda\, \mathbf{y}=\sum\lambda_u [y_i]^2.$$

Как видим, необходимым и достаточным условием положительной (отрицательной) определенности матрицы A является положительность (отрицательноть) всех ее собственных значений λ_i.

Пример использования квадратичных форм при отыскании экстремумов функций многих переменных

Применение квадратичных форм проиллюстрируем на примере задачи отыскания экстремумов функций, имещей важное значение при решении многих прикладных задач.

Пусть $\mathbf{u}=(u_1,..., u_n)$-многомерная переменная и $f(\mathbf{u})= f(u_1,..., u_n)$ — функция от $\mathbf{u}$, определенная в области $a_i \le u_i \le b_i$, $i=1, 2, \dots, n$ и допускающая разложение в сходящийся ряд Тейлора в каждой точке $\mathbf{c}=(c_1,...., c_n)$ внутри этой области. Это разложение имеет вид:

$$f(\mathbf{u}) = f(\mathbf{c}) + \sum_i \frac{\partial f}{\partial c_i}\Delta u_i + \frac{1}{2}\sum_i\sum_j \frac{\partial^2 f}{\partial c_i\,\partial c_j}\Delta u_i\,\Delta u_j + \dots;$$

где

$$\Delta u_i = u_i - c_i, \quad i=\overline{1, n};$$

$$\frac{\partial f}{\partial c_i} = \left.\frac{\partial f}{\partial u_i}\right|_{u_i=c_i}, \; i=\overline{1, n};$$

$$\frac{\partial^2 f}{\partial c_i\,\partial c_j} = \left.\frac{\partial^2 f}{\partial u_i\,\partial u_j}\right|_{u_i=c_i,\; u_j=c_j}, \quad i,\, j=\overline{1, n}.$$

Как известно, экстремумы подобных функций могут достигаться внутри области определения только в стационарных точках, т.е. в точках, в которых производные df/du_i, $i=1, 2, \dots, n$ обращаются в нуль. Пусть $\mathbf{u}=\mathbf{c}$ — стационарная точка. Тогда $df/dc_i=0$, $i=1, 2, \dots, n$ и разложение функции $f(\mathbf{u})$ в ряд Тейлора в окрестности точки $\mathbf{u}=\mathbf{c}$ принимает вид:

$$f(\mathbf{u}) = f(\mathbf{c}) + \frac{1}{2}\sum_i\sum_j \frac{\partial^2 f}{\partial c_i\,\partial c_j}\Delta u_i\,\Delta u_j + \dots \tag{25}$$

Введем обозначение:

$$x_i=\Delta u_i= u_i-c_i;$$

$$a_{ij} = a_{ji} = \frac{1}{2}\,\frac{\partial^2 f}{\partial c_i\,\partial c_j}\,.$$

Тогда ряд (25) запишется в виде

$$f(\mathbf{u}) = f(\mathbf{c}) + Q(x) + \ldots \qquad (26)$$

где

$$Q(x) = \sum_i \sum_j a_{ij} x_i x_j$$

представляет собой квадратичную форму $\mathbf{x}'A\mathbf{x}$ с вещественной симметрической матрицей A.

Учитывая, что при $\mathbf{u}=\mathbf{c}$ $x_i=0$, $i=1, 2, \ldots, n$ из (26) видим, что в стационарной точке $\mathbf{u}=\mathbf{c}$ имеет место относительный минимум, если $Q(\mathbf{x})>0$, и относительный максимум, если $Q(\mathbf{x})<0$, т.е. $-Q(\mathbf{x})>0$. В случае, если $Q(\mathbf{x})$ вблизи стационарной точки может принимать как положительные, так и отрицательные значения, имеет место стационарная точка более сложного характера, требующая дальнейшего специального исследования.

Таким образом, мы видим, что выяснение характера стационарной точки функции многих переменных требует выяснения того, есть ли положительно (отрицательно) определенной матрица A квадратичной форма $Q(\mathbf{x})$.

Приведение квадратичной формы к главным осям.

Применим последнюю теорему предыдущего раздела к доказательству следующей матричной теоремы:

Для всякой симметрической матрицы A можно найти такую ортогональную матрицу Q, которая приводит матрицу A к диагональному виду, т.е. матрица $Q^{-1}AQ$, полученная трансформированиям матрицы A матрицей Q, будет диагональной.

В самом деле, пусть дана симметрическая матрица A порядка n. Если $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ - некоторый ортонормированный базис n-мерного евклидова пространства E_n, то матрица A задает в этом базисе симметрическое преобразование φ. Как доказано, в E_n существует ортонормированный базис $\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_n$, составленный из собственных векторов преобразования φ; в этом базисе φ задается диагональной матрицей B. Тогда,

$$B = Q^{-1}AQ, \qquad (27)$$

где Q — матрица перехода от базиса $\boldsymbol{f}$ к базису $\boldsymbol{e}$,

$$\boldsymbol{e} = Q\boldsymbol{f}. \qquad (28)$$

Эта матрица, как матрица перехода от одного ортонормированного базиса к друному такому же базису, будет ортогональной. Теорема доказана.

Так как для ортогональной матрицы Q ее обратная матрица равна транспонированной, $Q^{-1}=Q'$, то равенство (27) можно переписать в виде

$$B=Q'AQ.$$

Известно, что именно так преобразуется симметрическая матрица A квадратичной формы, подвергнутой линейному преобразованию неизвестных с матрицей Q. Учитывая то, что линейное преобразование неизвестных с ортогональной матрицей является ортогональным преобразованием и что диагональную матрицу имеет квадратичная форма, приведенная к каноническому виду, мы на основании предыдущей теоремы получаем следующую теорему о приведении действительной квадратичной формы к главным осям:

Всякая действительная квадратичная форма $f(x_1, x_2, \ldots x_n)$ некоторым ортогональным преобразованием неизвестных может быть приведенная к каноническому виду.

Хотя может существовать много различных ортогональных преобразований неизвестных, которые приводят данную квадратичную форму к каноническому виду, однако сам этот канонический вид определяется однозначно:

Каково бы не было ортогональное преобразование, которое приводит к каноническому виду квадратичную форму $f(x_1, x_2, \ldots x_n)$ с матрицей A, коэффициентами этого канонического вида будут характеристические корни матрицы A, взятые с их кратностями.

Пусть, в самом деле, форма f некоторым ортогональным преобразованием приведена к каноническому виду

$$f(x_1, x_2, \ldots x_n)=\mu_1 y_1^2+\mu_2 y_2^2+ \ \ldots \ +\mu_n y_n^2.$$

Это ортогональное преобразование оставляет инвариантной сумму квадратов неизвестных, а потому, если λ - новое неизвестное, то

$$f(x_1, x_2, \ldots x_n)-\lambda \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} \mu_i y_i^2 - \lambda \sum_{i=1}^{n} y_i^2 .$$

Переходя к определителям этих квадратичных форм и учитывая то, что после выполнения линейного преобразования определитель квадратичной формы увеличивается на квадрат определителя преобразования, а квадрат определителя ортогонального преобразования равен единице, мы приходим к равенству

$$|A-\lambda E| = \begin{vmatrix} \mu_1-\lambda & 0 & \ldots & 0 \\ 0 & \mu_2-\lambda & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & \mu_n-\lambda \end{vmatrix} = \prod_{i=1}^{n} (\mu_i-\lambda),$$

из которого вытекает утверждение теоремы.

Этому результату можно придать также матричную формулировку:

Какова бы не была ортогональная матрица, которая приводит к диагональному виду симметрическую матриuу A, на главной диагонали полученной диагональной матрицы будут стоять характеристические корни матрицы A, взятые с их кратностями.

Практическое разыскание ортогонального преобразования, приводящего квадратичную форму к главным осям. В некоторых задачах необходимо знать не только тот канонический вид, к которому приводится действительная квадратичная форма ортогональным преобразованием, но и самое ортогональное преобразование, которое осуществляет это приведение. Было бы затруднительно разыскивать это преобразование, используя доказательство теоремы о приведении к главным осям, и мы хотим указать другой путь. Нужно лишь научиться находить ортогональную матрицу Q, которая приводит данную симметрическую матрицу A к диагональному виду, или, что то же самое, находить ее обратную матрицу Q^{-1}. Ввиду (28) это будет матрица перехода от базиса ***e*** к базису ***f***, т.е. ее строки являются координатными строками (в базисе ***e***) ортонормированной системы из n собственных векторов симметрического преобразования φ, определяемого матрицей A в базисе ***e***. Остается найти такую систему собственных векторов.

Пусть λ_0 — любой характеристический корень матрицы A и пусть его кратность равна k_0. Мы знаем, что совокупность координатных строк всех собственных векторов преобразования φ, относящихся к собственному значению λ_0, совпадает с совокупностью ненулевых решений системы линейных однородных уравнений

$$(A - \lambda_0 E)X=0; \qquad (29)$$

симметричность матрицы A разрешает написать здесь A вместо A'. Из доказанных выше теорем существования ортогональной матрицы, которая приводит симметрическую матрицу A к диагональной и единственности этого диагонального вида вытекает, что для системы (29) можно найти k_0 линейно независимых решений. Такую систему решений ищем известными методами, а потом ортогонализируем и нормируем полученную систему.

Беря в качестве λ_0 поочередно все различные характеристические корни симметрической матрицы A и учитывая то, что сумма кратностей этих корней равна n, мы получим систему из n собственных векторов преобразования φ, заданных их координатами в базисе ***e***. Для доказательства того, что это будет искомая ортонормированная система собственных векторов, необходимо доказать следующую лемму:

Собственные векторы симметрического преобразования φ, *относящиеся к различным собственным значениям, между собой ортогональные.*

Пусть, в самом деле,

$$\mathbf{b}\varphi = \lambda_1 \mathbf{b}, \qquad \mathbf{c}\varphi = \lambda_2 \mathbf{c},$$

причем $\lambda_1 \neq \lambda_2$, так как

$$(\mathbf{b}\varphi, \mathbf{c}) = (\lambda_1 \mathbf{b}, \mathbf{c}) = \lambda_1 (\mathbf{b}, \mathbf{c}),$$
$$(\mathbf{b}, \mathbf{c}\varphi) = (\mathbf{b}, \lambda_2 \mathbf{c}) = \lambda_1 (\mathbf{b}, \mathbf{c}),$$

то из

$$(\mathbf{b}\varphi, \mathbf{c}) = (\mathbf{b}, \mathbf{c}\varphi)$$

следует

$$\lambda_1 (\mathbf{b}, \mathbf{c}) = \lambda_2 (\mathbf{b}, \mathbf{c}),$$

или, ввиду $\lambda_1 \neq \lambda_2$,

$$(\mathbf{b}, \mathbf{c}) = 0,$$

что и требовалось доказать.

Пары форм. Пусть дана пара действительных квадратичных форм от n неизвестных, $f(x_1, x_2, ..., x_n)$ и $g(x_1, x_2, ..., x_n)$. Существует ли такое невырожденное линейное преобразование неизвестных $x_1, x_2, ..., x_n$, которое одновременно приводило бы обе эти формы к каноническому виду?

В общем случае ответ будет отрицательным. Рассмотрим, например, пару форм

$$f(x_1, x_2) = x_1^2, \qquad g(x_1, x_2) = x_1 x_2.$$

Пусть существует невырожденное линейное преобразование

$$\left.\begin{aligned} x_1 &= c_{11}y_1 + c_{12}y_2 \\ x_2 &= c_{21}y_1 + c_{22}y_2 \end{aligned}\right\} \qquad (30)$$

которое приводит обе эти формы к каноническому виду. Для того чтобы форма f могла быть приведена преобразованием (30) к каноническому виду, один из коэффициентов c_{11}, c_{l2} должен быть равен нулю, иначе вошел бы член $2c_{11}c_{12}y_1y_2$. Меняя, если нужно, нумерацию неизвестных y_1, y_2, можно положить, что $c_{12} = 0$ и поэтому $c_{11} \neq 0$. Мы получим теперь, однако, что,

$$g(x_1, x_2) = c_{11}y_1 (c_{21}y_1 + c_{22}y_2) = c_{11}c_{21}y_1^2 + c_{11}c_{22}y_1y_2.$$

Так как форма g также должна была перейти в канонический вид, то $c_{11}c_{22}=0$, т.е. $c_{22}=0$, что вместе из $c_{12}=0$ противоречит невырожденности линейного преобразования (30).

Ситуация будет иной, если мы положим, что хотя бы одна из наших форм, например $g(x_1, x_2, ..., x_n)$, является положительно определенной. Тогда справедлива теорема:

Если f и g - пара действительных квадратичных форм от n неизвестных, причем вторая из них положительно определенная, то существует невырожденное линейное преобразование, которое одновременно приводит форму g к нормальному виду, а форму f к каноническому виду.

Для доказательства теоремы выполним сначала невырожденное линейное преобразование неизвестных $x_1, x_2, ..., x_n$,

$$X=TY,$$

которое приводит положительно определенную форму g к нормальному виду,

$$g(x_1, x_2, ..., x_n)=y_1^2+y_2^2+ \ ... \ +y_n^2.$$

Форма f перейдет при этом в некоторую форму φ от новых неизвестных,

$$f(x_1, x_2, ..., x_n) = \varphi(y_1, y_2, ...,y_n).$$

Сделаем теперь ортогональное преобразование неизвестных $y_1, y_2, ...,y_n$

$$Y=QZ,$$

которое приводит форму φ к главным осям,

$$\varphi(y_1, y_2, ...,y_n)= \lambda_1 z_1^2+\lambda_2 z_2^2+ \ ... \ +\lambda_n z_n^2.$$

Это преобразование переводит сумму квадратов неизвестных $y_1, y_2, ..., y_n$ в сумму квадратов неизвестных $z_l, z_2, ..., z_n$. В результате мы получаем

$$f(x_1, x_2, ..., x_n)= \lambda_1 z_1^2+\lambda_2 z_2^2+ \ ... \ +\lambda_n z_n^2$$

$$g(x_1, x_2, ..., x_n)=z_1^2+z_2^2+ \ ... \ +z_n^2,$$

т.е. линейное преобразование

$$X=(TQ)Z$$

является искомым.

6.5. Примеры решения задач

Пример. Привести к главным осям квадратичную форму

$$f(x_1,x_2,x_3, x_4)=2x_1x_2+2x_lx_3 — 2x_1x_4 — 2x_2x_3+2x_2x_4+2x_3x_4.$$

Матрица A этой формы имеет вид

$$A=\begin{pmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{pmatrix}.$$

Найдем ее характеристический многочлен:

$$|A-\lambda E|=\begin{vmatrix} -\lambda & 1 & 1 & -1 \\ 1 & -\lambda & -1 & 1 \\ 1 & -1 & -\lambda & 1 \\ -1 & 1 & 1 & -\lambda \end{vmatrix}=(\lambda-1)^3(\lambda+3).$$

Таким образом, матрица A имеет трехкратный характеристический корень 1 и простой характеристический корень —3. Мы уже можем, следовательно, написать тот канонический вид, к которому форма f приводится ортогональным преобразованием:

$$f=y^2_1+y^2_2+y^2_3+3y^2_4.$$

Найдем ортогональное преобразование, которое осуществляет это приведение.

Система линейных однородных уравнений (29) при $\lambda_0=1$ принимает вид

$$\begin{cases} -x_1+x_2+x_3-x_4=0, \\ x_1-x_2-x_3+x_4=0, \\ x_1-x_2-x_3+x_4=0, \\ -x_1+x_2+x_3-x_4=0. \end{cases}$$

Ранг этой системы равен 1, и поэтому для нее можно найти три линейно независимых решения. Ими будут, например, векторы

$$b_1=(1, 1, 0, 0),$$
$$b_2=(1, 0, 1, 0),$$
$$b_3=(-1, 0, 0, 1).$$

Ортогонализируя эту систему векторов, мы получим систему векторов

$$c_1=b_1=(1,\ 1,\ 0,\ 0),$$
$$c_2=-\frac{1}{2}c_1+b_2=\left(\frac{1}{2},-\frac{1}{2},\ 1,\ 0\right),$$
$$c_3=\frac{1}{2}c_1+\frac{1}{3}c_2+b_3=\left(-\frac{1}{3},\frac{1}{3},\frac{1}{3},\ 1\right).$$

С другой стороны, система линейных однородных уравнений (29) принимает при $\lambda_0=-3$ вид

$$\begin{cases} 3x_1+x_2+x_3-x_4=0, \\ x_1+3x_2-x_3+x_4=0, \\ x_1-x_2+3x_3+x_4=0, \\ -x_1+x_2+x_3+3x_4=0. \end{cases}$$

Ранг этой системы равен 3. Ее ненулевым решением служит вектор

$$c_4=(1, -1, -1,\ \ 1,).$$

Система векторов $\boldsymbol{c_1}$, $\boldsymbol{c_2}$, $\boldsymbol{c_3}$, $\boldsymbol{c_4}$ ортогональная. Нормируя ее, мы придем к ортонормированной системе векторов

$$c_1' = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0\right),$$
$$c_2' = \left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \sqrt{\frac{2}{3}}, 0\right),$$
$$c_3' = \left(-\frac{1}{2\sqrt{3}}, \frac{1}{2\sqrt{3}}, \frac{1}{2\sqrt{3}}, \frac{\sqrt{3}}{2}\right),$$
$$c_4' = \left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right).$$

Таким образом, форма *f* приводится к главным осям ортогональным преобразованием

$$y_1 = \frac{1}{\sqrt{2}} x_1 + \frac{1}{\sqrt{2}} x_2,$$
$$y_2 = \frac{1}{\sqrt{6}} x_1 - \frac{1}{\sqrt{6}} x_2 + \sqrt{\frac{2}{3}} x_3,$$
$$y_3 = -\frac{1}{2\sqrt{3}} x_1 + \frac{1}{2\sqrt{3}} x_2 + \frac{1}{2\sqrt{3}} x_3 + \frac{\sqrt{3}}{2} x_4,$$
$$y_4 = \frac{1}{2} x_1 - \frac{1}{2} x_2 - \frac{1}{2} x_3 + \frac{1}{2} x_4.$$

Следует отметить, что выбор системы линейно независимых собственных векторов, которые относятся к кратному собственному значению, является довольно неоднозначным, а поэтому существует много различных ортогональных преобразований, которые приводят форму *f* к каноническому виду. Мы описали лишь одно из них.

Индивидуальные тестовые задачи

1. Пусть в пространстве L_2 дано два ортонормированных базиса $\{\boldsymbol{e_1}, \boldsymbol{e_2}\}$ и $\{\boldsymbol{e_{1'}}, \boldsymbol{e_{2'}}\}$. Написать выражения векторов одного базиса через векторы другого и формулы преобразования координат произвольного вектора при переходе от первого базиса к второму, если

а) векторы второго базиса получены из векторов первого поворотом на угол α и перенумерацией базисных векторов;

б) $\boldsymbol{e_{1'}}=-\boldsymbol{e_1}$, $\boldsymbol{e_{2'}}=\boldsymbol{e_2}$.

2. Написать матрицу Г перехода от ортонормированного базиса $\{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}\}$ пространства L_3 к другому ортонормированному базису $\{\boldsymbol{e_{1'}}, \boldsymbol{e_{2'}}, \boldsymbol{e_{3'}}\}$, если

а) $\boldsymbol{e_{1'}}=\boldsymbol{e_2}$, $\boldsymbol{e_{2'}}=\boldsymbol{e_2}$, $\boldsymbol{e_{3'}}=\boldsymbol{e_3}$;

б) $\boldsymbol{e}_{1'}=\boldsymbol{e}_3$, $\boldsymbol{e}_{2'}=\boldsymbol{e}_1$, $\boldsymbol{e}_{3'}=\boldsymbol{e}_2$.

3. Как изменится матрица перехода от одного базиса к другому, если

а) поменять местами два вектора первого базиса?

б) поменять местами два вектора второго базиса?

в) записать векторы обеих базисов в обратном порядке?

4. Доказать, что матрицы

$$\Gamma_1=\begin{pmatrix} \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{pmatrix}, \quad \Gamma_2=\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{\sqrt{2}}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \end{pmatrix}$$

являются ортогональными.

5. Вектор $\boldsymbol{x}$ пространства L_n относительно ортонормированного базиса $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n\}$ имеет координаты $x_1, \dots, x_n$. Как выбрать в L_n новый базис, чтобы относительно него координатами вектора $\boldsymbol{x}$ стали числа 0, ..., 0, $|x|$?

6. Пусть $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n\}$ - базис пространства L_n и L_k — некоторое подпространство L_n размерности k. Доказать, что L_k может быть задано как совокупность всех векторов $\boldsymbol{x}\in L_n$, координаты x_i которых относительно базиса $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n\}$ удовлетворяют системе уравнений вида

$$a_{ij}x_j=0, \quad i, j = 1, \dots, n.$$

7. Дана система векторов $\boldsymbol{e}_1 = (1, 0, 1)$; $\boldsymbol{e}_2 = (1, 1, 0)$; $\boldsymbol{e}_{3=}$ (2, 3, 1).

а) Показать, что векторы $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ образовывают базис трехмерного пространства.

б) Сделать ортогонализацию данного базиса.

в) Найти координаты вектора $\boldsymbol{x}$=(4, —1, 5) в исходном и ортонормированном базисах.

8. В трехмерном пространстве даны точки $\boldsymbol{x}$=(1, 0, 2), $\boldsymbol{y}$ = (3,4,0), $\mathbf{z}$=(l, 2, 3). Для этих точек проверить аксиому треугольника в пространствах $C^{(3)}{}_2$, $C^{(3)}{}_1$, $C^{(3)}$.

9. На плоскости (x, y) нарисовать круг с центром в 0 и радиусом 1 в пространствах $C^{(2)}{}_2$, $C^{(2)}{}_1$, $C^{(2)}$.

10. Доказать, что открытый шар в просторные $R^{(n)}$ является выпуклым множеством.

11. Кокие из множеств [0, 3], [5, 7], [0, 3]$\bigcup$ [5, 7] в пространстве вещественных чисел R выпуклы, а какие нет?

7. Внешние формы в евклидовом пространстве

7.1. Начальные сведения из алгебры внешних форм

7.1.1. Условные обозначения. Альтернатор

1. В дальнейшем нам часто придется записывать суммы произвольного числа слагаемых. Объясним обозначения, которыми мы будем пользоваться для сжатости таких записей (с некоторыми обозначениями читатель уже знаком).

Если все слагаемые занумерованы по порядку: $a_1, a_2, \ldots, a_n$, то любое из них мы будем писать в виде a_i. Сумма всех слагаемых в этом случае будет обозначаться $\sum a_i$; таким образом:

$$\sum a_i = a_1 + a_2 + \ldots + a_n.$$

2. Далее мы будем иметь дело также с системами величин, которые обозначены несколькими индексами (например, a^i_k). Как правило, у нас будут встречаться суммы таких величин с отождествленными индексами, которые называют *индексами суммирования;* например,

$$\sum a^i_i = a^1_1 + a^2_2 + \ldots + a^n_n,$$

или

$$\sum a^{ik}_{ik} = \sum a^{1k}_{1k} + \sum a^{2k}_{2k} + \ldots + \sum a^{nk}_{nk}.$$

Обычно один из индексов суммирования мы будем писать сверху, другой - снизу. Во втором из предыдущих примеров имеются два индекса суммирования. Они независимы, соответственно чему обозначенные разными буквами.

3. Если индексов много, то их обозначают одной буквой с подиндексом. Например, $a^{i_1 i_2 \ldots i_k}$ ($i_1, i_2, \ldots, i_k = 1, 2, \ldots, n$) есть короткое обозначение некоторой системы величин в числе n^k. Пусть $b_{j_1 j_2 \ldots j_k}$ - другая аналогичная система величин. Тогда, например,

$$\sum a^{\alpha_1 \alpha_2 \cdots \alpha_k} b_{\alpha_1 \alpha_2 \ldots \alpha_k} = \\ = a^{11 \cdots 1} b_{11 \ldots 1} + \ldots + a^{i_1 i_2 \cdots i_k} b_{i_1 i_2 \ldots i_k} + \ldots \\ \ldots + a^{nn \cdots n} b_{nn \ldots n}$$

означает сумму всевозможных произведений $a^{i_1 i_2 \dots i_k}$ на $b_{i_1 i_2 \dots i_k}$ (в каждом слагаемом оба сомножителя имеют один и тот же набор индексов).

4. Кроме индексов суммирования, могут быть индексы, которые в суммировании не принимают участие; их называют *свободными.* Обозначение свободных индексом должно быть унифицировано во всех членах соотношений, включающих суммы, например,

$$\sum a_{i\alpha}^{\alpha} = \sum b_{i\alpha}^{\alpha}. \tag{1}$$

Здесь свободный индекс и слева и справа обозначен одной и той же буквой i. Соотношение (1) означает наличие нескольких равенств, общее число которых n. Они выходят последовательно при i=1, 2, ..., n.

5. В некоторых случаях мы будем писать суммы, совсем не употребляя индексов. Например,

$$\sum A = \dots + A + \dots$$

Такая запись означает, что нас интересует только сам факт наличия некоторой суммы, одно из слагаемых которой обозначено буквой A.

6. Проиллюстрируем все сказанное на примере сумм, в которых участвует так называемый альтернатор.

Альтернатор обозначается символом

$$\delta_{j_1 j_2 \dots j_k}^{i_1 i_2 \dots i_k},$$

где $i_1, i_2, \dots, i_k, j_1, j_2, \dots, j_k$ принимают значение 1, 2, . . ., n, и определяется следующими условиями:

$$\delta_{j_1 j_2 \dots j_k}^{i_1 i_2 \dots i_k} = \pm 1,$$

если $i_1, i_2, \dots, i_k$ есть некоторая перестановка значений индексов $j_1, j_2, \dots, j_k$, считая, что все эти значения различны; при этом берется + 1, если указанная перестановка четная, и — 1 - если нечетная. Во всех остальных случаях

$$\delta_{j_1 j_2 \dots j_k}^{i_1 i_2 \dots i_k} = \pm 0,$$

(т.е. если среди значений $i_1, i_2, \dots, i_k$, или среди значений $j_1, j_2, \dots, j_k$ есть одинаковые, а также если среди значений $i_1, i_2, \dots, i_k$ есть такие, каких нет среди $j_1, j_2, \dots, j_k$ и наоборот).

Пример. Пусть $A=(a_{ij})$ — квадратная $n \times n$-матрица. При $k=n=2$ рассмотрим сумму

$$D = \sum \delta_{12}^{i_1 i_2} a_{1i_1} a_{2i_2} = \delta_{12}^{11} a_{11} a_{21} + \delta_{12}^{12} a_{11} a_{22} + \delta_{12}^{21} a_{12} a_{21} + \delta_{12}^{22} a_{12} a_{22}.$$

Имеем

$$D = a_{11} a_{22} - a_{12} a_{21} = \det A.$$

Вообще при $k = n$ имеем

$$\sum \delta^{i_1 i_2 \cdots i_n}_{1\,2 \ldots n} a_{1 i_1} a_{2 i_2} \ \ldots \ a_{n i_n} = \det A.$$

Точно так же

$$\sum \delta^{i_1 i_2 \cdots i_n}_{1\,2 \ldots n} a_{i_1 1} a_{i_2 2} \ \ldots \ a_{i_n n} = \det A.$$

7. В частности, при $k = 1$ и при любом n альтернатор представляет собой символ Кронекера: $\delta_j^i = 1$, если $i = j$, $\delta_j^i = 0$, если $i \neq j$. В суммах этот символ действует как тождественный оператор; например,

$$\sum \delta_j^i a_i = a_j.$$

7.1.2. Сопряженные линейные пространства

1. Пусть L и L^* — два действительных линейных пространства.

Пусть с каждой парой элементов $a \in L^*$, $x \in L$ сопоставлено действительное число; обозначим его через (a, x). Определенную тем самым на $L^* \times L$ функцию мы назовем *сверткой,* если соблюдены следующие условия.

1) *Линейность по первому аргументу:*

$$(\alpha a_1 + \beta a_2,\ x) = \alpha (a_1,\ x) + \beta (a_2,\ x)$$

для любых

$$\alpha,\ \beta \in \mathbb{R},\ a_1,\ a_2 \in L^*,\ x \in L$$

(R, как обычно, обозначает *множество действительных чисел).*

2) *Линейность по второму аргументу:*

$$(a,\ \alpha x_1 + \beta x_2) = \alpha (a,\ x_1) + \beta (a,\ x_2)$$

для любых

$$\alpha,\ \beta \in \mathbb{R},\ a \in L^*,\ x_1,\ x_2 \in L.$$

3) *Невырожденность по первому аргументу:* если $(a, x) = 0$ при данном a и при любом $x \in L$, то $a = \theta^*$ (где θ^* — нулевой элемент в L^*).

4) *Невырожденность по второму аргументу:* если $(a, x)=0$ при любом $a \in L^*$ и при данном x, то $x = \theta$ (где θ — нулевой элемент в L**).**

Если на $L^* \times L$ свертка задана, то линейные пространства L и L^* мы будем называть *сопряженными* друг другу; легко видеть, что отношение сопряженности двух линейных пространств есть взаимным.

2. Предположим теперь, что L и L^* - конечномерные пространства одной и той же размерности = n. Выберем в L и L^* какие-нибудь *базисы;* обозначим их соответственно через $\tilde{e}_1, \ldots, \tilde{e}_n$ и $\tilde{e}^1, \ldots, \tilde{e}^n$. Для произвольных элементов $a \in L^*$, $x \in L$ напишем разложения

$$a = a_1 \tilde{e}^1 + \ldots + a_n \tilde{e}^n, \quad x = x^1 \tilde{e}_1 + \ldots + x^n \tilde{e}_n. \qquad (2)$$

Вследствие (2) имеем следующее общее выражение свертки:

$$(a, x) = \sum(\tilde{e}^i, \tilde{e}_j)a_i x^j. \tag{3}$$

Из (3) видно, что свертка будет определена на $L^*\times L$, если мы зададим матрицу сверток базисных элементов, т.е. матрицу чисел $(\tilde{e}^i, \tilde{e}_j)$. (Отсюда видно, что для данного L можно построить бесконечно много различных сопряженных пространств L^* (точнее говоря, по-разному сопряженных с L). Однако можно естественным образом определить понятие эквивалентности пространств, сопряженных с данным L так, что любые два пространства L^*_1, L^*_2, сопряженные с L, окажутся эквивалентными.) Легко увидеть, что для обеспечения обеих условий невырожденности 3) и 4) п.1 необходимо и достаточно, чтобы эта матрица была невырожденной; таким образом,

$$\det(\tilde{e}^i, \tilde{e}_j) \neq 0. \tag{4}$$

3. В некоторых специальных базисах $\tilde{e}^i$, $\tilde{e}_j$ матрицу $(\tilde{e}^i, \tilde{e}_j)$ можно сделать единичной. Вместе с тем упростится выражение (3). Имеет место следующая теорема.

Теорема. *Пусть на $L^* \times L$ как угодно задана свертка (a, x) и в L^* как угодно задан базис $\tilde{e}^1, \dots, \tilde{e}^n$; тогда в L найдется единственный базис $e_1, \dots, e_n$ такой, что*

$$(a, x) = \sum(\tilde{e}^i, \tilde{e}_j)a_i x^j. \tag{5}$$

где δ^i_j — символ Кронекера. Роли L^* и L можно поменять.

Доказательство теоремы вытекает из следующего утверждения: для любого набора чисел $\alpha^1, \dots, \alpha^n$ найдется единственный вектор $u \in L$, такой, что $(\tilde{e}^1, u) = \alpha^1, \dots, (\tilde{e}^1, u) = \alpha^n$. Чтобы убедиться в этом, разложим искомый вектор u по какому-нибудь базису: $u=\lambda^1\tilde{e}^1, \dots, \lambda^n\tilde{e}^n$. Мы получим для $\lambda^1, \dots, \lambda^n$ систему уравнений первой степени с главной матрицей $(\tilde{e}^i, \tilde{e}_j)$; полученная система однозначно разрешима вследствие (4).

Беря теперь в качестве $\alpha^1, \alpha^2, \dots, \alpha^n$ набор чисел 1,0, . . ., 0, найдем по предыдущему вектор u. Положим $e_1=u$. Аналогично по набору 0, 1, 0, ..., 0 найдем e_2 и т.д. Получении векторы $e_1, e_2, \dots, e_n$ удовлетворяют равенствам (5). Из этих же равенств следует, что векторы $e_1, e_2, \dots, e_n$ линейно-независимые.

Определение. Два базиса, из которых один принадлежит пространству L, другой — пространству L^*, называются *взаимными* или *дуальными,* если они удовлетворяют равенствам (5). В дальнейшем мы будем взаимные базисы обозначать более простым образом без пометки тильдой. Соответственно имеем

$$(e^i, e_j) = \delta^i_j, \quad \text{де } e^i \in L^*, \ e_j \in L. \tag{6}$$

4. Если разложения (2) даны по взаимным базисам, то

$$(a, x) = a_1x^1 + \ldots + a_nx^n. \qquad (7)$$

Доказательство. Выражение (7) следует с (3) и (6).

5. Будем исходить теперь из данного линейного пространства L, считая его, как и раньше, действительным и n-мерным. Обозначим через a произвольную *линейную форму* в пространстве L, т.е. действительную функцию точки $x \in L$, которая удовлетворяет условию линейности

$$a(\alpha x' + \beta x'') = \alpha a(x') + \beta a(x''), \qquad (8)$$

для любых

$$\alpha,\ \beta \in \mathbb{R},\ x',\ x'' \in L.$$

Во множестве всех линейных форм пространства L естественно вводятся линейные операции. А именно, если a, b — две произвольные формы, λ, μ — любые действительные числа, то в качестве формы $\lambda a+\mu b$ берется функция, значение которой на произвольном векторе $x \in L$ определяется равенством

$$(\lambda a + \mu b)(x) = \lambda a(x) + \mu b(x). \qquad (9)$$

Линейность такой функции непосредственно усматривается из (8) и (9).

На этот раз обозначим через L^* линейное пространство, элементами которого являются всевозможные линейные формы, данные на L, а линейные операции определены согласно (9). Заметим, что нулевым элементом в L^* служит форма θ^*, которая равна нулю на любом $x \in L$.

Легко показать, что L^* имеет размерность n, равную размерности L. Поэтому любая система линейно-независимых форм, взятых в числе n, составляет базис в L^*.

6. Назначим свертку двух произвольных элементов $a \in L^*$ и $x \in L$, полагая

$$(a,x) = a(x), \qquad (10)$$

т.е. в качестве (a, x) мы берем сейчас число, равное значению формы $a \in L^*$ на элементе $x \in L$. Требования, которые предъявляются к свертке согласно п. 1, при этом соблюдены (проверка условий 1)- 4) п. 1 не представляет затруднений).

Пространство L^* является сопряженным пространством L согласно определению п. 1. Далее на протяжении ряда подразделов мы будем под L^* иметь в виду именно это конкретное сопряженное пространство, т.е., состоящее из линейных форм пространства L.

7. Из теоремы п. 3 и из выражения (10) непосредственно следует, что каковы бы не были линейно-независимые формы $e^1(x), \ldots, e^n(x)$, $x \in L$, в пространстве L найдется единственный базис $e_1, \ldots, e_n$ такой, что

$$e^i(e_j) = \delta^i_j. \qquad (11)$$

Справедливо также утверждение, что для любого базиса e_1 ..., e_n в L обнаружится единственная линейно-независимая система форм $e^1(x)$, ..., $e^n(x)$, подчиненная условиям (11).

В дальнейшем через $e^i(x)$ и e_j или через e^i и e_j всегда обозначаются взаимные базисы в L^* и L (для которых соблюдено (11)).

8. Пусть произвольный вектор $x \in L$ разложен по базису e_1, ..., e_n

$$x = x^1 e_1 + \ldots + x^n e_n.$$

Тогда

$$e^j(x) = x^1 e^j(e_1) + \ldots + x^j e^j(e_j) + \ldots + x^n e^j(e_n).$$

Отсюда и вследствие (11) имеем

$$e^j(x) = x^j. \qquad (12)$$

Это значит, что координатную запись форм $e^j(x)$ при употреблении взаимного с ними базиса оказывается особенно простой: все коэффициенты форм $e^j(x)$ равны нулю, кроме одного, который занимает j-е место и равен единице. Вместе с тем можно сказать, что координаты любого вектора по базису e_1, ..., e_n суть значения на этом векторе форм взаимного базиса. То же самое выражается также равенством

$$x = e^1(x)\, e_1 + \ldots + e^n(x)\, e_n. \qquad (13)$$

7.1.3. Разложение полилинейной формы в сумму произведений линейных форм

1. Обозначим через L^k декартово k-кратное произведение пространства L на себя: $L^k = L \times L \times \ldots \times L$ (k раз). По определению пространства L^k его элементами являются упорядоченные наборы векторов из L, взятых в числе k; таким образом, $(x_1, \ldots, x_k) \in L^k$, если каждый из векторов x_1, ..., x_k принадлежит L. *Полилинейную форму* от векторных аргументов определим как действительную функцию a в L^k при условии линейности по каждому (векторному) аргумента

$$a(\alpha x_1' + \beta x_1'', x_2, \ldots, x_k) = \alpha a(x_1', x_2, \ldots, x_k) + \beta a(x_1'', x_2, \ldots, x_k).$$

Здесь условие линейности записано для первого аргумента.

2. Рассмотрим произвольную полилинейную формуа. Возьмем любые линейные формы e^1 ..., e^n в числе n при единственном условии их линейной независимости.

Теорема. *Численное значение $a(x_1, \ldots, x_k)$ формы a может быть представленное в виде*

$$a(x_1, \ldots, x_k) = \sum a_{j_1 \ldots j_k} e^{j_1}(x_1) \ldots e^{j_k}(x_k), \tag{14}$$

где $a_{j_1 \cdots j_k}$ *— коэффициенты, которые определяются данной формой a, а также выбором системы линейно-независимых форм* $e^1, \ldots, e^n$.

Доказательство. Пусть $e_1, \ldots, e_n$ — базис в L, взаимный с базисом, $e^1, \ldots, e^n$, в L^*. Запишем каждый из векторов $x_1, \ldots, x_k$ следующим образом:

$$x_1 = e^1(x_1)e_1 + \ldots + e^n(x_1)e_n,$$
$$\cdots\cdots\cdots\cdots\cdots$$
$$x_k = e^1(x_k)e_1 + \ldots + e^n(x_k)e_n.$$

Подставим эти разложения в левую часть (14) и воспользуемся линейностью формы $a(x_1, \ldots, x_k)$ по каждому аргументу. Мы получим правую часть выражения (14), где положено

$$a_{j_1 \ldots j_k} = a(e_{j_1}, \ldots, e_{j_k}). \tag{15}$$

Теорема доказана.

3. Выражение (14) само по себе, т.е. будучи уже доказанным, не предусматривает использования какого-нибудь базиса. Если же пользоваться базисом $e_1, \ldots, e_n$, который употребляется в доказательстве, то формы $e^1(x), \ldots, e^n(x)$ можно выразить в нем известным нам специальным образом; а именно,

$$e^j(x_k) = x_k^j,$$

где справа написаа j-я координата вектора x_k в базисе $e_1, \ldots, e_n$. В этом базисе выражение (14) принимает вид

$$a(x_1, \ldots, x_k) = \sum a_{j_1 \ldots j_k} x_1^{j_1} \ldots x_k^{j_k}. \tag{16}$$

Мы получаем представление формы $a(x_1, \ldots, x_k)$ в базисе $e_1, \ldots, e_n$. Этот базис в L можно взять заранее и вполне произвольно. Взаимный с ним базис найдется, но при употреблении записи (16) может не использоваться.

4. Из равенства (15) следует, что для данной формы a и для данных базисных форм $e^1 \ldots, e^n$, разложение (14) единственно, т.е. численные значение коэффициентов $a_{j_1 \cdots j_k}$ однозначно определены индексами $j_1, \ldots, j_k$.

7.1.4. Пространство полилинейных форм

1. Полилинейные формы, определенные в пространстве $L^*=L\times\ldots\times L$, образуют линейное пространство, если для произвольных

двух таких форм a и b определить линейные операции согласно равенству

$$(\alpha a+\beta b)(x_1, \ldots, x_k)=\alpha a(x_1, \ldots, x_k)+\beta b(x_1, \ldots, x_k).$$

Это линейное пространство мы обозначим через $\mathrm{T}^k(L)$ или просто через T^k и будем называть k-кратным тензорным произведением сопряженного просторную L^* на себя. В символьном виде это будет выглядеть так

$$\mathrm{T}^k=\mathrm{T}^k(L)=L^*\otimes L^*\otimes \ldots \otimes L^*.$$

Заметим, что само пространство L^* есть T^1. Поэтому

$$\mathrm{T}^k=\mathrm{T}^1\otimes \mathrm{T}^1\otimes \ldots \otimes \mathrm{T}^1.$$

Полилинейные формы как элементы пространства T^k называются *k-тензорами*, подробнее, - *ковариантными тензорами валентности k* в пространстве L. Линейные формы как элементы пространства $\mathrm{T}^1=L^*$ называются *одновалентными ковариантными тензорами,* или *ковариантными векторами,* или *ковекторами*. Действительные числа мы будем называть *тензорами нулевой валентности.* Действительную ось R обозначим соответственно сказанному через T^0.

2. Поменяв ролями L и L^*, получим аналогично предыдущему *контравариантные тензоры валентности k,* как полилинейные формы, отображающие $(L^*)^k$ в R.

3. *Тензорным произведением* или просто *произведением тензора a на тензор b*

$$a\in \mathrm{T}^k(L), \quad b\in \mathrm{T}^l(L)$$

называется тензор, который обозначается $a\otimes b$, принадлежит пространству $\mathrm{T}^{k+l}(L)$ и определяется в виде полилинейной формы равенством

$$(a\otimes b)(x_1, \ldots, x_k, x_{k+1}, \ldots, x_{k+l})= \\ =a(x_1, \ldots, x_k)\,b(x_{k+1}, \ldots, x_{k+l}). \tag{17}$$

Здесь слева и справа написанные численные значения полилинейных форм при любом выборе векторов $x_1, \ldots, x_k, x_{k+1}, \ldots, x_{k+l}\in L$, взятых независимо друг от друга.

4. Тензорное произведение обладает следующими свойства:

$$(a_1+a_2)\otimes b=a_1\otimes b+a_2\otimes b,$$
$$a\otimes(b_1+b_2)=a\otimes b_1+a\otimes b_2,$$
$$(\alpha a)\otimes b=\alpha(a\otimes b),$$
$$(a\otimes b)\otimes c=a\otimes(b\otimes c).$$

Доказательство этих свойств тривиально.

Последнее свойство позволяет писать $a \otimes b \otimes c$ без указания ассоциаций. Вместе с тем определено тензорное произведение любого числа любых тензорных сомножителей.

5. Легко убедиться, что, вообще говоря, $a \otimes b$ не совпадает з $b \otimes a$ (чтобы убедиться в этом, достаточно написать (17) для $b \otimes a$).

6. Попутно заметим, что равенство тензоров a и b следует понимать как тождественное совпадение полилинейных форм $a{:}L^k \to \mathrm{R}$ и $b{:}\ L^k \to R$. Равенство $a = 0$ означает отображение $a{:}\ L^k \to 0$.

7. Пусть $e^1, \ldots, e^n$ линейно-независимые линейные формы в L. Мы рассматриваем их как одновалентные тензоры, т.е. как элементы пространства $\mathrm{T}^1 = L^*$. Тогда

$$e^{j_1} \otimes e^{j_2} \otimes \ \ldots \ \otimes e^{j_k} \in \mathrm{T}^k. \tag{18}$$

Имеет место следующая теорема

Теорема. *Множество всех тензорных произведений* (18), *т.е. отвечающих всевозможным наборам индексов* $j_1, \ldots, j_k$ ($j_1=1, 2, \ldots, n; \ldots;$ $j_k=1, 2, \ldots, n$) *составляет базис в* T^k.

Доказательство. Теорема уже доказана в п. 2—5 п. 7.1.3. Установленные там предложения означают, что для любого $a \in T^k$ имеет место единственное разложение

$$a = \sum a_{j_1 \ldots j_k} e^{j_1} \otimes \ \ldots \ \otimes e^{j_k} \tag{19}$$

(см. формулу (14). А это и значит, что

$$e^{j_1} \otimes \ldots \otimes e^{j_k}$$

составляют базис.

Следствие. *Размерность пространства* T^k равна n^k.

Замечание. Коэффициенты $a_{j_1 \ldots j_k}$ разложения (19) называются *координатами тензора a* (по базису $e^{j_1} \otimes \ldots \otimes e^{j_k}$ в T^k).

8. Итак, мы имеем счетную последовательность пространств

$$\mathrm{T}^0,\ \mathrm{T}^1,\ \mathrm{T}^2,\ \ldots,\ \mathrm{T}^k,\ \ldots,$$

построенных по данному линейному пространству L. В каждом T^k определены линейные операции, в каждой паре T^k, T^l (допуская $k=l$) определена операция тензорного произведения:

$$a \otimes b,\ a \in \mathrm{T}^k,\ b \in \mathrm{T}^l,\ a \otimes b \in \mathrm{T}^{k+l}$$

Элементы этих пространств, как объекты указанных операций, называются *ковариантными тензорами* (над L). Если $a \in \mathrm{T}^k$, то a называется *тензором валентности* k. Пространство T^k является n^k-мерным.

7.1.5. Альтернация полилинейных форм

1. *Альтернация* полилинейной формы *a* обозначается через [*a*] и определяется (независимо от базиса) следующим образом:

$$[a]\left(x_1, \ldots, x_k\right) = \frac{1}{k!}\sum \delta^{i_1 \ldots i_k}_{1 \ldots k} a\left(x_{i_1}, \ldots, x_{i_k}\right); \tag{20}$$

здесь справа принимает участие альтернатор, о котором говорилось в п. 7.1.1. В частности, имеем при k=1, 2, 3 соответственно

$$[a](x) = a(x),$$

$$[a](x_1, x_2) = \frac{1}{2!}(a(x_1, x_2) - a(x_2, x_1)),$$

$$[a](x_1, x_2, x_3) = \frac{1}{3!}(a(x_1, x_2, x_3) + a(x_2, x_3, x_1) +$$

$$+ a(x_3, x_1, x_2) - a(x_2, x_1, x_3) - a(x_1\ x_3, x_2) - a(a_3, x_2, x_1)).$$

Равенство (20) следует понимать с учетом следующего условия: если те же аргументы x_1 ..., x_k, которые стоят слева, написаны в другом (не натуральному) порядка, то в таком же порядке должны быть написаны нижние индексы альтернатора справа.

1) Альтернация полилинейной формы сама есть полилинейная форма. А именно (для первого аргумента)

$$[a]\left(\alpha x_1' + \beta x_1'', x_2, \ldots, x_k\right) =$$

$$= \alpha [a]\left(x_1', x_2, \ldots, x_k\right) + \beta [a]\left(x_1'', x_2, \ldots, x_k\right).$$

2) Альтернация полилинейной формы кососимметрична по любой паре аргументов. Например, для первой пары аргументов

$$[a](x_1, x_2, x_3, \ldots, x_k) = -[a](x_2, x_1, x_3, \ldots, x_k). \tag{21}$$

Эти свойства непосредственно вытекают из определения альтернатора, из равенства (20) и из высказанного выше условия по поводу порядка расположения индексов в левой и правой частях равенства (20).

2. Полилинейная форма $\omega = \omega\ (x_1, x_2, \ldots, x_k)$ называется *косой* или *внешней,* если $[\omega] = \omega$. *Косой является всякая форма, которая имеет не меньше двух аргументов* (*k*≥2) *и кососимметричная по любой паре своих аргументов.* В самом деле, в случае кососимметричности по любой паре аргументов имеем

$$\delta^{i_1 \ldots i_k}_{1 \ldots k} a\left(x_{i_1}, \ldots, x_{i_k}\right) = a\left(x_1, \ldots, x_k\right)$$

для любого набора индексов i_1 ... i_k. Таким образом, все слагаемые суммы в правой части равенства (20) в этом случае одинаковы и равны $a(x_1, ..., x_k)$. Так как число слагаемых равно $k!$, то

$$[a](x_1, \ldots, x_k) = a(x_1, \ldots, x_k). \tag{22}$$

Обратно, если $k \geq 2$ и форма косая (совпадает со своей альтернацией), то она кососиметрична по любой паре своих аргументов. Это утверждение следукт из косой симметрии альтернации (см. равенства (21) и (22)). Кроме этих форм, к числу внешних следует отнести всякую линейную форму, т.е. всякую форму при $k=1$ (поскольку для всякой линейной формы $[\omega]=\omega$).

3. Для любой формы повторная альтернация совпадает с однократной

$$[[a]](x_1, \ldots, x_k) = [a](x_1, \ldots, x_k).$$

Доказательство. Утверждение очевидно, так как альтернация кососиметрична и полилинейна.

4. Пусть $k \geq 2$. Рассмотрим коэффициенты разложения полилинейной формы ω по базису в T^k, отвечающему некоторому базису $e^1..., e^n \in L^*$. Если $e_1, ..., e_n$ — взаимный базис в L, то

$$\omega_{i_1 i_2 i_3 \ldots i_k} = \omega\left(e_{i_1}, e_{i_2}, e_{i_3}, \ldots, e_{i_k}\right).$$

Если форма ω косая, то при обмене местами любых двух ее аргументов она меняет знак. Следовательно, *коэффициенты косой формы кососимметричны по любой паре индексов;* например, для первой пары индексов

$$\omega_{i_1 i_2 i_3 \ldots i_k} = -\omega_{i_2 i_1 i_3 \ldots i_k}.$$

Обратно, *если в каком-либо базисе коэффициенты полилинейной формы* ω *кососимметричны, по любой паре индексов, то форма* ω *является косой* (доказательство предлагаем читателю).

5. Имеют место следующие предложения; они легко усматриваются, и мы приведем их без доказательства.

1) *Если у косой формы два аргумента принимают одинаковые значения, то форма обращается в нуль.*

2) *Если аргументы косой формы линейно-зависимы, то форма равна* 0.

3) *Если число аргументов $k>n$, то косая форма тождественно равна нулю, т.е. равна* 0 *на любом наборе своих аргументов.*

6. В дальнейшем нам потребуется следующая лемма.

Лемма. *Если две формы удовлетворяют тождесту*

$$a(x_1, \ldots, x_k, x_{k+1}, \ldots, x_{k+l}) = \\ = b(x_{k+1}, \ldots, x_{k+l}, x_1, \ldots, x_k), \quad (23)$$

то

$$[a](x_1, \ldots, x_k, x_{k+1}, \ldots, x_{k+l}) = \\ = [b](x_{k+1}, \ldots, x_{k+1}, x_1, \ldots, x_k). \quad (24)$$

Доказательство. Возьмем произвольный член суммы, которая стоит в левой части (24)

$$\delta^{\alpha_1 \ldots \alpha_k \alpha_{k+1} \ldots \alpha_{k+l}}_{1 \ldots k \, (k+1) \ldots (k+l)} a\left(x_{\alpha_1} \ldots x_{\alpha_k} x_{\alpha_{k+1}} \ldots x_{\alpha_{k+l}}\right).$$

Для произвольного члена правой части имеем

$$\delta^{\beta_1 \ldots \beta_l \beta_{l+1} \ldots \beta_{l+k}}_{(k+1) \ldots (k+l) \, 1 \ldots k} b\left(x_{\beta_1} \ldots x_{\beta_l} x_{\beta_{l+1}} \ldots x_{\beta_{l+k}}\right) = \\ = \delta^{\beta_{l+1} \ldots \beta_{l+k} \beta_1 \ldots \beta_l}_{1 \ldots k \, (k+1) \ldots (k+l)} b\left(x_{\beta_1} \ldots x_{\beta_l} x_{\beta_{l+1}} \ldots x_{\beta_{l+k}}\right).$$

Установим взаимно однозначное соответствие между членами сумм в левой и правой частях (24), полагая соответствующими члены, для которых

$$\beta_{l+1} = \alpha_1, \ \ldots, \ \beta_{l+k} = \alpha_k, \quad \beta_1 = \alpha_{k+1}, \ \ldots, \ \beta_l = \alpha_{k+l}.$$

Вследствие (23) соответствующие члены будут равны. Тем самым лемма доказана.

7.1.6. Второе выражение альтернации

1. Пусть имеются линейные формы u_1 ..., u_k. Перемножим их тензорно в произвольном порядке. Мы получим полилинейную форму

$$u_{s_1} \otimes \ldots \otimes u_{s_k}.$$

Напишем числовое значение альтернации этой формы:

$$\left[u_{s_1} \otimes \ldots \otimes u_{s_k}\right](x_1, \ldots, x_k) = \\ = \frac{1}{k!} \sum \delta^{l_1 \ldots l_k}_{1 \ldots k} u_{s_1}\left(x_{l_1}\right) \ldots u_{s_k}\left(x_{l_k}\right). \quad (25)$$

Выражение (25) неудобно для использования. Дело в том, что оно представляет не альтернацию $u_{s_1} \otimes \ldots \otimes u_{s_k}$, по данным u_1, ..., u_k, а ее численное значение на произвольном наборе аргументов x_1, ..., x_k. От этого набора мы не можем отвлечься, поскольку в правой части (25) приходится следить за порядком расположения аргументов в каждом слагаемом. Однако выражение $[u_{s_1} \otimes ... \otimes u_{s_k}](x_1, ..., x_k)$ легко освободить от этого недостатка. А именно, имеет место тождество

$$\sum \delta^{j_1 \dots j_k}_{1 \dots k} u_{s_1}(x_{j_1}) \dots u_{s_k}(x_{j_k}) = \\ = \sum \delta^{i_1 \dots i_k}_{s_1 \dots s_k} u_{i_1}(x_1) \dots u_{i_k}(x_n). \quad (26)$$

Важно заметить, что суммирование в правой части (26) выполняется не по номерам аргументов $x_1, \dots, x_k$, а по номерам самих форм $u_1, ..., u_k$. В каждом слагаемом правой части (26) сомножители записаны в натуральном порядке аргументов $x_1, ..., x_k$. Из (25) и (26) получаем

$$[u_{s_1} \otimes \dots \otimes u_{s_k}](x_1, \dots, x_k) = \\ = \frac{1}{k!} \sum \delta^{i_1 \dots i_k}_{s_1 \dots s_k} (u_{i_1} \otimes \dots \otimes u_{i_k})(x_1, \dots, x_k). \quad (27)$$

Доказательство тождества (26). В сумме (26) слева достаточно учитывать лишь те слагаемые, в которых набор $j_1, ..., j_k$ является перестановкой набора 1, ..., k. Будем считать слагаемые в сумме (26) слева различными, если они отвечают различным расположениям индексов $j_1, ..., j_k$ (не обращая внимания на численные значения слагаемых). Аналогично определим различные слагаемые суммы (26) справа. Россматриваемые с такой точки зрения, все слагаемые этих сумм должны считаться попарно различными, как слева, так и справа.

Возьмем в сумме (26) слева слагаемое, отвечающей некоторой перестановке $j_1, ..., j_k$, и рассмотрим произведение

$$u_{s_1}(x_{j_1}) \dots u_{s_k}(x_{j_k}).$$

Сделаем здесь перемену мест сомножителей, располагая их в порядке номеров аргументов; получим

$$u_{i_1}(x_1) \dots u_{i_k}(x_k).$$

Очевидно, что

$$u_{s_1}(x_{j_1}) \dots u_{s_k}(x_{j_k}) = u_{i_1}(x_1) \dots u_{i_k}(x_k). \quad (28)$$

Таким способом с перестановкой $j_1, ..., j_k$ сопоставляется перестановка $i_1, \dots, i_k$. Одновременно мы сопоставим со взятым слагаемым суммы (26) слева то слагаемое суммы (26) справа, которое отвечает перестановке $i_1, ..., i_k$. Установленное соответствие между слагаемыми сумм (25) и (26) будет взаимно однозначным, поскольку двум разным перестановкам $j_1, ..., j_k$ сопоставляются также разные перестановки $i_1, ..., i_k$.

Легко убедиться, что

$$\delta^{j_1 \dots j_k}_{1 \dots k} = \delta^{i_1 \dots i_k}_{s_1 \dots s_k}. \tag{29}$$

Из (28) и (29) следует, что в суммах (26) соответствующие слагаемые слева и справа численно совпадают. Тем самым тождественность доказана.

2. Пусть имеется полилинейная форма с численным значением

$$a(x_1, \dots, x_k) = \sum a_{\alpha_1 \dots \alpha_k} e^{\alpha_1}(x_1) \dots e^{\alpha_k}(x_k).$$

Из определения альтернации вытекает, что альтернация суммы форм равна сумме их альтернации и что числовой коэффициент можно выносить за знак альтернации. Поэтому

$$[a](x_1, \dots, x_k) = \\ = \sum a_{\alpha_1 \dots \alpha_k} [e^{\alpha_1} \otimes \dots \otimes e^{\alpha_k}](x_1, \dots, x_k). \tag{30}$$

Здесь мы можем теперь считать, что

$$[e^{\alpha_1} \otimes \dots \otimes e^{\alpha_k}](x_1, \dots, x_k) = \\ = \frac{1}{k!} \sum \delta^{\alpha_1 \dots \alpha_k}_{i_1 \dots i_k} (e^{i_1} \otimes \dots \otimes e^{i_k})(x_1, \dots, x_k). \tag{31}$$

Выражение (31) непосредственно следует из (27); нужно только учесть изменение обозначений (в частности, что номера базисных форм поставлены сверху).

7.1.7. Альтернация тензоров

1. Пусть тензор a есть полилинейная форма со значением $a(x_1, \dots, x_k)$. Тогда *альтернацией* $[a]$ *тензора* a мы назовем альтернацию формы a. Заметим, что при этом мы не можем рассматривать тензор a формально, т.е. просто как элемент пространства T^k, поскольку альтернация определена путем перестановок аргументов $x_1 \dots, x_k$; таким образом мы вынуждены использовать свойства формы как функции. Но, если выбрать базисные формы $e^1,\dots,e^n$, то альтернацию тензора $a \in T^k$ можно выразить с помощью тех операций, которые введены в T^k и в парах T^k, T^l.

А именно, тензор a может быть записан в виде

$$a = \sum a_{j_1 \dots j_k} e^{j_1} \otimes \dots \otimes e^{j_k}. \tag{32}$$

Из (32) и из формулы (30) имеем

$$[a] = \sum a_{j_1 \dots j_k} [e^{j_1} \otimes \dots \otimes e^{j_k}], \tag{33}$$

где правая часть определена согласно (31):

$$[e^{j_1}\otimes \dots \otimes e^{j_k}]=\frac{1}{k!}\sum \delta^{j_1 \dots j_k}_{i_1 \dots i_k}(e^{i_1}\otimes \dots \otimes e^{i_k}). \tag{34}$$

Замечание. В дальнейшем в записи произведения тензоров мы позволим себе часто опускать знак $\otimes$.

В таком случае формула (33), например, запишется проще:

$$[a]=\sum a_{j_1 \dots j_k}[e^{j_1} \dots e^{j_k}].$$

7.1.8. Внешнее произведение внешних форм

1. В этому подразделе мы определим некоторое действие над внешними (косыми) формами, называемое их внешним произведением. Условимся называть *степенью* внешней формы число ее (векторных) аргументов. Таким образом, если дана внешняя форма ω=ω(x_1, ..., x_k), то ее степень =*k*. Внешнюю форму степени *k* часто называют *k-формой.*

2. Пусть дано две внешних формы $\omega^k{}_1$, $\omega^l{}_2$, где *k*, *l* — степени данных форм.

Определение. *Внешним произведением* формы $\omega^k{}_1$ на форму $\omega^l{}_2$ называется внешняя форма, которая обозначается и выражается согласно равенству

$$\omega_1^k \wedge \omega_2^l=\frac{(k+l)!}{k!\,l!}\left[\omega_1^k\otimes\omega_2^l\right]. \tag{35}$$

Внешнее произведение $\omega^k{}_1\wedge\omega^l{}_2$ есть внешняя (коса) форма, поскольку в правой части (35) произведение $\omega^k{}_1\otimes\omega^l{}_2$ проальтернировано.

3. Для альтернации имеют место соотношения:

$$[\Sigma\omega]=\sum[\omega], \qquad [\alpha\omega]=\alpha[\omega].$$

Отсюда получаются соответствующие свойства внешнего произведения:

a) $\left(\alpha\omega_1^k\right)\wedge\omega_2^l=\omega_1^k\wedge\left(\alpha\omega_2^l\right)=\alpha\left(\omega_1^k\wedge\omega_2^l\right)$,

b) $\left(\omega_1^k+\omega_2^k\right)\wedge\omega_3^l=\omega_1^k\wedge\omega_3^l+\omega_2^k\wedge\omega_3^l$.

Доказательства этих свойств вытекают из определение п. 1, и мы проводить их не будем.

c) $\omega_1^k\wedge\omega_2^l=(-1)^{kl}\,\omega_2^l\wedge\omega_1^k$.

Докажем это свойство. Очевидно, что

$$\omega_1^k(x_1, \dots, x_k)\,\omega_2^l(x_{k+1}, \dots, x_{k+l})=$$
$$=\omega_2^l(x_{k+1}, \dots, x_{k+l})\,\omega_1^k(x_1, \dots, x_k).$$

Отсюда и вследствие леммы п. 6 п. 7.1.5

$$(\omega_1^k \wedge \omega_2^l)(x_1, \ldots, x_k, x_{k+1}, \ldots, x_{k+l}) = \\ = (\omega_2^l \wedge \omega_1^k)(x_{k+1}, \ldots, x_{k+l}, x_1, \ldots, x_k).$$

Следовательно, ввиду кососимметричности правой части предыдущего равенства

$$(\omega_1^k \wedge \omega_2^l)(x_1, \ldots, x_k, x_{k+1}, \ldots, x_{k+l}) = \\ = (-1)^{kl}(\omega_2^l \wedge \omega_1^k)(x_1, \ldots, x_l, x_{l+1}, \ldots, x_{l+k}).$$

Следствие. Если $l=k$ — *нечетное и* $\omega^l{}_2 = \omega^k{}_1$, то $\omega^k{}_1 \wedge \omega^l{}_2 = 0$.

Частный случай. Для двух линейных форм u, v имеем

$$(u \wedge v)(x_1, x_2) = -(v \wedge u)(x_1, x_2).$$

d) Свойство ассоциативности

$$(\omega_1^k \wedge \omega_2^l) \wedge \omega_3^m = \omega_1^k \wedge (\omega_2^l \wedge \omega_3^m).$$

Доказательство этого свойства будет дано позже (см. п. 7.1.9). Сначала установим тождество

$$[[e^{i_1} \ldots e^{i_k}][e^{j_1} \ldots e^{j_l}]] = [e^{i_1} \ldots e^{i_k} e^{j_1} \ldots e^{j_l}]. \qquad (36)$$

Здесь $e^{i_1}, \ldots, e^{i_j}$ суть базисные формы.

Для доказательства тождества (36) заметим, что

$$[e^{i_1} \ldots e^{i_k}] = \frac{1}{k!} \sum \delta^{i_1 \ldots i_k}_{\alpha_1 \ldots \alpha_k} e^{\alpha_1} \ldots e^{\alpha_k}.$$

$$[e^{j_1} \ldots e^{j_l}] = \frac{1}{l!} \sum \delta^{j_1 \ldots j_l}_{\beta_1 \ldots \beta_l} e^{\beta_1} \ldots e^{\beta_l}.$$

Перемножая и применяя альтернацию, получим

$$[[e^{i_1} \ldots e^{i_k}][e^{j_1} \ldots e^{j_l}]] = \\ = \frac{1}{k!\, l!} \sum \delta^{i_1 \ldots i_k}_{\alpha_1 \ldots \alpha_k} \delta^{j_1 \ldots j_l}_{\beta_1 \ldots \beta_l} [e^{\alpha_1} \ldots e^{\alpha_k} e^{\beta_1} \ldots e^{\beta_l}].$$

При перестановке в правой части предыдущего равенства двух индексов среди $\alpha_1, \ldots, \alpha_k$ или $\beta_1, \ldots, \beta_l$, будет меняться знак альтернации. Но при этом будет одновременно меняться знак соответствующего альтернатора. Поэтому в каждом члене индексы $\alpha_1, \ldots, \alpha_k$, $\beta_1, \ldots, \beta_l$ можно привести к стандартному расположению $i_1, \ldots, i_k\ j_1, \ldots, j_l$; тогда все члены справа окажутся одинаковыми и число их будет $k!\ l!$. Таким образом, правая часть примет вид

$$[e^{i_1} \ldots e^{i_k} e^{j_1} \ldots e^{j_l}].$$

Тождество (36) доказано.

7.1.9. Внешнее произведение базисных форм

1. Имеем для базисных форм (первой степени)

$$e^i \wedge e^j = 2!\,[e^i e^j]. \tag{37}$$

Отсюда

$$(e^i \wedge e^j) \wedge e^k = \frac{3!}{2!\,1!}\,[e^i \wedge e^j,\ e^k] = \\ = 3!\,[[e^i e^j]\,e^k] = 3!\,[e^i e^j e^k] \tag{38}$$

(см. тождество (36)). Аналогично

$$e^i \wedge (e^j \wedge e^k) = 3!\,[e^i e^j e^k].$$

Следовательно,

$$(e^i \wedge e^j) \wedge e^k = e^i \wedge (e^j \wedge e^k) = e^i \wedge e^j \wedge e^k.$$

Это равенство выражает ассоциативное свойство внешнего произведения базисных форм. Но чтобы доказать свойство ассоциативности в общем виде (см. п. 7.1.8, свойство d)), т.е. для внешнего произведения трех любых внешних форм, предыдущее равенство приходится обобщить.

2. Прежде всего, пользуясь тождеством (38) и применяя индукцию, мы получим

$$e^{i_1} \wedge \ \dots \ \wedge e^{i_k} = k!\,[e^{i_1} \dots e^{i_k}]. \tag{39}$$

Используя формулы (36) и (39), найдем

$$(e^{i_1} \wedge \ \dots \ \wedge e^{i_k}) \wedge (e^{j_1} \wedge \ \dots \ \wedge e^{j_l}) = \\ = e^{i_1} \wedge \ \dots \ \wedge e^{i_k} \wedge e^{j_1} \wedge \ \dots \ \wedge e^{j_l}. \tag{40}$$

Отсюда

$$((e^{i_1} \wedge \dots \wedge e^{i_k}) \wedge (e^{j_1} \wedge \dots \wedge e^{j_l})) \wedge (e^{s_1} \wedge \dots \wedge e^{s_m}) = \\ = (e^{i_1} \wedge \dots \wedge e^{i_k}) \wedge ((e^{j_1} \wedge \dots \wedge e^{j_l}) \wedge \\ \wedge (e^{s_1} \wedge \dots \wedge e^{s_m})). \tag{41}$$

3. Из последнего равенства ассоциативное свойство d) п. 7.1.8 вытекает непосредственно. Достаточно каждую форму $\omega^k{}_1$, $\omega^l{}_2$, $\omega^m{}_3$ разложить по надлежащему базису и выполнить внешние произведения этих форм почленно; тогда из (41) получится нужное соотношение

$$(\omega_1^k \wedge \omega_2^l) \wedge \omega_3^m = \omega_1^k \wedge (\omega_2^l \wedge \omega_3^m).$$

Разложению внешних форм по базису посвящен следующий подраздел.

7.1.10. Пространство внешних форм данной степени и базис в нем

1. Внешние формы данной степени k (коротко: k-формы) представляют линейное пространство, которое является подпространством в T^k. В самом деле, если $\omega_1, \omega_2 \in T^k$ и

$$[\omega_1] = \omega_1, \quad [\omega_2] = \omega_2,$$

то

$$[\alpha_1\omega_1 + \alpha_2\omega_2] = \alpha_1 [\omega_1] + \alpha_2 [\omega_2] = \alpha_1\omega_1 + \alpha_2\omega_2.$$

Линейное пространство k-форм обозначим через Λ^k*;* имеем $\Lambda^k \subset T^k$. Элементы Λ^k называются также *косыми тензорами.*

Отметим, что $\Lambda^1 = T^1 = L^*$.

2. Пусть e^1 ..., e^n - базис в L^*. Тогда, как мы знаем, всевозможные произведения

$$e^{i_1} \otimes \ldots \otimes e^{i_k}$$

составляют базис в T^k. Соответственно, для произвольного k-тензора, т.е. для произвольной формы $\omega \in T^k$ имеем

$$\omega = \sum \omega_{i_1 \ldots i_k} e^{i_1} \ldots e^{i_k}. \tag{42}$$

Здесь мы воспользовались договоренностью опускать для краткости записи знак $\otimes$. Числа $\omega_{i_1 \ldots}\omega_{i_k}$ суть коэффициенты разложения (42), или координаты тензора ω. Если тензор ω — косой и $k \geq 2$, то его координаты $\omega_{i_1 \ldots}\omega_{i_k}$ обладают косой симметрией по любой пари индексов; например

$$\omega_{i_1 i_2 i_3 \ldots i_k} = -\omega_{i_2 i_1 i_3 \ldots i_k}.$$

Обратно, если $\omega_{i_1 \ldots}\omega_{i_k}$ — кососиметричен по всем парам индексов, то ω — косой тензор (т.е. является косой формой). Тем самым мы имеем характеристику подпространства Λ^k в пространстве T^k с помощью только таких соотношений, которые специфичны для T^k (точнее, с помощью линейных операций в T^k и операции тензорного перемножения). Заметим, что выбор базиса $e^1,\ldots,e^n$ при этом безразличный.

3. Теперь мы укажем базис в подпространстве Λ^k. Прежде всего, учтем, что если форма ω косая, то она равна своей альтернации.

Поэтому из (42) выходит равенство

$$\omega = \sum \omega_{i_1 \ldots i_k} [e^{i_1} \ldots e^{i_k}]. \tag{43}$$

Пользуясь формулой (39), форму (43) можно представить в виде

$$\omega = \frac{1}{k!}\sum \omega_{i_1 \dots i_k} e^{i_1} \wedge \dots \wedge e^{i_k}, \tag{44}$$

или

$$\omega = *\sum \omega_{i_1 \dots i_k} e^{i_1} \wedge \dots \wedge e^{i_k}, \tag{45}$$

где звездочка перед знаком суммы означает, что суммирование ведется по индексам $i_1 \dots i_k$ при условии, что $i_1 < i_2 < \dots < i_k$.

Теорема. *Всевозможные наборы внешних произведений*

$$e^{i_1} \wedge \dots \wedge e^{i_k}$$

при условии, что $i_1 < i_2 < \dots < i_k$, *представляют базис в* Λ^k, *т.е. в пространстве внешних форм данной степени k.*

Доказательство. Ввиду наличие разложения (45), достаточно доказать, что указанные наборы внешних произведений линейно независимы. Предположим, что есть тождественное соотношение

$$*\sum \omega_{i_1 \dots i_k} e^{i_1} \wedge \dots \wedge e^{i_k} = 0, \tag{46}$$

где $\omega_{i_1 \dots} \omega_{i_k}$ - некоторые числа; при этом $i_1 < i_2 < \dots < i_k$.

Заметим, что разложение вида (42) всегда может быть восстановлено по заранее данному разложению вида (45). Поэтому из (46) имеем

$$\sum \omega_{i_1 \dots i_k} e^{i_1} \dots e^{i_k} = 0.$$

Отсюда $\omega_{i_1 \dots} \omega_{i_k}$ =0. Таким образом, $\omega_{i_1 \dots} \omega_{i_k}$ =0 как следствие (46). Теорема доказана.

Следствие. *Пространство всех внешних форм степени k имеет размерность* C^k_n.

Пример. Пусть k=2. Тогда разложение внешней формы ω будет иметь вид

$$\begin{aligned}\omega = *\sum \omega_{i_1 i_2} e^{i_1} \wedge e^{i_2} = \\ = \omega_{12} e^1 \wedge e^2 + \omega_{13} e^1 \wedge e^3 + \dots + \omega_{1n} e^1 \wedge e^n + \\ + \omega_{23} e^2 \wedge e^3 + \dots + \omega_{2n} e^2 \wedge e^n + \\ + \dots \quad \dots \quad \dots \quad + \\ + \omega_{(n-1)n} e^{n-1} \wedge e^n.\end{aligned}$$

7.1.11. Вычисление одночленных форм

1. Разложим векторы $x_1, \dots, x_k$ по базису $e_1, \dots, e_n$:

$$x_1 = x_1^1 e_1 + \ldots + x_1^n e_n,$$
$$\cdots\cdots\cdots\cdots$$
$$x_k = x_k^1 e_1 + \ldots + x_k^n e_n.$$

Базисные формы $e^1, \ldots, e^n$, как и раньше, возьмем так, что $e^i(x)$ равна i -й координате аргумента x.

2. Пусть $i_1 < i_2 < \ldots < i_k$; тогда

$$(e^{i_1} \wedge \ldots \wedge e^{i_k})(x_1, \ldots, x_k) = \\ = \sum \delta^{i_1 \ldots i_k}_{\alpha_1 \ldots \alpha_k} e^{\alpha_1}(x_1) \ldots e^{\alpha_k}(x_k) = \\ = \sum \delta^{i_1 \ldots i_k}_{\alpha_1 \ldots \alpha_k} x_1^{\alpha_1} \ldots x_k^{\alpha_k} = V^{i_1 \ldots i_k},$$

где $V^{i_1,\ldots,i_k}$ — минор k-го порядка матрицы X, составленной из координат векторов $x_1, \ldots, x_k$:

$$X = \begin{Vmatrix} x_1^1 & x_1^2 & \ldots & x_1^n \\ \cdot & \cdot & \cdot & \cdot \\ x_k^1 & x_k^2 & \ldots & x_k^n \end{Vmatrix};$$

минор $V^{i_1,\ldots,i_k}$ определен столбцами, номера которых суть $i_1, \ldots, i_k$.

Замечание. Если пространство евклидово и $e_1, \ldots, e_n$ — ортонормированный базис, то число $V^{i_1,\ldots,i_k}$ есть k-мерный ориентированный об′єм k-мерного параллелепипеда, который построен на проекциях векторов $x_1, \ldots, x_k$ на координатную k-мерную плоскость базисных векторов $e_{i_1,\ldots,} e_{i_k}$.

7.1.12. Координатное выражение внешней формы

1. Из п. 7.1.11 и из формулы (45) следует координатная запись значения внешней формы:

$$\omega(x_1, \ldots, x_k) = * \sum \omega_{i_1 \ldots i_k} V^{i_1 \ldots i_k}. \tag{47}$$

В более подробном виде

$$\omega(x_1, \ldots, x_k) = * \sum \omega_{i_1 \ldots i_k} \begin{vmatrix} x_1^{i_1} & \ldots & x_1^{i_k} \\ \cdot & \cdot & \cdot \\ x_k^{i_1} & \ldots & x_k^{i_k} \end{vmatrix}.$$

7.1.13. Специальные обозначения

1. Вернемся к разложению внешней формы согласно (45)

$$\omega = * \sum \omega_{i_1 \dots i_k} e^{i_1} \wedge \dots \wedge e^{i_k}. \tag{48}$$

Так как $e^i(x)=x^i$, то символ x^i мы будем рассматривать как обозначение формы e^i. Ввиду этого наряду с формулой (48) будем употреблять также формулу

$$\omega = * \sum \omega_{i_1 \dots i_k} x^{i_1} \wedge \dots \wedge x^{i_k}. \tag{49}$$

Формулы (48) и (49) отличаются только способом записи. Во избежание недоразумений еще раз подчеркнем, что в равенстве (49) $x^{i_1}, \dots, x^{i_k}$ следует понимать, как обозначение линейных форм, которые могут быть взяты от разных векторных аргументов (каждая от своего). Например,

$$(x^1 \wedge x^2)(x_1, x_2) = (e^1 \wedge e^2)(x_1, x_2) =$$
$$= e^1(x_1)e^2(x_2) - e^2(x_1)e^1(x_2) = x_1^1 x_2^2 - x_1^2 x_2^1 = \begin{vmatrix} x_1^1 & x_1^2 \\ x_2^1 & x_2^2 \end{vmatrix};$$

здесь x^1, x^2 — формы, x_1, x_2 — векторы, x^i_j — их координаты.

2. Отметим еще равенство, которое соответствует разложению определителя $V^{i_1,\dots,i_k}$ по элементам последней строки:

$$(x^{i_1} \wedge x^{i_2} \wedge \dots \wedge x^{i_{k-1}} \wedge x^{i_k})(x_1, x_2, \dots, x_{k-1}, x_k) =$$
$$= (-1)^{k-1}\{(x^{i_2} \wedge x^{i_3} \wedge \dots \wedge x^{i_k})\, x_k^{i_1} -$$
$$- (x^{i_1} \wedge x^{i_3} \wedge \dots \wedge x^{i_k})\, x_k^{i_2} + \dots$$
$$\dots + (-1)^{k-1}(x^{i_1} \wedge x^{i_2} \wedge \dots \wedge x^{i_{k-1}})\, x_k^{i_k}\}(x_1, \dots, x_{k-1}); \tag{50}$$

в правой части написаны внешние произведения линейных форм, которые в каждом слагаемом берутся последовательно от векторных аргументов $x_1 \dots, x_{k-1}$; коэффициенты

$$x_k^{i_1}, \dots, x_k^{i_k}$$

суть числа (координаты вектора x_k).

3. Запись внешней формы в виде (49) особенно удобна при переходе к новым координатам.

7.1.14. Преобразование внешней формы при переходе к новым координатам

1. Пусть выполняется переход к новым координатам:

$$\left.\begin{aligned} x^1 &= P^1_{1'}x^{1'} + \ldots + P^1_{n'}x^{n'}, \\ &\cdots\cdots\cdots\cdots \\ x^n &= P^n_{1'}x^{1'} + \ldots + P^n_{n'}x^{n'}; \end{aligned}\right\} \tag{51}$$

здесь $x^1, ..., x^n$— координаты произвольного вектора x в старом базисе, $x^{1'}, ..., x^{n'}$ — координаты этого вектора x в новом базисе. Числа P^i_j суть коэффициенты преобразования. Однако согласно предыдущему подразделу можно считать, что написанные в этих равенствах слева символы суть линейные формы $e^1, ..., e^n$, взаимные со старым базисом (т.е. которые удовлетворяют условиям $e^i(x) = x^1$, где x^i есть i-я координата вектора x). Аналогичным образом по отношению к новому базису можно рассматривать члены правых частей. При такой точке зрения на формулы (51) приведение каждого члена $x^{i_1} \wedge \ldots \wedge x^{i_k}$ внешней формы (49) к новым координатам можно выполнить путем почленного внешнего умножения правых частей формул (51), беря из них последовательно те, которые имеют номера $i_1, \ldots, i_k$. Выполняя указанное действие, получим

$$x^{i_1} \wedge \ldots \wedge x^{i_k} = * \sum D^{i_1 \ldots i_k}_{j'_1 \ldots j'_k} x^{j'_1} \wedge \ldots \wedge x^{j'_k},$$

где

$$D^{i_1 \ldots i_k}_{j'_1 \ldots j'_k} = \begin{vmatrix} P^{i_1}_{j'_1} & \ldots & P^{i_1}_{j'_k} \\ \cdot & \cdot \,\cdot\,\cdot & \cdot \\ P^{i_k}_{j'_1} & \ldots & P^{i_k}_{j'_k} \end{vmatrix}.$$

В конкретных случаях вместо применения этой готовой формулы удобнее непосредственно проводить действия по правилам внешней алгебры. Например, если

$$x = \alpha u + \beta v,$$
$$y = \gamma u + \delta v,$$

то

$$\begin{aligned} x \wedge y &= (\alpha u + \beta v) \wedge (\gamma u + \delta v) = \\ &= \alpha\gamma (u \wedge u) + \alpha\delta (u \wedge v) + \beta\gamma (v \wedge u) + \beta\delta (v \wedge v) = \\ &= (\alpha\delta - \beta\gamma)(u \wedge v). \end{aligned}$$

7.2. Внешнее дифференцирование

7.2.1. Касательные пространства

1. Пусть E — евклидово n-мерное пространство, т.е. линейное n-мерное пространство, в котором задано положительно определенное скалярное произведение.

Зафиксируем какой-нибудь элемент $x\in E$ будем рассматривать всевозможные пары (x, u), где u - произвольный элемент из E. В множестве этих пар введем линейные операции:

$$(x, u_1) + (x, u_2) = (x, u_1 + u_2), \qquad (52)$$

$$\alpha(x, u) = (x, \alpha u). \qquad (53)$$

Полученное тем самым линейное пространство пар (x, u) обозначим через T_x. Первый (зафиксированный) элемент x пары (x, u) мы будем называть *точкой.* Пару (x, u), как элемент T_x, будем называть *вектором,* приложенным к точке x. Впрочем, мы разрешим себе говорить также, что вектор u приложен к точке x, если он является вторым элементом пары (x, u).

Линейное пространство T_x назовем *касательным пространством* к E в точке x. Вследствие (52) и (53) биективное отображение E на T_x, при котором произвольному вектору $u\in E$ отвечает пара (x, u), является линейным изоморфизмом. С помощью этого изоморфизма все свойства E, как линейного пространства, переносятся на каждое касательное пространство T_x. Больше того, по тому же изоморфизму мы будем переносить из E в T_x и евклидовы свойства E; впрочем, достаточно сказать, что *скалярным произведением* двух элементов (x, u) и (x, v) пространства T_x мы назовем число, равное скалярному произведению элементов $u, v \in E$.

2. Все, что было сейчас определено формально, можно высказать наглядным образом. А именно: 1) любой элемент $x\in E$ можно зафиксировать и назвать *точкой;* 2) любой элемент $u\in E$ можно сопоставить с точкой x и назвать *вектором, приложенным к точке* x. Можно также сказать, что точка x является *началом* приложенного к ней *вектора* u. Тогда точку $y= x+u$ следует считать его *концом.* 3) Линейное пространство T_x векторов, приложенных к точке x, называется *касательным пространством к* E *в точке* x.

3. Согласно сказанному евклидово пространство E можно рассматривать как пространство точек; онявляется метрическим пространством с метрикой

$$\rho(x, y) =\|y-x\|= \sqrt{(y-x,\, y-x)}\ .$$

Соответственно в E определены все понятия, которые относятся к метрическим пространствам, в частности, в E определено понятие открытого множества и области.

4. Пусть (приложенные) векторы $(x, u) \in T_x$ и $(y, u) \in T_y$ отвечают одному и тому же (свободному) вектору $u \in E$. Тогда говорят, что один из них (любой) получен *параллельным перенесением* другого; например, вектор (y, u) получен параллельным перенесением вектора (x, u) из касательного пространства T_x в касательное пространство Ty.

Очевидно, что приложенные векторы, получаемые во всех касательных пространствах параллельным перенесением какого-нибудь одного вектора, удовлетворяют обычным условиям отношения эквивалентности. Поэтому их часто называют *равными* векторами (хотя они принадлежат разным пространствам). Имея в виду возможность однозначно строить равные векторы в разных касательных пространствах, говорят, что в евклидовом пространстве E имеет место *абсолютный параллелизм векторов.*

5. Вместе с линейным касательным пространством T_x определено спряженное ему пространство T^*_x; оно называется *кокасательным пространством* к евклидовому просторную E в точке x. Кроме того, определено при любом k линейное пространство

$$\overset{\otimes}{T}{}^k_x = T^*_x \otimes \ldots \otimes T^*_x$$

(справа — k сомножителей). В каждом $\overset{\otimes}{T}{}^k_x$ определено подпространство Λ^k_x, аналогично тому, как в п.7. 1.10 определено подпространство Λ^k пространства T^k.

6. Заметим, наконец, что наличие евклидовой структуры в E нами, в сущности, не использовано. Все сказанное относится и к аффинному пространству. Мы предполагаем пространство E евклидовым только с той целью, чтобы в дальнейшем мы могли в нужных случаях применять евклидовы понятия без специальных предостережений.

7.2.2. Внешние дифференциальные формы

1. Обозначим через U некоторую область в евклидовом пространстве E. Пусть для любой точки $x \in U$ в касательном пространстве T_x заданаполилинейная форма a.

Тогда говорят, что дана форма в области U. Иначе можно сказать, что с каждой точкой x и с каждым упорядоченным набором векторов $\xi_1, \ldots, \xi_k \in T_x$ сопоставлено число из R, или, что дано отображение

$$a:\ T^k_x \to \mathbb{R}, \tag{54}$$

где T^k_x есть k-кратное произведение T_x на себя. Элементами T_x можно считать упорядоченные наборы $(x, \xi_1, ..., \xi_k)$. Отображение (54) предполагается линейным по каждому векторному аргументу, т.е. по каждой составлющей набора $(x, , \xi_1, ..., \xi_k)$, не считая точки x.

Форму (54) называют *дифференциальной формой* в области U. В соответствии с п. 7.1.3 можно написать

$$a \in \overset{\otimes}{T}{}^k_x.$$

2. Если форма (54) совпадает со своей альтернацией, то ее называют *внешней дифференциальной k-формой* в области U. Для обозначения внешних дифференциальных форм наиболее употребляема буква ω.

Соответственно можно написать $\omega \in \Lambda^k_x$.

3. К дифференциальным формам применимые все алгебраические понятия и методы, о которых говорилось в п. 7.1. Достаточно считать, что точка x зафиксирована, и рассматривать как пространства L касательное пространство T_x. В частности, для внешней формы можно дать координатное представление.

4. Пусть в пространстве E задана координатная система с базисом $e_1, ..., e_n$, пусть $e^1, ..., e^n$ - линейные формы, которые составляют взаимный с ним базис в сопряженном пространстве E^*. Будем считать, что одновременно с этим в каждом касательном пространстве T_x задан базис, который состоит из векторов $e_1, ..., e_n$, приложенных к точке x (для этих приложенных векторов мы сохраним обозначение символами $e_1, ..., e_n$). Формы $e^1, ..., e^n$ будем рассматривать как элементы сопряженного пространства T_x^*, для чего достаточно полагать, что их аргументы приложены к точке x. Эти постоянные, точнее говоря, не зависящие от x формы, мы будем называть *координатными* (для данной координатной системы) или *проектующими,* поскольку по условию взаимности

$$e^j(\xi) = \xi^j, \tag{55}$$

что есть проекция вектора ξ на координатную ось с номером j. Координатные формы принято обозначать через dx^j. Таким образом, наряду с записью (55) можно писать

$$dx^j(\xi) = \xi^j. \tag{56}$$

5. Согласно п. 7.1.10 имеет место координатная запись внешней дифференциальной формы $\omega \in \Lambda^k_x$:

$$\omega = * \sum \omega_{i_1 \dots i_k}(x)\, e^{i_1} \wedge \dots \wedge e^{i_k}, \tag{57}$$

или

$$\omega = * \sum \omega_{i_1 \dots i_k}(x)\, dx^{i_1} \wedge \dots \wedge dx^{i_k}. \tag{58}$$

Формулы (57) и (58) отличаются только способом записи (подобно формулам (45) и (51). Обе эти формулы определяют ω как внешнюю форму в касательном пространстве T_x. Коэффициенты $\omega_{i_1 \dots i_k}$ (x) суть числовые функции точки x (кроме того, они зависят от выбора координатной системы).

Обычную функцию точки $(f\colon U \to \mathrm{R})$ считают формой нулевой степени. Для формы нулевой степени равенство (58) имеет вид $\omega = \omega(x)$.

6. Рассмотрим примеры. Пусть E — трехмерное евклидово пространство, и пусть в нем дано поле векторов: $p = p(x)$, т.е. в каждой точке $x \in E$ прилжен вектор $p(x)$. Тогда в E определено две внешних дифференциальных формы. Одна из них - линейная форма, есть скалярное произведение: $\omega^1 = (p(x), \xi)$. Если $p = p(x)$ толковать как силу, приложенную в точке x, то ω^1 означает элементарную работу, которую выполняет сила p при перемещении точки приложения из x в конец вектора $x+\xi$. Другая - внешняя дифференциальная форма второй степени, есть смешанное произведение

$$\omega^2 = (p(x),\ \xi_1,\ \xi_2).$$

Форме ω^2 часто дают гидродинамическое истолкование, полагая, что ω^2 есть поток вектора $p(x)$ через элементарную плоскость $(\xi_1,\ \xi_2)$. Отметим, наконец, внешнюю дифференциальную форму третьей степени

$$\omega^3 = (\xi_1,\ \xi_2,\ \xi_3).$$

Эта форма есть ориентированный объем параллелепипеда, построенного на векторах $\xi_1, \xi_2, \xi_3 \in T_x$.

Нетрудно написать координатные представления форм ω^1, ω^2, ω^3. А именно, если $p(x) = \{P(x),\ Q(x),\ R\{x)\}$, то

$$\omega^1 = P\,dx^1 + Q\,dx^2 + R\,dx^3,$$
$$\omega^2 = R\,dx^1 \wedge dx^2 + (-Q)\,dx^1 \wedge dx^3 + P\,dx^2 \wedge dx^3,$$
$$\omega^3 = dx^1 \wedge dx^2 \wedge dx^3.$$

Первое из этих равенств усматривается с очевидностью:

$$\omega^1(\xi) = P\xi^1 + Q\xi^2 + R\xi^3 = (P\,dx^1 + Q\,dx^2 + R\,dx^3)(\xi).$$

Чтобы понять второе и третье, довольно написать общеизвестные выражения смешанного произведения:

$$\omega^2(\xi_1, \xi_2) = \begin{vmatrix} P & Q & R \\ \xi_1^1 & \xi_1^2 & \xi_1^3 \\ \xi_2^1 & \xi_2^2 & \xi_2^3 \end{vmatrix} =$$

$$= R \begin{vmatrix} \xi_1^1 & \xi_1^2 \\ \xi_2^1 & \xi_2^2 \end{vmatrix} + (-Q) \begin{vmatrix} \xi_1^1 & \xi_1^3 \\ \xi_2^1 & \xi_2^3 \end{vmatrix} + P \begin{vmatrix} \xi_1^2 & \xi_1^3 \\ \xi_2^2 & \xi_2^3 \end{vmatrix},$$

$$\omega^3(\xi_1, \xi_2, \xi_3) = \begin{vmatrix} \xi_1^1 & \xi_1^2 & \xi_1^3 \\ \xi_2^1 & \xi_2^2 & \xi_2^3 \\ \xi_3^1 & \xi_3^2 & \xi_3^3 \end{vmatrix}.$$

Отсюда получается предыдущая запись ω^2, ω^3 в символике внешних форм, если использовать п. 7.1.13.

7.2.3. Внешний дифференциал

1. *Внешний дифференциал* формы нулевой степени, т.е. функции, по определению есть ее обычный дифференциал.

Пусть в области $U \subset E$ задана (действительная) дифференцируемая функция

$$f\colon U \to \mathbb{R}. \tag{59}$$

Тогда в каждой точке $x \in U$ существует дифференциал функции

$$dy = f'(x)\xi. \tag{60}$$

Здесь $f'(x)$ — линейная форма, которая определена в касательном пространстве T_x; dy — ее значение на векторе $\xi \in T_x$. Сама форма $f'(x)$ есть производная данной функции в точке x; ее обозначают также через $Df(x)$.

Мы имеем в виду сейчас функции с числовыми значениями. Согласно этому, считая действительную ось R линейным пространством, вместо (60) можно написать

$$f'(x)\colon T_x \to \mathbb{R},$$

имея в виду, что это отображение линейное.

2. Напомним определение производной. Будем считать, что данная функция f определена в области: $U(f\colon U \to R)$. Пусть A — некоторое линейное отображение $T_x \to \mathrm{R}$ (точка x зафиксирована при условии $x \in U$). Обозначим через z произвольную (переменную) точку области U. Имеем $z - x \in T_x$; следовательно, определен образ $A(z - x)$ вектора $z - x$ при отображении A. Данная функция f называется

дифференцируемой в точке x, если существует такое линейное отображение $A\colon T_x \to R$, что

$$f(z)-f(x) = A(z-x) + o(z-x); \qquad (61)$$

здесь «*o* малое» от $z - x$ понимается как обычн: оно удовлетворяет неравенству

$$\| o(z-x) \| \leqslant \psi(z, x) \| z-x \|, \qquad (62)$$

где $\psi(z, x)$ — некоторая неотрицательная функция, которая стремится к нулю при $z \to x$. В данном случае $o(z - x)$ есть функция с числовыми значениями; ее норма $\| o(z - x) \|$ есть просто модуль $|o(z - x)|$.

Линейное отображение $A\colon T_x \to R$, которое удовлетворяет условию (61), называется *производной* функцией f в точке x символически $A=Df(x)$ или $A= f'(x)$.

3. Поскольку линейные отображения в R мы называем формами (с числовыми значениями), то *производная функции* в данной точке есть *форма*.

Очевидно, $f'(x) \in \Lambda^1_x$, где $\Lambda^1_x = T_x$ есть кокасательное пространство в точке x. Соответственно можно сказать, что *производная* есть *ковектор*.

Значение производной на произвольном векторе $\xi \in T_x$ называется *дифференциалом функции*: $dy=Df(x)\xi$, или $df= Df(x)\xi$; можно писать также $dy=f'(x)\xi$ или $df=f'(x)\xi$. Впрочем, дифференциалом часто называют производную (что не верно).

4. Как видно из определений, данных в п. 2, производная и дифференциал инвариантны, т.е. не зависят от выбора координат. Однако их координатные представления от выбора координатной системы зависят.

Пусть в E введена система координат с базисом $e_1, \ldots, e_n$. Тогда в каждом касательном пространстве T_x будет введен базис, который состоит из векторов $e_1, \ldots, e_n$, приложенных к точке x. Тем самым произвольный вектор $\xi \in T_x$ получает координатное представление

$$\xi = \{\xi^1, \ldots, \xi^n\}.$$

Одновременно получает координатное представление дифференциал функции

$$df = f_{x^1}\xi^1 + \ldots + f_{x^n}\xi^n. \qquad (63)$$

и производная (как ковектор из T^*_x)

$$f'(x) = \{f_{x^1}(x), \ldots, f_{x^n}(x)\}. \qquad (64)$$

Здесь $f_{x^1}(x), \ldots, f_{x^n}(x)$ — частные производные в точке x. Они зависят от выбора системы координат, что ясно по их определению и усматривается также из очевидной формулы

$$f_{x^i}(x) = f'(x)\, e_i. \qquad (65)$$

или

$$f_{x^i}(x) = Df(x)\, e_i.$$

Последняя формула дает основание писать вместо f_{x^i} символ $D_i f$ или просто D_i.

Замечание. Заметим, что правая часть (63) есть свертка элемента ξ касательного пространства T_x с элементом $f'(x)$ сопряженного пространства T^*_x.

Замечание. Уже сейчас можно понять, что рассматривать и различать касательные пространства имеет смысл, хотя, казалось бы, они не отличаются от исходного пространства E. В самом деле, например, производная $f'(x)$ по нашему определению есть линейное отображение $T_x \to$ R. Можно было бы определить $f'(x)$ как линейное отображение $E \to R$, но все равно тогда мы должны были бы указать, что оно зависит и как, именно, зависит от точки x. А это, в сущности, и означает, что мы имеем дело с парами (x, u), $u \in E$ или, наглядно говоря, с приложенными векторами.

5. Для данной системы декартовых прямоугольных координат зададим систему функций $\pi^1(x), \ldots, \pi^n(x)$ от (переменной) точки $x \in E$ условиями $\pi^i(x) = x^i$, где $x^1, \ldots, x^n$ — координаты точки x. По очевидным причинам эти функции называются *проектирующими*. Легко понять, что

$$d\pi^i(x) = D\pi^i(x)\,\xi = \xi^i = dx^i(\xi) \qquad (66)$$

(см. 7.2.2 п. 2). Таким образом, производные проектирующих функций суть проектирующие формы: $D\pi^i(x) = dx^i$.

Обратим внимание, что формулу (63) можно написать в виде

$$df = (D_1\, dx^1 + \ldots + D_n\, dx^n)(\xi),$$

где справа $D_1, \ldots, D_n$ можно рассматривать как коэффициенты разложения формы Df по формам $dx^1 \ldots, dx^n$. Можно написать также

$$df = D_1\, d\pi^1(x) + \ldots + D_n\, d\pi^n(x),$$

поскольку $d\pi^i(x) = dx^i(\xi)$. Пишут очень часто еще

$$df = D_1 dx^1 + \ldots + D_n dx^n.$$

Однако, поскольку $dx^1 \ldots, dx^n$ означают формы, а не значение форм, то последняя запись на самом деле выражает не дифференциал функции, а ее производную (употребление символа d для обозначения производной (вместо дифференциала) вошло в традицию.)

Соответственно этой записи имеем вместо (66)

$$d\pi^i(x) = dx^i = e^i. \tag{67}$$

6. Теперь мы определим дифференциал полилинейной формы; тем самым, в частности, будут определены высшие дифференциалы функции, так как ее первый дифференциал есть линейная форма.

Пусть в некоторой области U дана дифференциальная полилинейная форма с числовыми значениями

$$a = a(x, \xi_1, \ldots, \xi_k), \tag{68}$$

где

$$x \in U, \ \xi_1, \ldots, \xi_k \in T_x.$$

Иначе говоря, для любого $x \in U$ дано полилинейное отображение

$$a\colon T_x^k \to \mathbb{R}.$$

Зафиксируем $\xi_{1,\ldots}$, ξ_k в области U. Это значит, что мы выбираем как-нибудь $\xi_{1,\ldots}$, ξ_k в некоторой точке $x \in U$, а в других точках области U берем в качестве $\xi_{1,\ldots}$, ξ_k векторы, которые получаются параллельным перенесением вібранных (см. п. 4 п. 7.2.1). Если при любых фиксированных $\xi_{1,\ldots}$, ξ_k форма a становится дифференцируемой функцией в области U (т.е. дифференцируемой функцией переменной точки $x \in U$), то говорят, что эта форма дифференцируема в области U. Предполагая форму a дифференцируемой, рассмотрим дифференциал функции, которая получается из формы a путем фиксирования $\xi_{1,\ldots}$, ξ_k:

$$da = Da(x, \xi_1, \ldots, \xi_k)\eta, \qquad \eta \in T_x. \tag{69}$$

Позволим теперь векторам η, $\xi_{1,\ldots}$, ξ_k принимать любые значения; тогда правая часть (69) будет полилинейной формой от векторов η, $\xi_{1,\ldots}$, ξ_k, число которых — k+1. Эта форма называется *дифференциалом* данной формы a.

7. Пусть теперь дана *внешняя* дифференциальная форма

$$\omega = \omega(x, \xi_1, \ldots, \xi_k). \tag{70}$$

Ее дифференциал будет полилинейной формой, которая обладает косой симметрией по аргументам $\xi_{1,\ldots}$, ξ_k, являющимся аргументами самой формы ω, т.е. без участия η.

8. *Внешним дифференциалом k-формы,* ω называется альтернация ее дифференциала, который определен в п. 6, взятая с коэффициентом k +1:

$$d\omega = (k+1)[D\omega]. \tag{71}$$

Вследствие этой альтернации форма $d\omega$ обладает косой симметрией по всем своим векторным аргументам вместе с η.

Если форма задана своим координатным представлением

$$\omega = * \sum \omega_{i_1 \dots i_k}(x)\, dx^{i_1} \wedge \dots \wedge dx^{i_k}, \tag{72}$$

то из (71), (72) и из определения внешнего произведения двух форм получаем

$$d\omega = * \sum d\omega_{i_1 \dots i_k}(x) \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k}. \tag{73}$$

Здесь

$$d\omega_{i_1 \dots i_k}(x) = \sum_{\alpha=1}^{n} D_\alpha \omega_{i_1 \dots i_k}(x)\, dx^\alpha; \tag{74}$$

$D_\alpha \omega_{i_1 \dots i_k}(x)$ — частные производные.

9. Следует заметить, что согласно принятой символике dx^α, $dx^{i_1}, \dots, dx^{i_k}$ суть формы, а не значения форм. Поэтому в силу равенства (73) $d\omega$ точнее следовало бы называть внешней производной, а не внешним дифференциалом. Производной является также $d\omega_{i_1 \dots i_k}(x)$. Внешний дифференциал мы получим, если подсчитаем $d\omega$ на произвольном наборе векторов η, $\xi_1, \dots, \xi_k \in T_x$. При этом $d\omega_{i_1 \dots i_k}(x)$ подсчитывается на векторе η. Однако мы сохраним традиционную терминологию.

10. Из определения внешнего дифференциала согласно п. 8 непосредственно следует, что внешний дифференциал в евклидовом пространстве определен нами инвариантно (поскольку в его определении вообще координатная система не принимает участие). Координатное представление $d\omega$ от выбора системы координат, разумеется, зависит.

7.2.4. Основные свойства внешнего дифференциала

1. Из определения внешнего дифференциала или из его координатного представления (69) непосредственно следует, что операция внешнего дифференцирования является линейной. Именно, если ω_1, ω_2 — любые дифференциальные k-формы, дифференцируемые в области U, и α, β— любые числа, то k-форма $\alpha\omega_1+\beta\omega_2$ также дифференцируема в U, причем

$$d(\alpha\omega_1 + \beta\omega_2) = \alpha\, d\omega_1 + \beta\, d\omega_2. \tag{75}$$

Доказательство этого утверждения только что написанного тождества мы предоставим читателю.

2. Даее, имеет место следующее свойство: если ω, σ — любые внешние дифференциальные формы, дифференцируемые в области U,

то внешнее произведение $\omega \wedge \sigma$ есть внешняя дифференциальная форма, которая также дифференцируема в области U, причем

$$d(\omega \wedge \sigma) = d\omega \wedge \sigma + (-1)^k \omega \wedge d\sigma, \tag{76}$$

где k — степень формы ω.

Доказательство формулы (76) проведем при помощи какого-нибудь координатного представления (тем самым (76) будет доказано вообще, поскольку сами формы и их внешние дифференциалы инвариантны). Имеем:

$$\omega = {*}\sum \omega_{i_1 \ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k},$$
$$\sigma = {*}\sum \sigma_{j_1 \ldots j_l} dx^{j_1} \wedge \ldots \wedge dx^{j_l},$$

Отсюда

$$\begin{aligned} d(\omega \wedge \sigma) &= d\big({*}\sum \omega_{i_1 \ldots i_k} \sigma_{j_1 \ldots j_l} dx^{i_1} \wedge \ldots \\ &\qquad \ldots \wedge dx^{i_k} \wedge dx^{j_1} \wedge \ldots \wedge dx^{j_l}\big) = \\ &= {*}\sum \big((d\omega_{i_1 \ldots i_k}) \sigma_{j_1 \ldots j_l} + \omega_{i_1 \ldots i_k} d\sigma_{j_1 \ldots j_l}\big) \wedge \\ &\qquad \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k} \wedge dx^{j_1} \wedge \ldots < dx^{j_l} = \\ &= {*}\sum \big(d\omega_{i_1 \ldots i_k} \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k}\big) \wedge \\ &\qquad \wedge \big({*}\sum \sigma_{j_1 \ldots j_l} dx^{j_1} \wedge \ldots \wedge dx^{j_l}\big) + \\ &\quad + (-1)^k \big({*}\sum \omega_{i_1 \ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}\big) \wedge \\ &\qquad \wedge \big({*}\sum d\sigma_{j_1 \ldots j_l} \wedge dx^{j_1} \wedge \ldots \wedge dx^{j_l}\big). \end{aligned} \tag{77}$$

В правой части равенства (77) перед вторым членом пришлось поставить $(—1)^k$, поскольку линейную форму $d\sigma_{j_1 \ldots j_l}$ нужно было переставить с k-формой $dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ (см. 7.1.8). Вместе с равенством (77) доказано и равенство (76).

3. **Л е м м а**. *Если f — дважды дифференцируемая в некоторой области U функция (рассматриваемая как форма нулевой степени), то внешний дифференциал $d(df)=0$.*

Доказательство. По определению внешних дифференциалов имеем

$$df = \sum \frac{\partial f}{\partial x^i} dx^i,$$

$$d(df) = \sum d\left(\frac{\partial f}{\partial x^i}\right) \wedge dx^i =$$

$$= \sum \frac{\partial^2 f}{\partial x^i \partial x^j} dx^j \wedge dx^i = 0.$$

Эта сумма равна нулю потому, что вторая частная производная не меняется при перестановке индексов i, j, а внешнее произведение $dx^i \wedge dx^j$ при такой перестановке меняет знак.

4. **Теорема**. *Какой бы не была дважды дифференцируемая в области U внешняя форма* ω, *всегда*

$$d^2\omega = 0. \tag{78}$$

Здесь $d^2\omega = d(d\omega)$.

Доказательство. Воспользуемся координатным представлением формы в в какой-нибудь системе координат

$$\omega = * \sum \omega_{i_1 \ldots i_k}(x)\, e^{i_1} \wedge \ldots \wedge e^{i_k},$$

где $e^1, \ldots, e^n$ — проектирующие формы, соответствующие выбранной координатной системе. Согласно пп. 5 и 7 п. 7.2.3

$$e^i = d\pi^i(x), \tag{79}$$

где $\pi^i(x)$ — проектирующие функции. Из (79) и вследствие предыдущей леммы

$$de^i = d(\pi^i(x)) = 0. \tag{80}$$

С другой стороны, по той же лемме

$$d\left(d\omega_{i_1 \ldots i_k}(x)\right) = 0. \tag{81}$$

Применяя (многократно) правила дифференцирования, которые даны в пп. 1 и 2, и пользуясь равенствами (80), (81), получим $d^2\omega = 0$.

5. Сама формулировка доказанной теоремы показывает ее фундаментальное значение для внешнего дифференцирования (хотя бы уже потому, что в силу этой теоремы все дифференциальное исчисление внешних форм сводится только к первым дифференциалам). Дальше мы покажем роль этой теоремы в конкретных и принципиальных вопросах.

7.2.5. Примеры внешнего дифференцирования

1. Рассмотрим сначала линейную дифференциальную форму на евклидовой плоскости

$$\omega = P\, dx^1 + Q\, dx^2;$$

здесь P и Q — функции точки x. Имеем

$$d\omega = dP \wedge dx^1 + dQ \wedge dx^2 =$$
$$= \left(\frac{\partial P}{\partial x^1} dx^1 + \frac{\partial P}{\partial x^2} dx^2\right) \wedge dx^1 + \left(\frac{\partial Q}{\partial x^1} dx^1 + \frac{\partial Q}{\partial x^2} dx^2\right) \wedge dx^2 =$$
$$= \left(\frac{\partial Q}{\partial x^1} - \frac{\partial P}{\partial x^2}\right) dx^1 \wedge dx^2.$$

В связи с этими выражениями ω, $d\omega$ полезно вспомнить известную в элементарном анализе формулу Грина.

2. Теперь рассмотрим линейную дифференциальную форму в пространстве

$$\omega = Pdx^1 + Q\,dx^2 + R\,dx^3; \qquad (82)$$

здесь P, Q, R — также функции точки x. Можно считать, что эти функции определяют вектор

$$p(x) = \{P(x), Q(x), R(x)\}.$$

Имеем:

$$d\omega = dP \wedge dx^1 + dQ \wedge dx^2 + dR \wedge dx^3 =$$
$$= \left(\frac{\partial P}{\partial x^1} dx^1 + \frac{\partial P}{\partial x^2} dx^2 + \frac{\partial P}{\partial x^3} dx^3\right) \wedge dx^1 +$$
$$+ \left(\frac{\partial Q}{\partial x^1} dx^1 + \frac{\partial Q}{\partial x^2} dx^2 + \frac{\partial Q}{\partial x^3} dx^3\right) \wedge dx^2 +$$
$$+ \left(\frac{\partial R}{\partial x^1} dx^1 + \frac{\partial R}{\partial x^2} dx^2 + \frac{\partial R}{\partial x^3} dx^3\right) \wedge dx^3 =$$
$$= \left(\frac{\partial Q}{\partial x^1} - \frac{\partial P}{\partial x^2}\right) dx^1 \wedge dx^2 + \left(\frac{\partial R}{\partial x^1} - \frac{\partial P}{\partial x^3}\right) dx^1 \wedge dx^3 +$$
$$+ \left(\frac{\partial R}{\partial x^2} - \frac{\partial Q}{\partial x^3}\right) dx^2 \wedge dx^3.$$

Введем в рассмотрение ротацию вектора $p(x)$:

$$\operatorname{rot} p = \left\{ \frac{\partial R}{\partial x^2} - \frac{\partial Q}{\partial x^3};\ \frac{\partial P}{\partial x^3} - \frac{\partial R}{\partial x^1};\ \frac{\partial Q}{\partial x^1} - \frac{\partial P}{\partial x^2} \right\}.$$

Тогда

$$d\omega(\xi_1, \xi_2) = (\operatorname{rot} p, \xi_1, \xi_2), \qquad (83)$$

т.е. $d\omega(\xi_1, \xi_2)$ представляет собой поток ротации через элементарную площадку (ξ_1, ξ_2) (см. п. 6 п. 7.2.2). В связи с выражениями (82) и (83), полезно вспомнить известную в элементарном анализе формулу Стокса.

3. Пусть теперь ω - внешняя дифференциальная форма второй степени

$$\omega = Rdx^1 \wedge dx^2 - Q\,dx^1 \wedge dx^3 + P\,dx^2 \wedge dx^3. \qquad (84)$$

Согласно п. 6 п. 7.2.2 $\omega(\xi_1, \xi_2)$ есть поток вектора $p(x)$ через элементарную площадку (ξ_1, ξ_2). Имеем:

$$d\omega = \left(\frac{\partial R}{\partial x^1}dx^1 + \frac{\partial R}{\partial x^2}dx^2 + \frac{\partial R}{\partial x^3}dx^3\right)\wedge dx^1 \wedge dx^2 -$$
$$-\left(\frac{\partial Q}{\partial x^1}dx^1 + \frac{\partial Q}{\partial x^2}dx^2 + \frac{\partial Q}{\partial x^3}dx^3\right)\wedge dx^1 \wedge dx^3 +$$
$$+\left(\frac{\partial P}{\partial x^1}dx^1 + \frac{\partial P}{\partial x^2}dx^2 + \frac{\partial P}{\partial x^3}dx^3\right)\wedge dx^2 \wedge dx^3 =$$
$$= \left(\frac{\partial P}{\partial x^1} + \frac{\partial Q}{\partial x^2} + \frac{\partial R}{\partial x^3}\right)dx^1 \wedge dx^2 \wedge dx^3.$$

Введем в рассмотрение дивергенцию поля *p(x):*

$$\operatorname{div} p = \frac{\partial P}{\partial x^1} + \frac{\partial Q}{\partial x^2} + \frac{\partial R}{\partial x^3}.$$

Тогда

$$d\omega = \operatorname{div} p\, dx^1 \wedge dx^2 \wedge dx^3. \qquad (85)$$

Форма $\omega(\xi_1, \xi_2, \xi_3)$ есть дивергенция поля *p(x)* в элементарном объеме (ξ_1, ξ_2, ξ_3). В связи с выражениями (84) и (85) полезно вспомнить известную формулу Гаусса - Остроградского.

4. В разделе 7.3 будет установлена некоторая общая интегральная формула, включающая как свои частные случаи все три интегральные формулы, которые упоминались нами в конце каждого из пп. 1, 2 и 3. Ее называют общей *формулой Стокса.*

7.2.6. Индуцированное отображение пространства внешних форм

1. Пусть задано линейное пространство L_1 и некоторое линейное отображение его в линейное пространство L_2:

$$\psi: L_1 \to L_2. \qquad (86)$$

Вместо (86) можно написать: $v=\psi(u)$, где $u\in L_1$, v — образ элемента u в пространстве $L_2(v\in L_2)$. Обозначим через $T^k(L_1)$ и $T^k(L_2)$ линейные пространства k-тензоров, определенных соответственно в L_1 и L_2. Как и раньше, будем представлять себе k-тензоры как полилинейные формы.

Для дальнейшего важно, что отображение (86) индуцирует по определенному стандарту некоторое линейное отображение $T^k(L_2)$ в $T^k(L_1)$. Это индуцированное линейное отображение мы обозначим через ψ^* (следовало бы писать ψ_k^*)

$$\psi^*: \mathrm{T}^k(L_2) \to \mathrm{T}^k(L_1). \qquad (87)$$

Отображение ψ^* определяется следующим образом. Пусть $a(v_1,..., v_k)$ — данная полилинейная форма в $T^k(L_2)$, где $v_1,..., v_k \in L_2$. Пусть теперь $u_1,..., u_k$ — любые векторы в L_1 ($u_1,..., u_k \in L_1$). С этими векторами мы сопоставим число $a(\psi(u_1), ..., \psi(u_k))$. Так как $a(v_1, ..., v_k)$- полилинейная форма и $v=\psi(u)$ — линейное отображение, то $a(\psi(u_1), ..., \psi(u_k))$ линейно зависит от каждого из аргументов $u_1,..., u_k$. Тем самым с каждой формой $a \in T^k(L_2)$ сопоставлена форма в $T^k(L_1)$; ее и обозначают через ψ^*a; $\psi^*a \in T^k(L_1)$. Согласно сказанному значение формы $\psi^*(a)$ для набора векторов $u_1,..., u_k \in L_1$ определяется равенства

$$(\psi^*a)(u_1, \ldots, u_k) = a(\psi(u_1), \ldots, \psi(u_k)). \tag{88}$$

Легко проверить, что для любых форм $a, b \in T^k(L_2)$ и для любого $\alpha \in R$ имеют место равенства:

$$\psi^*(a+b) = \psi^*a + \psi^*b, \quad \psi^*(\alpha a) = \alpha\psi^*a. \tag{89}$$

Тем самым отображение (87) действительно есть линейным. Отметим еще равенство

$$\psi^*\alpha = \alpha, \tag{90}$$

которое мы принимаем как определение ψ^* для форм нулевой степени.

2. Пусть теперь a и b — две формы с любым числом аргументов:

$$a \in T^k(L_2), \quad b \in T^l(L_2).$$

По определению тензорного произведения имеем:

$$(a \otimes b)(v_1, \ldots, v_k, v_{k+1}, \ldots, v_{k+l}) = \\ = a(v_1, \ldots, v_k)\, b(v_{k+1}, \ldots, v_{k+l});$$

здесь

$$v_j \in L_2 \quad (j = 1, \ldots, k, k+1, \ldots, k+l).$$

Отсюда для любых $u_j \in L_1$ следует:

$$\begin{aligned} &\psi^*(a \otimes b)(u_1, \ldots, u_k, u_{k+1}, \ldots, u_{k+l}) = \\ &\quad = (a \otimes b)(\psi(u_1), \ldots, \psi(u_k), \psi(u_{k+1}), \ldots, \psi(u_{k+l})) = \\ &\quad = a(\psi(u_1), \ldots, \psi(u_k))\, b(\psi(u_{k+1}), \ldots, \psi(u_{k+l})) = \\ &\qquad = (\psi^*a \otimes \psi^*b)(u_1, \ldots, u_k, u_{k+1}, \ldots, u_{k+l}). \end{aligned} \tag{91}$$

Кроме того, из (88) непосредственно получается равенство альтернаций

$$[\psi^*a](u_1, \ldots, u_k) = [a](\psi(u_1), \ldots, \psi(u_k)). \tag{92}$$

Числовые тождества (90) и (91) приводят к соответствующим тензорным равенствам:

$$\psi^*(a \otimes b) = \psi^*a \otimes \psi^*b, \tag{93}$$

$$[\psi^* a] = \psi^* [a]. \tag{94}$$

3. ***Следствие***. Если *a* — *внешняя форма в* $T^k(L_2)$, то $\psi^*(a)$ — *внешняя форма в* $T^k(L_1)$.

Доказательство. Пусть $[a] = a$. Тогда

$$[\psi^* a] = \psi^* [a] = \psi^* a,$$

что и утверждалось.

4. Доказанное утверждение можно выразить еще следующим способом: если $a \in \Lambda^k(L_2)$, то $\psi^* a \in \Lambda^k(L_1)$. Иначе говоря, линейное отображение ψ индуцирует линейное отображение

$$\psi^*: \Lambda^k(L_2) \to \Lambda^k(L_1). \tag{95}$$

Отображение (95) является, очевидно, сужением отображения (87) на

$$\Lambda^k(L_2) \subset \mathrm{T}^k(L_2);$$

мы сохранили для него символ ψ^*.

Далее мы рассматриваем внешние формы и поэтому вместо *a* и *b* пишем ω и σ.

5. Из (93) и (94) следует также равенство

$$\psi^*(\omega \wedge \sigma) = \psi^*\omega \wedge \psi^*\sigma. \tag{96}$$

В самом деле, если

$$\omega \in \Lambda^k(L_2), \;\; \sigma \in \Lambda^l(L_2),$$

то

$$\psi^*(\omega \wedge \sigma) = \psi^*\left(\frac{(k+l)!}{k!\,l!}[\omega \otimes \sigma]\right) = \\ = \frac{(k+l)!}{k!\,l!}[\psi^*\omega \otimes \psi^*\sigma] = \psi^*\omega \wedge \psi^*\sigma.$$

Кроме того, если $\alpha \in R$, то согласно (89)

$$\psi^*(\alpha\omega) = \alpha\psi^*(\omega).$$

6. Теперь мы займемся внешними дифференциальными формами. Изложенные сейчас вещи мы перенесем из алгебры внешних форм в дифференциальное исчисление.

Пусть в области $U \in E$ задана функция $y=f(x)$, $x \in U$, значения которой суть элементы некоторого евклидова пространства $\tilde{E}$ ($y \in \tilde{E}$). Иначе говоря, дано отображение области *U* евклидова пространства *E* в евклодово пространство $\tilde{E}$:

$$f: U \to \tilde{E}. \tag{97}$$

Размерности *E* и $\tilde{E}$ могут быть любыми; мы обозначим их соответственно через *n* и *m*.

Предположим, что отображение *f* дифференцируемо в каждой точке области *U*. Тогда в произвольной точке $x \in U$ существует производная отображения *f*, которая представляет собой линейное отображение T_x в

$\tilde{E}$. Напомним, что T_x обозначает касательное пространство к E у точке x (см. п. 7.2.1). Через $\tilde{T}_y$ мы будем иногда обозначать касательное пространство к $\tilde{E}$ в точке $y=f(x)$.

Производную отображения f в точке x будем обозначать через $f'(x)$. Что касается определения производной, то мы ограничимся отсылкой к п. 2 п. 7.2.3, где дано определения производной функции с числовыми значениями. Оно почти без изменений переносится на случай производной общего отображения: довольно всюду в п. 2 вместо отображений $U \to$ R и $T_x \to$ R иметь в виду соответственно отображения $U \to \tilde{E}$ и $T_x \to \tilde{E}$.

Дифференциал отображения f в точке x пишется в виде

$$\eta = f'(x)\,\xi. \tag{98}$$

Здесь ξ — произвольный вектор из T_x, η (дифференциал отображения) — его образ в $\tilde{T}_y$.

Производная и дифференциал отображения являются инвариантными объектами (их определение не используют координатных систем). Но допустим, что в E и $\tilde{E}$ введены координаты (будем считать - декартовы прямоугольные, хотя это не везде существенно). Тогда производная и дифференциал получат координатное представление. Прежде всего, вместо (97) можно написать координатное представление самого отображения f:

$$y^1 = f^1(x^1 \ldots, x^n), \ldots, y^m = f^m(x^1 \ldots, x^n). \tag{99}$$

Отсюда известным путем выводится координатное представление дифференциала (98):

$$\left.\begin{array}{l} \eta^1 = f_1^1(x)\,\xi^1 + \ldots + f_n^1(x)\,\xi^n, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ \eta^m = f_1^m(x)\,\xi^1 + \ldots + f_n^m(x)\,\xi^n, \end{array}\right\} \tag{100}$$

где

$$(\xi^1, \ldots, \xi^n) = \xi, \quad (\eta^1, \ldots, \eta^m) = \eta,$$

$f^i_j(x)$ — частная производная функции $f^i(x)= f^i(x^1, \ldots,x^n)$ по x^j, вычисленная в точке x.

Равенства (100) дают координатное представление производной в виде функциональной (якобиевой) матрицы

$$f'(x) = \begin{pmatrix} f_1^1(x) & \ldots & f_n^1(x) \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot \\ f_1^m(x) & \ldots & f_n^m(x) \end{pmatrix}. \tag{101}$$

Далее символ $f'(x)$ (или просто f') в равной мере употребляется как для обозначения самих отображений, так и для их координатных представлений, т.е. для матриц. В случае $m=n$ символ det $f'(x)$ будет обозначать определитель якобиевой матрицы, квадратной, поскольку $m = n$.

7. Согласно п. 4 линейное отображение $f'(x)$ индуцирует линейное отображение внешних дифференциальных форм из пространства

$$\Lambda^k(\tilde{T}_y),\ y=f(x),$$

в пространство $\Lambda^k(T_x)$. Его следовало бы (согласно п. 4) обозначать через f'^*. Однако с целью упрощения запись принята писать f^*. Подведем итог.

8. Итак, данное (вообще говоря, нелинейное) дифференцированное в области U отображение

$$f\colon U \to \tilde{E}$$

определяет в каждой точке $x \in U$ в качестве своей производной линейное отображение $f'(x)$ касательного пространства T_x в касательное пространство $\tilde{T}_y$ $(y=f(x))$:

$$f'(x)\colon T_x \to \tilde{T}_y.$$

Это последнее отображение индуцирует (при любому k и в любой точке x) линейное отображение пространства $\Lambda^k(\tilde{T}_y)$ в пространство $\Lambda^k(T_x)$:

$$f^*\colon\ \Lambda^k(\tilde{T}_y) \to \Lambda^k(T_x).$$

Таким образом, с каждой k-формой $\omega \in \Lambda^k(\tilde{T}_y)$ сопоставляется форма $f^*\omega \in \Lambda^k(T_x)$.

9. Из пп. 1—5 вытекают следующие алгебраические свойства отображения f^*:

$$1) \qquad f^*(dy) = f'_1(x)dx^1 + \ldots + f'_n(x)dx^n. \qquad (102)$$

В самом деле, обозначая $f'(x)\xi$ через $\eta(\xi)$, имеем согласно (100):

$$f^*(dy^i)(\xi) = dy^i(\eta(\xi)) = \eta^i = f^i_1(x)\xi^1 + \ldots + f^i_n(x)\xi^n =$$
$$= \left(f^i_1(x)\,dx^1 + \ldots + f^i_n(x)\,dx^n\right)(\xi),$$

что и означает (102).

Вместо (102) можно писать коротко

$$f^*(dy^i) = df^i(x); \qquad (103)$$

здесь $df^i(x)$ — производная функции $f^i(x)$.

Ясно, что (102) и (103) определяются не только отображением $y=f(x)$, но и выбранными в E, $\tilde{E}$ координатными системами.

Следующие три свойств отображения f^* записываются в виде инвариантных соотношений:

2) $$f^*(\omega_1+\omega_2)=f^*(\omega_1)+f^*(\omega_2)$$
$$\omega_1,\ \omega_2\in\Lambda^k(\tilde{T}_y);$$

3) $$f^*(\omega\wedge\sigma)=f^*\omega\wedge f^*\sigma;$$

4) $$f^*(g)=g\circ f,\qquad g\in\Lambda^0(\tilde{U}).$$

Запись $g\in\Lambda^0(\tilde{U})$ означает, что g есть форма нулевой степени, т.е. функция, которая задана в области $\tilde{U}\subset\tilde{E}$.

Свойства 2) и 3) верны, поскольку они не отличаются от соотношений (89) и (96); см. пп. 1 и 5. Свойство 4) не отличается от соотношения (90). В самом деле, значение $g^\circ f$ в точке x равно значению g в точке $y = f(x)$.

Из 3) и 4), в частности, имеем

5) $$f^*(g\omega) = (g^\circ f)f^*\omega.$$

10. Операция f^* определена инвариантно. Указанные выше свойства этой операции позволяют легко получить ее координатное представление. Пусть дано координатное представление формы $\omega\in\Lambda^k(\tilde{T}_y)$:

$$\omega=*\sum\omega_{i_1\ldots i_k}(y)\,dy^{i_1}\wedge\ \ldots\ \wedge dy^{i_k}. \tag{104}$$

Тогда на основании свойств 2) -5) получаем

$$f^*\omega=*\sum\left(\omega_{i_1\ldots i_k}\circ f\right)\left(f^*\left(dy^{i_1}\right)\wedge\ \ldots\ \wedge f^*\left(dy^{i_k}\right)\right). \tag{105}$$

Отсюда и вследствие (101) имеем координатное представление $f^*\omega$.

Указание. Формула (105) означает, что координатное представление $f^*\omega$ получается простой формальной подстановкой. А именно, нужно в правой части выражения (104) заменить $y^1, \ldots, y^m$ их выражениями (99) и записать $f^*(dy^1), \ldots, f^*(dy^m)$ по формуле (103).

11. В частности, если $m = n$ и $k = n$, то матрица f' является квадратной, а форма имеет одночленный вид.

$$\omega=g\,dy^1\wedge\ \ldots\ \wedge dy^n, \tag{106}$$

где $g = g(y)$. В этом случае

$$f^*\omega=(g\circ f)(\det f')\,dx^1\wedge\ \ldots\ \wedge dx^n. \tag{107}$$

12. Имеет место теорема

Теорема.

$$f^*(d\omega)=d(f^*\omega). \tag{108}$$

Доказательство сначала проведем для форм нулевой степени. Пусть g — функция, заданная в области $\tilde{U}\subset\tilde{E}$.

Имеем

$$f^*(g) = g \circ f.$$

Отсюда

$$df^*(g) = dg \circ df = dg \cdot f' = f^*(dg). \qquad (109)$$

Пусть теперь ω - произвольная *k*-форма, которая задана в $\tilde{U} \subset \tilde{E}$:

$$\omega = * \sum \omega_{i_1 \dots i_k} dy^{i_1} \wedge \dots \wedge dy^{i_k}.$$

Тогда

$$f^*\omega = * \sum \left(f^*\omega_{i_1 \dots i_k}\right) f^*\left(dy^{i_1}\right) \wedge \dots \wedge f^*\left(dy^{i_k}\right).$$

Вследствие (103)

$$df^*(dy^i)=0.$$

Отсюда и на основании (109)

$$d\left(f^*\omega\right) = * \sum d\left(f^*\omega_{i_1 \dots i_k}\right) \wedge f^*\left(dy^{i_1}\right) \wedge \dots \wedge f^*\left(dy^{i_k}\right) =$$
$$= * \sum f^*\left(d\omega_{i_1 \dots i_k}\right) \wedge f^*\left(dy^{i_1}\right) \wedge \dots \wedge f^*\left(dy^{i_k}\right) = f^*(d\omega).$$

Теорема доказана.

13. Пусть, как и раньше, f: $U \to \tilde{E}$, $U \subset E$; пусть $\tilde{U}$ обозначает образ области U в пространстве $\tilde{E}$. Предположим, что $m=n$ и что отображение f: $U \to \tilde{U}$ дифеоморфно. Допустим, что в E и $\tilde{E}$ введены декартовы прямоугольные координаты. Тогда f получает координатное представление: $y^i = f^i(x^1, \dots, x^n)$ $(u=1, \dots, n)$. Тем самым набор чисел $(x^1, \dots, x^n) = x \in U$ определяет точку $y=(y^1, \dots, y^n) = f(x) \in \tilde{U}$. Числа $(x^1, \dots, x^n)$ называют *криволинейными координатами* точки $y \in \tilde{U}$.

Пусть в области $\tilde{U} \subset \tilde{E}$ задана форма ω и дано ее координатное представление в координатах $y^1, \dots, y^m$. Тогда координатное представление формы $f^*\omega$ в пространстве E в координатах $(x^1, \dots, x^n)$ называется записью формы ω в криволинейных координатах $(x^1, \dots, x^n)$ в пространстве $\tilde{E}$.

14. **Пример**. Пусть роль $\tilde{E}$ играет двумерная плоскость с декартовыми координатами (y^1, y^2). Роль E — двумерная плоскость с декартовими координатами (ρ,θ) и отображение f дано формулами

$$y^1 = \rho\cos\theta, \quad y^2 = \rho\sin\theta \quad (\rho > 0).$$

Тогда (ρ,θ) — криволинейные (полярные) координаты в плоскости $\tilde{E}$. Рассмотрим форму $\omega = dy^1 \wedge dy^2$. Ее запись в криволинейных (полярных) координатах будет

$$f^*\omega = (\cos\theta\, d\rho - \rho\sin\theta\, d\theta) \wedge (\sin\theta\, d\rho + \rho\cos\theta\, d\theta) =$$
$$= \rho\,(d\rho \wedge d\theta).$$

15. Пусть в декартовых координатах даны форма ω и ее внешний дифференциал $d\omega$. Их записи в криволинейных координатах будут соответственно $f^*\omega$ и $f^*d\omega$. Но $f^*d\omega=df^*\omega$. Поэтому вычисления внешнего дифференциала в криволинейных координатах можно проводить в пространстве $\tilde{E}$ непосредственно по координатной записи формы $f^*\omega$, не обращая внимания на криволинейный характер координат, в которых она записана.

16. В заключение этого раздела мы рассмотрим координатное представление индуцированного отображения f^*, причем для любых тензоров (т.е. для любых полилинейных форм, не обязательно внешних).

Пусть, как и в п. 13, дано $f\colon U \to \tilde{E}$ $(m = n)$, или $y^i = f^i(x^1, ..., x^n)$ $(i=1, 2, .., n)$. Вместе с f определена производная f' как линейное отображение $\eta=f'\xi$ векторов из касательного просторную T_x $(\xi\in T_x)$ в касательное пространство $T_y(y=f(x))$: $\xi\to\eta\in T_y$. В координатной записи имеем:

$$\eta^i=\sum\frac{\partial y^i}{\partial x^\alpha}\xi^\alpha;$$

частные производные подсчитаны в точке x.

Можно написать также

$$f^*\,dy^i=\sum\frac{\partial y^i}{\partial x^\alpha}dx^\alpha.$$

где $dx^1, ..., dx^n$ — координатные формы в T_x, $dy^l,..., dy^n$— координатные формы в T_y.

Пусть в T_y задан k-тензор, т.е. полилинейная форма $\in \overset{\otimes}{T}{}^k_y$:

$$a=\sum a_{i_1\,\dots\,i_k}dy^{i_1}\otimes\,\dots\,\otimes dy^{i_k}.$$

Отсюда и из предыдущего

$$f^*a=\sum a_{i_1\,\dots\,i_k}\frac{\partial y^{i_1}}{\partial x^{\alpha_1}}\cdots\frac{\partial y^{i_k}}{\partial x^{\alpha_k}}dx^{\alpha_1}\otimes\,\dots\,\otimes dx^{\alpha_k}.$$

Таким образом, если положить

$$f^*a=\sum\tilde{a}_{\alpha_1\,\dots\,\alpha_k}dx^{\alpha_1}\otimes\,\dots\,\otimes dx^{\alpha_k},$$

то f^* представится в координатах формулами

$$\tilde{a}_{\alpha_1\,\dots\,\alpha_k}=\sum a_{i_1\,\dots\,i_k}\frac{\partial y^{i_1}}{\partial x^{\alpha_1}}\cdots\frac{\partial y^{i_k}}{\partial x^{\alpha_k}}. \qquad (110)$$

Еще раз подчеркнем, что f^* зависит от k (ясно, что от k зависит и число формул (110) и число слагаемых в их правых частях).

Заметим, наконец, что f можно рассматривать не как отображение точек, а как преобразование координат. Тогда формулы (110) выражают «ковариантный закон» преобразование координат тензора при переходе к новой координатной системе.

7.3. Интегрирование внешних дифференциальных форм

7.3.1. Интеграл от внешней формы по сингулярному кубу

1. Пусть R^k — декартово координатное представление k-мерного евклидова пространства. Будем обозначать произвольную точку в R^k буквой t, а координаты ее — той же буквой t с надлежащими индексами

$$t=(t^1, \ldots, t^k) \in \mathbb{R}^k.$$

Обозначим через h так называемый стандартный куб в R^k, т.е. единичный координатный куб $[0, 1]^k$

$$h = [0,1]^k.$$

По определению

$$[0, 1]^k \Leftrightarrow 0 \leqslant t^i \leqslant 1 \qquad (i=1, 2, \ldots, k).$$

Рассмотрим произвольную область U, $U \in R^k$, содержащую куб h, и допустим, что в области U задана внешняя дифференциальная форма σ степени k. Координатная запись k-формы в k-мерном пространстве имеет одночленный вид

$$\sigma = g(t^1, \ldots, t^k)\, dt^1 \wedge \ldots \wedge dt^k. \tag{111}$$

Предположим, что σ непрерывна в U; это равносильно предположению непрерывности функции $g(t^1.....t^k)$.

При этом условии функция $g(t^1.....t^k)$ заведомо интегрируема в h.

2. Интеграл по кубу $h = [0, 1]^k$ от формы σ, заданной в пространстве R^k, определяется равенством

$$\int\limits_h \sigma = \int\limits_{[0,1]^k} g(t^1, \ldots, t^k)\, dt^1 \ldots dt^k, \tag{112}$$

где справа написан обычный k-кратный интеграл по $h = [0, 1]^k$

3. Пусть теперь ω— внешняя дифференциальная k-форма, заданная и непрерывная в некоторой области V пространства E. Обозначим размерность E через n, считая $n \geq k$.

Рассмотрим неперервно-дифференцируемое отображение

$$c\colon U \to V \subset E.$$

Вместе с ним определено его сужение на h, которое мы обозначим той же буквой c:

$$c: h \to V. \tag{113}$$

Отображение (113) называется *k-мерным сингулярным кубом* в пространстве E, или в области $V \subset E$. Подчеркнем, что сингулярный куб c представляет собой не образ куба h в пространстве E, а именно отображение (113). Можно сказать, что сингулярный куб c представляет собой множество пар вида (x, t), где $t \in h$, $x = c(t) \in E$; две пары (x_1, t_1) и (x_2, t_2) считаются различными, если различны хотя бы только t_1 и t_2. Название сингулярного куба (именно, прилагательное: «сингулярный») связано с тем, что отображение (113) может иметь особенности; например, не исключается, что h отобразится в одну точку.

Вместе с сингулярным кубом c определено для каждой точки $t \in h$ линейное отображение

$$c': T_t \to T_x, \tag{114}$$

т.е. производная от c; здесь $T_t = T_t(\mathrm{R}^k)$ — касательное пространство к R^k в точке t, $T_x = T_x(E)$ — касательное пространство к E у точке $x = c(t)$. Отображение (114) в свою очередь индуцирует известное нам линейное отображение

$$c^*: \Lambda^k(T_x) \to \Lambda^k(T_t). \tag{115}$$

Таким образом, с каждой k-формой ω в области $V \subset E$ сопоставляется k-форма $c^*\omega$ на стандартном кубе $h \subset R^k$. Форма $c^*\omega$ определена инвариантно в том смысле, что ее определение не использует координатных систем в E. Ее координатное представление в R^k является одночленным (см. (111).

4. *Интегралом от внешней дифференциальной k-формы* ω по сингулярному k-мерному кубу c в области V называется число, которое обозначается и определяется согласно следующему равенству:

$$\int_c \omega = \int_h c^*\omega. \tag{116}$$

Правая часть этого равенства уже определена равенством (112); в этом случае $\sigma = c^*\omega$. Для сохранности стандарта терминологии можно считать, что в правой части (116) буква h обозначает сингулярный куб I: $h \to h$, где I — тождественное отображение. Определение интеграла от ω по c, как мы видим, инвариантно. В частном случае $k=0$ нульмерным сингулярным кубом называется отображение в E стандартного куба нульмерного пространства, состоящего из одной точки

$$c: \{0\} \to E.$$

Пусть $c(0)$ — образ точки $\{0\}$ в $E(c(0)\in E)$. Если ω — нульмерная форма в E, т.е. функция $\omega=\omega(x)$, $x\in E$, то по определению

$$\int_c \omega = \omega(c(0)).$$

5. Предположим, что в пространстве E введены декартовы прямоугольные координаты; будем для точки пространства E и для ее координат употреблять букву x: $x = (x^1, \dots x^n)$.

Тогда отображение c получит координатное представление:

$$\left.\begin{array}{l} x^1 = c^1(t^1, \dots, t^k), \\ \dots\dots\dots\dots \\ x^k = c^k(t^1, \dots, t^k), \\ x^{k+1} = c^{k+1}(t^1, \dots, t^k), \\ \dots\dots\dots\dots \\ x^n = c^n(t^1, \dots, t^k). \end{array}\right\} \qquad (117)$$

Теперь все сказанное выше мы сведем в один (простой) рецепт для вычисления $c^*\omega$ и интеграла от ω по c. Для простоты записи возьмем в E одночленную k-форму

$$\omega = G(x^1, \dots, x^n)\, dx^1 \wedge \dots \wedge dx^k. \qquad (118)$$

Согласно п. 7.2.7 имеем

$$c^*\omega = (G\circ c)\, c^*(dx^1 \wedge \dots \wedge dx^k) = \\ = (G\circ c)\, c^* dx^1 \wedge \dots \wedge c^* dx^k, \qquad (119)$$

$$c^* dx^i = (D_1 c^i)\, dt^1 + \dots + (D_k c^i)\, dt^k. \qquad (120)$$

Таким образом, получение $c^*\omega$ фактически сводится к подстановке в (118) вместо $x^1, \dots x^n$ их выражений из (117).

6. Из (119) и (120)

$$c^*\omega = (G\circ c)\det\left(\frac{x^1, \dots, x^k}{t^1, \dots, t^k}\right) dt^1 \wedge \dots \wedge dt^k \qquad (121)$$

(см. п. 7.1.14 в частном случае $n = k$; см. также п. 11 п. 7.2.6).

Из (121) получаем для одночленной формы (4.221)

$$\int_c \omega = \int_{[0,1]^k} G(x^1(t), \dots, x^n(t)) \det\left(\frac{x^1, \dots, x^k}{t^1, \dots, t^k}\right) dt^1 \dots dt^k. \qquad (122)$$

Формула (122) написана с полной подробностью и может быть положена в основу фактического вычисления интегралов от внешних форм по сингулярному кубу.

7. Отметим простейший случай применения формулы (122) для интеграла от линейной формы

$$\omega = Pdx^1$$

где $P = P(x^1, x^2, x^3)$. В этом случае интегрирования должно вестись по одномерному сингулярному кубу c, который обычно называют ориентированной дугой

$$c\colon \begin{cases} x^1 = c^1(t), \\ x^2 = c^2(t), \\ x^3 = c^3(t). \end{cases}$$

Если мы для удобства записи положим $P(t)=P(c^1(t), c^2(t), c^3(t))$, то формула (122) в этом частном случае дает

$$\int_c \omega = \int_{[0,\,1]} P(t)\frac{dx^1}{dt}\,dt.$$

8. Формуле (122) можно придать более краткую запись. С этой целью обозначим через $\hat{c}$ отображение, которое состоит из c и последующего проектирования образа на координатное k-мерное подпространство первых по счету осей. Мы будем рассматривать это подпространство, отвлекаясь от остальных осей. Тогда для отображения $\hat{c}$ имеем координатное представление (см. (117)):

$$\begin{array}{l} x^1 = c^1(t^1, \ldots, t^k), \\ \cdots\cdots\cdots\cdots \\ x^k = c^k(t^1, \ldots, t^k). \end{array}$$

Для краткости напишем

$$\det\left(\frac{x^1, \ldots, x^k}{t^1, \ldots, t^k}\right) = \det \hat{c}'.$$

Отсюда из (122)

$$\int_c \omega = \int_{[0,\,1]^k} (G \circ c)(\det \hat{c}'); \qquad (123)$$

здесь снова полезно вспомнить п.11 п. 7.2.6. Равенство (123) и есть та формула, которую мы имели в виду написать. В ней, как часто делается, не написан элемент объема.

Замечание. Если одночленная k-форма имеет вид

$$\omega = G(x^1, \ldots, x^n)\,dx^{i_1} \wedge \ldots \wedge dx^{i_k},$$

то в правой части (123) в качестве $\hat{c}$ следует брать отображение

$$x^{i_1} = c^{i_1}(t^1, \ldots, t^k),$$
$$\cdots\cdots\cdots$$
$$x^{i_k} = c^{i_k}(t^1, \ldots, t^k).$$

Замечание. Интеграл от формы общего вида исчисляется согласно (123) почленно (с учетом предыдущего замечания).

9. Для дальнейшего нам нужно вспомнить теорему и формулу замены переменных в кратном интеграле.

Пусть U — некоторая область в R^n и $y=s(x)$ $(x \in U, y \in R^n)$ непрерывно-дифференцируемое отображение $U \to R^n$, инъективное, т.е. взаимно-однозначное на образ $s(U)$:

$$s: \begin{cases} y^1 = s^1(x^1, \ldots, x^n), \\ \cdots\cdots\cdots \\ y^n = s^n(x^1, \ldots, x^n). \end{cases}$$

Тогда, если $\det s'(x) \neq 0$ в области U, то для любой интегрируемой функции $f(y)$, $y \in s(U)$, имеет место равенство

$$\int\limits_{s(U)} f(y)\, dy^1 \ldots dy^n = \int\limits_U f(s(x)) |\det s'(x)|\, dx^1 \ldots dx^n.$$

В более короткой записи:

$$\int\limits_U (f \circ s) |\det s'| = \int\limits_{s(U)} f.$$

Нам эта формула потребуется в частном случае, когда $s(U)$ совпадает с U. В этом случае

$$\int\limits_U (f \circ s) |\det s'| = \int\limits_U f. \tag{124}$$

10. Пусть c и $\tilde{c}$ — два k-мерных сингулярных кубы:

$$c\colon\ h \to E, \qquad \tilde{c}\colon\ h \to E.$$

Предположим, что существует взаимно-однозначное гладкое отображение p стандартного куба h на себя, которое имеет положительный определитель ($\det p' > 0$) и для которого $\tilde{c} = c \circ p$. В таком случае говорят, что $\tilde{c}$ получен из c изменением параметризации. Мы будем говорить также, что $\tilde{c}$ эквивалентен c и будем писать: $\tilde{c} \sim c$.

Замечание. Так как $\tilde{c} = c \circ p$, то образ $\tilde{c}(h)$ куба h совпадает с его образом $c(h)$. Тем самым переход от c к $\tilde{c}$ означает только изменение прообразов произвольной точки $x \in c(h) = \tilde{c}(h)$. Пусть $x = c(t)$, $t \in h$ и $x = \tilde{c}(\tilde{t})$, $\tilde{t} \in h$. Если $t = (t^1, \ldots, t^k)$, то числа $t^1, \ldots, t^k$ можно назвать криволинейными координатами точки x в координатной системе

$c: h \to E$ (см. п. 13 п. 7.2.7); если $\tilde{t} = (\tilde{t}^1, \ldots, \tilde{t}^k)$, то числа $\tilde{t}^1, \ldots, \tilde{t}^k$ являются криволинейными координатами той же точки x в другой координатной системе $\tilde{c}: h \to E$. Тем самым, дело заключается в преобразовании криволинейных координат по формуле $t = p(\tilde{t})$. Криволинейные координаты часто называют параметрами. Отсюда — выражение, которое мы употребили выше: $\tilde{c}$ получен из c изменением параметризации.

11. Легко показать, что

1) $c \sim c$;

2) если $\tilde{c} \sim c$, то $c \sim \tilde{c}$;

3) если $\tilde{c} \sim c$, $\tilde{\tilde{c}} \sim \tilde{c}$, то $\tilde{\tilde{c}} \sim c$,

12. Будем говорить, что сингулярный куб $\tilde{c}$ эквивалентен сингулярному кубу c после изменения ориентации, и писать $\tilde{c} \sim -c$, если существует отображение p, для которого $\tilde{c} = c \circ p$ и $\det p' < 0$ при сохранении остальных условий.

13. Имеют место следующие предложения:

1) *Если $\tilde{c} \sim c$, то для любой k-формы* ω

$$\int_{\tilde{c}} \omega = \int_{c} \omega.$$

Доказательство. Возьмем для простоты в качестве формы ω одночленную форму (118). Имеем согласно формуле (123)

$$\int_{\tilde{c}} \omega = \int_{c \circ p} \omega = \int_{h} (c \circ p)^* \omega = \int_{h} (G \circ (c \circ p)) \det \widehat{(c \circ p)}' =$$
$$= \int_{h} ((G \circ c) \circ p) (\det \hat{c}' \circ p) \det p'.$$

Здесь учтено, что $\hat{p} = p$. Теперь, поскольку $\det p' > 0$, имеем по теореме о замене переменных

$$\int_{\tilde{c}} \omega = \int_{h} ((G \circ c) \circ p) \det \hat{c}' \circ p) \mid \det p' \mid =$$
$$= \int_{h} (G \circ c) \det \hat{c}' = \int_{h} c^* \omega = \int_{c} \omega.$$

2) *Если $\tilde{c} \sim -c$, то для любой k-формы* ω

$$\int_{\tilde{c}} \omega = -\int_{c} \omega.$$

Доказательство легко усматривается из предыдущего с учетом условия $\det p' < 0$.

14. Только что доказанные теоремы выражают особую роль косой симметрии формы для ее интегрирования. Именно косая симметрия формы вызывает появление определителя в правой части формулы (123), что вместе с теоремой о замене переменных обеспечивает инвариантность интеграла относительно изменения параметризации.

7.3.2. Понятие цепи. Интеграл от формы по цепи

1. В качестве наглядного источника общего понятия цепи в евклидовом пространстве E можно указать дугу A_0A_p, составленную из нескольких ориентированных дуг A_0A_1, A_1A_2, ..., $A_{p-1}A_p$. Каждая ориентированная дуга $A_{i-1}A_i$ представляет собой некоторый одномерный сингулярный куб; обозначим его для краткости через c_i. Тогда дугу A_0A_p можно рассматривать как набор одномерных сингулярных кубов $c_1, c_2, ..., c_p$. По некоторым соображениям, содержание которых выяснится чуть дальше, этот набор обозначают в виде суммы $c_1 + c_2 + ... + c_p$. Такую сумму назовем формальной (поскольку пока это лишь символ, который обозначает набор $c_1, c_2, ..., c_p$). Порядок записи слагаемых в формальной сумме для нас безразличен. Это естественно. В самом деле, переставляя в записи суммы, например, c_1 и c_2, мы в действительности никаких изменений с этими сингулярными кубами не делаем; поэтому $c_1 + c_2 + ... + c_p$ и $c_2 + c_1 + ... + c_p$ обозначают одну и ту же дугу A_0A_p. Формальная сумма $c_1 + c_2 + ... + c_p$ представляет собой пример одномерной цепи. Общее понятие одномерной цепи легко уяснить себе, хотя бы в главных чертах, исходя из этого примера путем некоторых обобщений. Прежде всего, мы предположим, что одномерные сингулярные кубы $c_1, c_2, ..., c_p$ могут быть выбраны произвольно (не обязательно в виде ориентированных дуг, которые в просттранстве последовательно приложены друг к другу). Затем мы будем брать сингулярные кубы $c_1 + c_2 + ... + c_p$ с произвольными действительными коэффициентами $\lambda_1, ..., \lambda_p$; эти коэффициенты пока будут писаться также формально. Таким образом, *одномерной цепью* мы назовем любую формальную сумму $\lambda_1 c_1 + \lambda_2 c_2 + ... + \lambda_p c_p$ (любую в том смысле, что одномерные сингулярные кубы $c_1, c_2, ..., c_p$ и действительные числа $\lambda_1, ..., \lambda_p$ могут быть выбраны как угодно). Содержательный смысл этих понятий выяснится путем их связи с теорией интегрирования.

2. Теперь мы будем рассматривать цепи любой размерности. Пусть $c_1, ..., c_p$ — некоторый набор k-мерных сингулярных кубов в E, $\lambda_1, ..., \lambda_p$

— набор действительных чисел; при этом мы считаем, что число λ_i сопоставлено с кубом c_i (i=1, 2, ,.., p). Совокупность таких двух наборов мы назовем *k-мерной цепью* в пространстве E. Обозначая цепь буквой C, запишем ее в виде формальной суммы

$$C = \lambda_1 c_1 + \ldots + \lambda_p c_p.$$

Пусть даны две цепи:

$$C = \lambda_1 c_1 + \ldots + \lambda_p c_p,$$
$$\tilde{C} = \mu_1 \tilde{c}_1 + \ldots + \mu_q \tilde{c}_q.$$

Определим линейные операции следующими равенствами:

$$C + \tilde{C} = \lambda_1 c_1 + \ldots + \lambda_p c_p + \mu_1 \tilde{c}_1 + \ldots + \mu_q \tilde{c}_q,$$
$$\alpha C = (\alpha\lambda_1) c_1 + \ldots + (\alpha\lambda_p) c_p.$$

Очевидно, что

$$(C + \tilde{C}) + \tilde{\tilde{C}} = C + (\tilde{C} + \tilde{\tilde{C}}),$$
$$1 \cdot C = C,$$
$$\alpha(\beta C) = (\alpha\beta) C.$$

Здесь всюду знак равенства употреблен в смысле тождественного совпадения объектов.

3. Пусть ω - внешняя дифференциальная форма степени k, которая определена в области $U \in E$; пусть c_1 — сингулярный куб ($c_1 : h \to E$) такой, что образ $c_1(h)$ лежит в U; пусть C — специальная k-мерная цепь вида $C = 1 \bullet c_1$:

Интеграл от ω по этой цепи определяется равенством

$$\int\limits_C \omega = \int\limits_{c_1} \omega.$$

Вспомним, что интеграл по сингулярному кубу нами уже определен раньше. В общем случае, если

$$C = \lambda_1 c_1 + \ldots + \lambda_p c_p,$$

положим, по определению,

$$\int\limits_C \omega = \lambda_1 \int\limits_{c_1} \omega + \ldots + \lambda_p \int\limits_{c_p} \omega.$$

4. ***Определение***. Две k -мерные цепи C и $\tilde{C}$ будем называть *равными*, если

$$\int\limits_{\tilde{C}} \omega = \int\limits_C \omega$$

для любой внешней дифференциальной формы ω степени k
Замечание. Отсюда и далее равенство

$$\tilde{C} = C$$

будет пониматься в смысле только что высказанного определения. Отметим, что равенства, которые были написаны в п. 2, остаются в силе.

Теорема. *Пусть* c *и* $\tilde{c}$ *— два* k*-мерных сингулярных кубы,* C *и* $\tilde{C}$ *— две* k*-мерные цепи:* $C = 1 \bullet c$, $\tilde{C} = 1 \bullet \tilde{c}$. *Тогда, если* $\tilde{c} \sim c$, *то* $\tilde{C} = C$, *если* $\tilde{c} \sim -c$, *то* $\tilde{C} = (-1)C$.

Замечание. Знак ~ означает эквивалентность в смысле определений п. 10 и 12 п. 7.3.1.

Доказательство теоремы непосредственно вытекает из п. 13, п. 7.3.1.

5. ***Определение***. Назовем k -мерную цепь θ *нулевой,* если

$$\int_\theta \omega = 0$$

для любой k -мерной формы ω.

Теорема. *Если* θ *— нулевая* k*-мерная цепь и* $c_1, \ldots, c_p$ *любые* k*-мерные сингулярные кубы, то*

$$\theta = 0 \bullet c_1 + \ldots + 0 \bullet c_p.$$

Обратно, цепь, написанная здесь справа, есть нулевой.

Доказательство очевидно.

6. Теперь легко убедиться, что k-мерные цепи, точнее, их классы эквивалентности, образуют линейное пространство.

Для этого следует проверить пять аксиом линейного пространства, остающиеся после трех аксиом, которые проверены в конце п. 2.

Имеем

$$C + \tilde{C} = \tilde{C} + C,$$

как прямое следствие определения п. 4. Далее,

$$C + \theta = C$$

и

$$C + (-1)C = \theta,$$

что следует из теоремы п. 5.

Наконец,

$$(\alpha + \beta)\, C = \alpha C + \beta C, \qquad \alpha\,(C + \tilde{C}) = \alpha C + \alpha \tilde{C}.$$

Чтобы установить эти соотношения, достаточно убедиться в равенстве соответствующих интегралов от произвольной формы ω степени k.

7. Обозначим линейное пространство k-мерных цепей в E через S^k Легко убедиться, что пространство S^k бесконечномерно. В самом деле, пусть p — любое натуральное число. Построим k -мерные сингулярные кубы $c_1, ..., c_p$ любым образом при соблюдении следующих двух условий:

1) $c_1(h)..... c_p(h)$ лежат соответственно в попарно непересекающихся областях $U_1, ..., U_p$;

2) цепи $C_1 = 1 \bullet c_1, .. ., C_p = 1 \bullet c_p$ являются ненулевыми.

Тогда цепи $C_1,..., C_p$ линейно-независимы. Чтобы доказать это, заметим, что существуют формы $\omega_1, ..., \omega_p$ степени k такие, что

$$\int\limits_{C_j} \omega_j \neq 0 \qquad (j = 1, \ldots, p).$$

Существование таких форм прямо следует из второго условия. Обозначим через ψ_j гладкую функцию в E, которая равна единице на образе $c_j(h)$ и равна нулю вне области U_j. Положим $\sigma_j = \psi_j\omega_j$. Тогда

$$\int\limits_{C_j} \sigma_j = \int\limits_{C_j} \omega_j \neq 0 \qquad (j = 1, \ldots, p).$$

Но при любому $s \neq j$ будет

$$\int\limits_{C_s} \sigma_j = 0.$$

Предположим, что имеет место соотношения

$$\lambda_1 C_1 + ... + \lambda_p C_p = \theta,$$

где θ - нулевая цепь. Тогда

$$\lambda_1 \int\limits_{C_1} \sigma_j + \ldots + \lambda_p \int\limits_{C_p} \sigma_j = \int\limits_{\theta} \sigma_j = 0,$$

где j =1, 2, ..., p. Отсюда и из предыдущего следует, что $\lambda_j = 0$. Тем самым линейная независимость цепей $C_1,. . ., C_p$ доказана.

8. Обозначим через W^k линейное пространство всех внешних дифференциальных форм степени k, определенных и гладких (бесконечно дифференцируемых) в евклидовом пространстве E. Таким образом, мы теперь предполаганм, что область, в которой даны объекты наших рассмотрений, совпадают со всем пространством (по сути дела, это ограничение существенно лишь в пп. 7.3.5 и 7.3.6). Легко убедиться, что пространство W^k бесконечномерно. В самом деле, пусть p — любое натуральное число. Построим формы $\omega_1, ..., \omega_p$ степени k как-нибудь при соблюдении следующих двух условий:

1) форма ω_j (j= 1, 2, . . ., p) не равна тождественно нулю;

2) форма ω_j (j= 1, 2, . . ., p) равна нулю во всех точах, лежащих вне некоторой области U_j, причем области U_1,. . ., U_p попарно не пересекаются.

Тогда формы ω_1, ..., ω_p линейно-независимые. Это очевидно, так как в силу двух указанных условий из соотношения

$$\lambda_1\omega_1 + \ \ldots \ + \lambda_p\omega_p = 0.$$

непосредственно следует, что

$$\lambda_1 = \ \ldots \ = \lambda_p = 0.$$

9. Если евклидово пространство E имеет размерность n, то определен набор бесконечномерных линейных пространств S^0, S^1, ..., S^n, а также W^0, W^1, ..., W^n. Согласно предыдущему S^k имеет в качестве своих элементов классы эквивалентности k-мерных цепей в E; элементами W^k являются формы степени k. Если $k>n$, то W^k состоит только из одной формы, тождественно равной нулю. Отсюда следует, что S^k при $k > n$ содержит только один класс эквивалентности, состоящий из нулевых цепей.

Замечание. Если C — некоторая k-мерная цепь, то мы позволим себе писать $C\in S^k$ (хотя следовало бы писать $C\in\{C\}\in S^k$, где $\{C\}$ — класс эквивалентности).

10. Зафиксируем k, считая $0\leq k\leq n$, и рассмотрим линейные пространства S^k и W^k. Пусть $C\in S^k$, $\omega\in W^k$. Будем обозначать через (ω, C) интеграл от формы ω по цепи C:

$$(\omega,\ C) = \int\limits_C \omega.$$

Имеем для любых

$$\alpha_1,\ \alpha_2 \in \mathbb{R},\ \ \omega,\ \omega_1,\ \omega_2 \in W^k,\ \ C,\ C_1,\ C_2 \in S^k:$$

$$1) \qquad (\alpha_1\omega_1 + \alpha_2\omega_2,\ C) = \alpha_1(\omega_1,\ C) + \alpha_2(\omega_2,\ C);$$

$$2) \qquad (\omega,\ \alpha_1 C_1 + \alpha_2 C_2) = \alpha_1(\omega,\ C_1) + \alpha_2(\omega,\ C_2).$$

Доказательство очевидно.

Далее:

3) Если (ω, C)=0 для любой цепи $C\in S^k$, то ω = 0. Доказательство от противного. Пусть

$$\omega = {*}\sum \omega_{i_1 \ldots i_k}(x)\, dx^{i_1} \wedge \ \ldots \ \wedge dx^{i_k}$$

и хотя бы один коэффициент этой формы в какой-нибудь точке отличен от нуля. Для определенности будем считать $\omega_{1\ldots k}(x_0)>0$, где x_0 — нулевая точка (0.....0).

Возьмем k-мерную координатную плоскость E^k: $x^{k+1}=\ldots=x^n=0$ и обозначим через h стандартный куб в E^k, через c — преобразование подобия пространства E^k с центром x_0 и с коэффициентом λ ($\lambda>0$). Имеем

$$c^*\omega=\omega_{1\ldots k}(\lambda x)\lambda^k\,dx^1\wedge\ \ldots\ \wedge dx^k,$$

где $x\in E^k$. При достаточно малом λ для $x\in h$ будет $\omega_{1\ldots k}(\lambda x)>0$. Поэтому, беря цепь $C=1\cdot c$, одержимо

$$(\omega,\ C)=\int\limits_h c^*\omega>0,$$

что противоречит предположению.

4) Если $(\omega, C)=0$ для любой формы $\omega\in W^k$, то $C=0$.

Последнее утверждение справедливо, поскольку именно так мы и определили равенство нулю k-мерной цепи.

11. Свойства 1) и 2) означают билинейный характер функции (ω, C), свойства 3) и 4) означают ее невирожденность. Невирожденная билинейная функция (ω, C) называется *сверткой* элементов ω и C ($\omega\in W^k$, $C\in S^k$).

Пространства W^k и S^k называются *сопряженными* относительно свертки (ω, C); см. п. 7.1.2.

12. Выше мы определили линейный оператор внешнего дифференцирования d. А именно, если ω — внешняя дифференциальная форма степени k, то при известных условиях определен ее внешний дифференциал $d\omega$, который является внешней дифференциальной формой степени $k+1$. Если $\omega\in W^k$, то $d\omega\in W^{k+1}$. Таким образом, мы имеем отображение

$$d: W^k\to W^{k+1}. \tag{125}$$

Наряду с этим существует линейный оператор ∂, который с произвольной k-мерной цепью C сопоставляет некоторую k — 1-мерную цепь ∂C, называемую *границей* цепи C. Если $C\in S^k$, то $\partial C\in S^{k-1}$. Таким образом, мы имеем отображение

$$\partial: S^k\to S^{k-1}. \tag{126}$$

Для удобства сравнения с соотношением (125) можно вместо (126) написать также

$$\partial: S^{k+1}\to S^k. \tag{127}$$

Согласно (125) и (127) операторы d и ∂ действуют в сопряженных пространствах. Более того, эти операторы оказываются сопряженными. Именно, для любой формы $\omega\in W^k$ и для любой цепи $C\in S^{k+1}$ справедливое равенство

$$(d\omega,\ C)=(\omega,\ \partial C), \tag{128}$$

которая и означает сопряженность операторов d и ∂. Равенство (128) называется *формулой Стокса.* Мы докажем ее чуть позднее. Сначала определим границу цепи, т.е. оператор ∂.

7.3.3. Граница цепи

1. Прежде всего определим границу сингулярного куба, точнее, границу цепи $C = 1 \bullet c$, где c — сингулярный куб.

Пусть в n-мерном евклидовом простстранстве E задан k-мерный сингулярный куб c: $h \to E$, где h, как обычно, стандартный куб в R^k. Буквой t мы, как и раньше, обозначаем произвольную точку пространства R^k: $t=(t^1, ..., t^k) \in R^k$. Рассмотрим в R^k координатную гиперплоскость, в которой лежат точки вида $t=(t^1,.., t^{i-1}, 0, t^{i+1}, ..., t^k)$. Обозначим ее через $R^{k-1}_{i,0}$ и введем в ней координатную систему, в которой числа $t^1,.., t^{i-1}, t^{i+1}, ..., t^k$ суть координаты точки $t \in R^{k-1}_{i,0}$. Тем самым $R^{k-1}_{i,0}$ становится k—1-мерным евклидовым пространством с данным координатным представлением. Одновременно в $R^{k-1}_{i,0}$ определится стандартный куб, который мы обозначим через $h_{i,0}$. Аналогичным образом рассмотрим гиперплоскость $R^{k-1}_{i,1}$, удерживающею точки вида $t=(t^1,.., t^{i-1}, 1, t^{i+1}, ..., t^k)$; в качестве координаты $t \in R^{k-1}_{i,1}$ примем числа $t^1,.., t^{i-1}, t^{i+1}, ..., t^k$. Определяемый при этом в $R^{k-1}_{i,1}$ стандартный куб обозначим через $h_{i,1}$. Ясно, что $h_{i,0}$, $h_{i,1}$ есть пара противоположных k— 1-мерных граней куба h. Обозначим через $c_{i,0}$ и $c_{i,1}$ отображение c: $h \to E$, суженное на грани $h_{i,0}$ и $h_{i,1}$:

$$c_{i,0}\colon\ h_{i,0} \to E, \quad c_{i,1}\colon\ h_{i,1} \to E.$$

Таким путем одновременно с k-мерным сингулярным кубом c определены k— 1-мерные сингулярные кубы $c_{i,0}$ и $c_{i,1}$ (i = l, 2, ..., k). Их можно назвать k — 1-мерными гранями сингулярного куба c. Граница ∂c сингулярного куба c (точнее, цепи $1 \bullet c$) определяется как k—1-мерная цепь, составленная из его k — 1-мерных граней с коэффициентами по следующему правилу:

$$\partial c = \sum_{i=1}^{k} \{(-1)^i c_{i,0} + (-1)^{i+1} c_{i,1}\}. \tag{129}$$

Если $(-1)^i c_{i,0}$ и $(-1)^{i+1} c_{i,1}$ назвать ориентированными (k—1-мерными) гранями сингулярного куба c, то можно сказать, что граница ∂c есть сумма ориентированных граней куба c. При k=2 наглядная иллюстрация определения границы дается рис. 4.23, где ориентированные одномерные грани изображены в виде внутренних стрелок (на рис. 1 сингулярный куб c есть тождественное отображение стандартного куба h на себя).

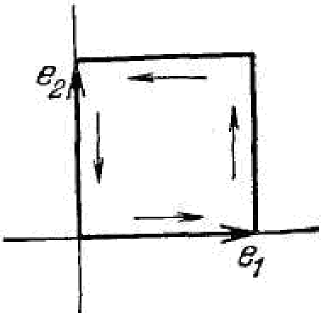


Рис. 1

Говоря о сингулярном кубе *c*, мы здесь всюду имеем в виду цепь 1 • *c* (как элемент линейного пространства S^k). Для произвольной *k*-мерной цепи

$$C = \lambda_1 c_1 + \ldots + \lambda_p c_p \tag{130}$$

граница определяется формулой

$$\partial C = \lambda_1 \partial c_1 + \ldots + \lambda_p \partial c_p. \tag{131}$$

2. Из определения ∂C непосредственно следует, что для любых

$$C_1,\ C_2 \in S^k,\ \alpha_1,\ \alpha_2 \in \mathbb{R}$$

$$\partial(\alpha_1 C_1 + \alpha_2 C_2) = \alpha_1 \partial C_1 + \alpha_2 \partial C_2.$$

Таким образом, ∂ есть линейный оператор, который отображает S^k в S^{k-1}. Основные свойства оператора ∂ будут изложены дальше (в частности, в п. 7.3.4 будет показано, что равные цепи имеют равные границы).

7.3.4. Доказательство формулы Стокса для цепи

1. Мы должны доказать формулу

$$\int_C d\omega = \int_{\partial C} \omega, \tag{132}$$

где

$$\omega \in W^{k-1}, \quad C \in S^k.$$

Достаточно рассмотреть цепь C=1• *c*, где *c* — некоторый *k*-мерный сингулярный куб в *E*. Мы имеем отображение

$$c\colon\ h \to E, \quad h \subset \mathbb{R}^k.$$

Как обычно, обозначим через *t* точку в R^k: $t = (t^1, \ldots, t^k)$.

Поскольку $\omega \in W^{k-1}$, то

$$c^*\omega = a(t^1, \ldots, t^k)\, dt^2 \wedge \ldots \wedge dt^k + \ldots, \tag{133}$$

где моготочие обозначает члены, которые содержат dt^1. Положим

$$\omega^{(1)} = a(t^1, \ldots, t^k)\, dt^2 \wedge \ldots \wedge dt^k. \tag{134}$$

Аналогично обозначим через $\omega^{(2)}, \ldots, \omega^{(k)}$ остальные (обозначенные многоточием) члены в правой части (133) так, что $\omega^{(i)}$ представляет собой некоторую функцию от $t^1, \ldots, t^k$, умноженную на внешнее произведение дифференциалов $dt^1, \ldots, dt^k$, где пропущен сомножитель dt^i. Тогда вместо (133) получаем

$$c^*\omega = \omega^{(1)} + \omega^{(2)} + \ldots + \omega^{(k)}. \tag{135}$$

Согласно теореме п. 7.2.6 (см. п. 12)

$$c^*(d\omega) = d(c^*\omega) = d\omega^{(1)} + \ldots + d\omega^{(k)}. \tag{136}$$

Напишем подробно $d\omega^{(1)}$:

$$d\omega^{(1)} = \frac{\partial a(t^1, \ldots, t^k)}{\partial t^1}\, dt^1 \wedge dt^2 \wedge \ldots \wedge dt^k. \tag{137}$$

Отсюда

$$\int_h d\omega^{(1)} = \int_h \frac{\partial a(t^1, \ldots, t^k)}{\partial t^1}\, dt^1\, dt^2 \ldots dt^k = \\ = \int_{h_1} a(1, t^2, \ldots, t^k)\, dt^2 \ldots dt^k - \\ - \int_{h_1} a(0, t^2, \ldots, t^k)\, dt^2 \ldots dt^k, \tag{138}$$

где h_1 — проекция куба h на координатную плоскость, которая содержит оси Ot^1 ..., Ot^k (h_1 рассматривается как область интегрирования).

Равенству (138) можно придать вид

$$\int_h d\omega^{(1)} = (-1) \int_{h_{1,0}} \omega^{(1)} + (-1)^2 \int_{h_{1,1}} \omega^{(1)},$$

где $h_{1,0}$ и $h_{1,1}$ определены в предыдущем параграфе. В общем случае, как легко проверить,

$$\int_h d\omega^{(i)} = (-1)^{i-1} \left\{ (-1) \int_{h_{i,0}} \omega^{(i)} + (-1)^2 \int_{h_{i,1}} \omega^{(i)} \right\}. \tag{139}$$

Множитель $(—1)^{i-1}$ появляется потому, что в общем случае в правой части формулы, аналогичной (137), на первом месте будет стоять dt^i; его придется переместить на i-е место. Заметим, дальше, что на $h_{i,0}$ и на $h_{i,1}$ (т.е. при $dt^i = 0$) значения $\omega^{(i)}$ и $c^*\omega$ совпадают. Поэтому из (139)

$$\int_h d\omega^{(i)} = (-1)^i \int_{h_{i,0}} c^*\omega + (-1)^{i+1} \int_{h_{i,1}} c^*\omega = \\ = (-1)^i \int_{c_{i,0}} \omega + (-1)^{i+1} \int_{c_{i,1}} \omega. \qquad (140)$$

Из (136) и (140)

$$\int_c d\omega = \int_h c^*(d\omega) = \sum_{i=1}^{k} \int_h d\omega^{(i)} = \\ = \sum_{i=1}^{k} \left\{ (-1)^i \int_{c_{i,0}} \omega + (-1)^{i+1} \int_{c_{i,1}} \omega \right\} = \int_{\partial c} \omega. \qquad (141)$$

Последнее равенство в соотношениях (141) имеет место согласно формуле (129). Тем самым формула Стокса доказана для сингулярного куба. После этого она тривиально переносится на произвольную цепь.

2. Из формулы Стокса следует утверждение: *равные цепи имеют равные границы.*

Доказательство. Пусть k-мерные цепи C_1, C_2 равны друг другу, пусть ω — произвольная форма степени k— 1. Так как $C_2 = C_1$, то

$$\int_{C_2} d\omega = \int_{C_1} d\omega.$$

Отсюда по формуле Стокса имеем

$$\int_{\partial C_2} \omega = \int_{\partial C_1} \omega.$$

Следовательно, $\partial C_2 = \partial C_1$.

7.3.5. Оператор проектирования

1. Пусть c: $h \to E$ — какой-нибудь k-мерный сингулярный куб в E. Будем рассматривать h как стандартный куб в R^k, считая при этом, что R^k есть координатная гиперплоскость в R^{k+1}. Произвольную точку в R^{k+1} запишем в виде $\tilde{t} = (t^1, ..., t^k, u)$, стандартный куб в R^{k+1} обозначим через $\tilde{h}$. В таком случае первоначально данный стандартный куб h в

R^k следует представлять себе в виде нижней грани куба $\tilde{h}$ (при u=0). Возьмем в евклидовом пространстве E произвольную точку O. По данному сингулярному k-мерному кубу c и по данной точке O определим некоторым специальным образом k+1-мерный сингулярный куб $\tilde{c}$. А именно, если для произвольной точки $t = (t^1, ..., t^k, 0) \in h$ дано $x = c(t) \in E$, то для произвольной точки $\tilde{t} = (t^1, ..., t^k, u)$ мы определим $\tilde{x} = \tilde{c}(\tilde{t})$ равенством

$$O\tilde{x} = (1 - u)\, Ox,$$

где Ox — вектор в E, который идет из точки O в точку x. Заметим, что при отображении $\tilde{c}$ вся верхняя грань куба $\tilde{h}$ (которая определяется равенством u=1), отображается в одну точку O.

Будем говорить, что k+1-мерный сингулярный куб $\tilde{c}$ проектирует из точки O данный k-мерный сингулярный куб c; будем обозначать $\tilde{c}$ также символом $O\times c$.

Пусть дана k-мерная цепь

$$C = \lambda_1 c_1 + \ldots + \lambda_p c_p. \tag{142}$$

Будем говорить, что k+1-мерная цепь

$$\tilde{C} = \lambda_1 \tilde{c}_1 + \ldots + \lambda_p \tilde{c}_p \tag{143}$$

проектирует из точки O данную k-мерную цепь C; будем обозначать $\tilde{C}$ также символом $O\times C$.

2. Имеет место следующая теорема:

Если

$$C_2 = C_1 \ (C_1,\ C_2 \in S^k),$$

то

$$O \times C_2 = O \times C_1.$$

Иначе говоря, равные цепи проектуються равными цепями.

Из определения $O\times C$ с помощью равенств (142) и (143) следует, что

$$O \times (C_1 + C_2) = O \times C_1 + O \times C_2 \tag{144}$$

и

$$O \times (\alpha C) = \alpha (O \times C), \qquad \alpha \in \mathbb{R}. \tag{145}$$

Ввиду этих соотношений сформулированная теорема равносильна следующему утверждению: если $C=\theta \in S^k$, то $O \times C = \theta \in S^{k+1}$, т.е. если C является k-мерной нулевой цепью, то $O\times C$ будет k+1-мерной нулевой цепью.

Замечание. Не следует думать, что последнее утверждение вытекает из равенства (145) при $\alpha=0$. На самом деле, чтобы воспользоваться равенством (145), нужно в его левой части произвольную k-мерную нулевую цепь θ заменить цепью $0\cdot C$, которая равна θ. Но то, что после этой замены левая часть (145) будет равна своему первоначальному значению, как раз и требуется доказать.

3. Доказательство теоремы будет проведено после некоторых предварительных конструкций. Прежде всего введем в E декартовы координаты с начальной точкой O. Тогда отображения c и $\tilde{c}$ представятся в координатах:

$$\left.\begin{array}{l} c: \begin{cases} x^1 = c^1(t^1, \ldots, t^k), \\ \cdots\cdots\cdots \\ x^n = c^n(t^1, \ldots, t^k), \end{cases} \\ \tilde{c}: \begin{cases} \tilde{x}^1 = (1-u)\, c^1(t^1, \ldots, t^n), \\ \cdots\cdots\cdots \\ \tilde{x}^n = (1-u)\, c^n(t^1, \ldots, t^n). \end{cases} \end{array}\right\} \qquad (146)$$

Рассмотрим произвольную форму степени $k + 1$:

$$\omega^{k+1} = *\sum \omega_{i_1 \ldots i_{k+1}}(x)\, dx^{i_1} \wedge \ldots \wedge dx^{i_{k+1}}.$$

Нам достаточно показать, что если C — нулевая k-мерная цепь и $\tilde{C} = O \times C$, то

$$\int\limits_{\tilde{C}} \omega^{k+1} = 0.$$

К этому и будут направлены наши выкладки. Чтобы упростить запись дальнейших соотношений, возьмем сначала в качестве ω^{k+1} форму одночленного вида

$$\omega^{k+1} = g(x)\, dx^{i_1} \wedge \ldots \wedge dx^{i_k} \wedge dx^{i_{k+1}}.$$

Здесь

$$g(x) = g(x^1, \ldots, x^n).$$

Положим

$$g(x, u) = g((1-u)\, x^1, \ldots, (1-u)\, x^n)$$

и рассмотрим форму $\omega^k(u)$ степени k, которая зависит от параметра u и определяется равенством

$$\omega^k(u) = g(x, u)\, dx^{i_1} \wedge \ldots \wedge dx^{i_k} \wedge e^{i_{k+1}}(x), \qquad (147)$$

где $e^{i_{k+1}}$ — известная нам базисная форма в касательном пространстве точки x. Запись $e^{i_{k+1}}$ вместо символа dx^{k+1} означает, что при подсчете внешнего произведения

$$dx^{i_1} \wedge \ldots \wedge dx^{i_k} \wedge e^{i_{k+1}}(x)$$

последний сомножитель берется на векторе, который равен радиус-вектору точки x. Полное объяснение этой записи дает формула (50), согласно которой

$$\begin{aligned} dx^{i_1} \wedge \ldots \wedge dx^{i_k} \wedge e^{i_{k+1}}(x) = \\ &= (-1)^k \{(dx^{i_2} \wedge dx^{i_3} \wedge \ldots \wedge dx^{i_{k+1}})\, x^{i_1} - \\ &- (dx^{i_1} \wedge dx^{i_3} \wedge \ldots \wedge dx^{i_{k+1}})\, x^{i_2} + \ldots \\ &\ldots + (-1)^k (dx^{i_1} \wedge dx^{i_2} \wedge \ldots \wedge dx^{i_k})\, x^{i_{k+1}}\}. \end{aligned} \tag{148}$$

Здесь $(\ldots, x^{i_1}, \ldots, x^{i_{k+1}}, \ldots)$ — координаты радиус-вектора точки x (те самые координаты, которые являются аргументами функции $g(x^1, \ldots, x^n)$). Введем для произвольного k-мерного сингулярной куба c функцию

$$f_c(u) = (\omega^k(u),\ c) = \int_c \omega^k(u). \tag{149}$$

Покажем, что имеет место равенство

$$(\omega^{k+1},\ O \times c) = -\int_0^1 (1-u)^k f_c(u)\, du. \tag{150}$$

Из него мы получим теорему, которая интересует нас.
Имеем

$$\begin{aligned} (\omega^{k+1},\ O \times c) &= \int_{O \times c} \omega^{k+1} = \int_{\tilde{h}} \tilde{c}^* \omega^{k+1} = \\ &= \int_{\tilde{h}} g(\tilde{c}(\tilde{t})) \det \left(\frac{\tilde{x}^{i_1}, \ldots, \tilde{x}^{i_k}, \tilde{x}^{i_{k+1}}}{t^1, \ldots, t^k, u} \right) dt^1 \ldots dt^k du. \end{aligned} \tag{151}$$

Из формул (146)

$$\det\left(\frac{\tilde{x}^{i_1}, \ldots, \tilde{x}^{i_k}, \tilde{x}^{i_{k+1}}}{t^1, \ldots, t^k, u}\right) =$$

$$= \begin{vmatrix} (1-u)\dfrac{\partial c^{i_1}}{\partial t^1} & \ldots & (1-u)\dfrac{\partial c^{i_1}}{\partial t^k} & (-1)\, c^{i_1} \\ (1-h)\dfrac{\partial c^{i_2}}{\partial t^1} & \ldots & (1-u)\dfrac{\partial c^{i_2}}{\partial t^k} & (-1)\, c^{i_2} \\ \cdot & \cdot & \cdot & \cdot \\ (1-u)\dfrac{\partial c^{i_{k+1}}}{\partial t^1} & \ldots & (1-u)\dfrac{\partial c^{i_{k+1}}}{\partial t^k} & (-1)\, c^{i_{k+1}} \end{vmatrix} =$$

$$= -(1-u)^k(-1)^k\left\{c^{i_1}\det\left(\frac{c^{i_2}, \ldots, c^{i_{k+1}}}{t^1, \ldots, t^k}\right) - \ldots \right.$$

$$\left. \ldots + (-1)^k c^{i_{k+1}}\det\left(\frac{c^{i_1}, \ldots, c^{i_k}}{t^1, \ldots, t^k}\right)\right\}. \tag{152}$$

Отсюда и из (151)

$$(\omega^{k+1}, O \times c) = -\int_0^1 (1-u)^k \times$$

$$\times\left[\int_C (-1)^k \{g(x, u)\, x^{i_1}(dx^{i_2} \wedge \ldots \wedge dx^{i_{k+1}}) - \ldots\}\right] du. \tag{153}$$

В правой части (153) мы написали внутри фигурных скобок только один член. Остальные легко получаются из (152); легко усмотреть, что внутри фигурных скобок находится правая часть (148), умноженная на $g(x, u)$. Таким образом, равенство (150) вытекает из (148) и (153).

Возвратимся к случаю, когда ω^{k+1} — произвольная форма степени k+1. Положим

$$\omega^k(u) =$$
$$= * \sum \omega_{i_1 \ldots i_{k+1}}(x, u)\, dx^{i_1} \wedge \ldots \wedge dx^{i_k} \wedge e^{i_{k+1}}(x), \tag{154}$$

где

$$\omega_{i_1 \ldots i_{k+1}}(x, u) = \omega_{i_1 \ldots i_{k+1}}((1-u)\, x)$$

и снова определим $f_c(u)$ по формуле (149), беря теперь в качестве $\omega^k(u)$ выражение (154). Формула (150) сохранит силу. Это ясно, поскольку левая часть (150) линейна относительно ω^{k+1}, а правая - относительно $\omega^k(u)$. Рассмотрим цепь

$$C = \lambda_1 c_1 + \ldots + \lambda_p c_p.$$

Введем функции

$$\left.\begin{aligned} f_{c_j}(u) &= \int_{c_j} \omega^k(u), \\ f(u) &= \lambda_1 f_{c_1}(u) + \ldots + \lambda_p f_{c_p}(u) = \\ &= \int_C \omega^k(u) = (\omega^k(u),\ C). \end{aligned}\right\} \tag{155}$$

Тогда из (144), (145) и (150) имеем

$$(\omega^{k+1}, O \times C) = \lambda_1 (\omega^{k+1}, O \times c_1) + \ldots + \lambda_p (\omega^{k+1}, O \times c_p) =$$
$$= -\lambda_1 \int_0^1 (1-u)^k f_{c_1}(u)\,du - \ldots - \lambda_p \int_0^1 (1-u)^k f_{c_p}(u)\,du =$$
$$= -\int_0^1 (1-u)^k f(u)\,du. \tag{156}$$

4. Докажем теорему п. 2 во второй формулировке.

Пусть C — нулевая k-мерная цепь. Возьмем произвольно ω^{k+1}. Из (155) имеем $f(u)=0$ при любому u, $0 \le u \le 1$ (поскольку C — нулевая). Отсюда и из (156) имеем

$$(\omega^{k+1}, \quad O \times C) = 0.$$

Следовательно, $O \times C$ — нулевая $k+1$-мерная цепь. Теорема доказана.

5. Положим

$$\Pi(C) = (-1)^{k+1}(O \times C). \tag{157}$$

Из равенств (144), (145) и из только что доказанной теоремы следует, что $\Pi(C)$ есть линейный оператор, определенный на все k-мерных цепях, зависящий от выбора точки O. Его значением является $k+1$-мерная цепь, которая лишь знаком отличается от проектируемой цепи $O \times C$ Таким образом,

$$\Pi:\ S^k \to S^{k+1}.$$

Заметим, что S^k и S^{k+1} по точному смыслу своего определения имеют в качестве элементов не цепи, а классы их эквивалентности. Поэтому доказательство теоремы п. 2 играет фундаментальную роль для определения оператора П. Только на основании этой теоремы П определен как оператор, который отображает S^k в S^{k+1}. Мы назовем линейный оператор П *оператором проектирования.*

6. Имеет место следующее тождество:

$$C = \partial\Pi(C) + \Pi(\partial C), \tag{158}$$

где ∂ - оператор, который дает границу цепи.

Доказательство с очевидностью усматривается для одного сингулярного куба *с.* В самом деле, применяя обозначения п. 7.13.3, получаем по формуле (129):

$$\partial(O \times c) = \sum_{i=1}^{k+1} \{(-1)^i \tilde{c}_{i,0} + (-1)^{i+1} \tilde{c}_{i,1}\}, \tag{159}$$

где при i= 1, 2, ..., k

$$\tilde{c}_{i,0} = O \times c_{i,0}, \quad \tilde{c}_{i,1} = O \times c_{i,1}.$$

При $i=k+1$ имеем $\tilde{c}_{k+1,0}=c$ (т.е. первоначально данный k-мерный сингулярный куб), а $\tilde{c}_{k+1,1}$ представляет собой нулевой k -мерный куб. Таким образом, из формулы (159) с учетом (144) и (145) следует, что

$$\partial(O \times c) = \\ = \sum_{i=1}^{k} \{(-1)^i (O \times c_{i,0}) + (-1)^{i+1} (O \times c_{i,1})\} + (-1)^{k+1} c = \\ = O \times \partial c + (-1)^{k+1} c. \tag{160}$$

Теперь, если дана произвольная цепь

$$C = \lambda_1 c_1 + \ldots + \lambda_p c_p,$$

то из (160) стандартным образом получаем для цепи

$$\partial(O \times C) = O \times \partial C + (-1)^{k+1} C. \tag{161}$$

Поскольку

$$\partial\Pi(C) = (-1)^{k+1} \partial(O \times C), \quad \Pi(\partial C) = (-1)^k \{O \times \partial C\},$$

то (158) доказано как следствие (161) и (157).

7. Теперь мы введем в рассмотрение некоторый линейный оператор

$$I\colon W^{k+1} \to W^k,$$

который во многих отношениях аналогичен оператору П.

А именно, пусть во всем пространстве R^n дана форма $\omega \in W^{k+1}$. Как и раньше, рассмотрим зависимую от параметра *и* форму $\omega^k(u) \in W^k$, определяемую равенством (154). Оператор I задается формулой

$$I(\omega) = (-1)^k \int_0^1 (1-u)^k \omega^k(u)\, du. \tag{162}$$

Согласно этому определению $I(\omega)$ есть внешняя форма в R^n степени k ($I(\omega) \in W^k$).

Замечание. Для определения $I(\omega)$ нет необходимости иметь ω во всем просторанстве R^n; достаточно, чтобы форма ω была задана в какой-нибудь области, звездной относительно точки *O* (поскольку $\omega^k(u)$ определяется по лучам, которые выходят из точки *O*). В такой области будут верны и дальнейшие выводы.

8. **Теорема**. *Операторы* Π *и I являются сопряженными.*

В самом деле, имеем для цепи $C\in S^k$ и для $\omega\in W^{k+1}$

$$(I(\omega),\ C)=\int\limits_C I(\omega)=(-1)^k\int\limits_0^1\left\{(1-u)^k\int\limits_C\omega^k(u)\right\}du=$$
$$=(-1)^k\int\limits_0^1(1-u)^k f(u)\,du. \tag{163}$$

Но по формуле (150)

$$(\omega,\ O\times C)=-\int\limits_0^1(1-u)^k f(u)\,du.$$

Отсюда

$$(I(\omega),\ C)=(-1)^{k+1}(\omega,\ O\times C). \tag{164}$$

Из (163) и (157) имеем для любых $\omega\in W^{k+1},\ C\in S^k$

$$(I(\omega),\ C)=(\omega,\ \Pi(C)), \tag{165}$$

что и означает сопряженность операторов

$$\Pi:\ S^k\to S^{k+1},\qquad I:\ W^{k+1}\to W^k.$$

9. Вследствие (165) следующие два тождества равносильны

$$C=\partial\Pi(C)+\Pi(\partial C), \tag{166}$$

$$\omega=dI(\omega)+I(d\omega). \tag{167}$$

Так как тождество (166) уже доказано, выведем из него тождество (167).

Доказательство тождества (167). Имеем согласно (165) для любой цепи $C\in S^{k+1}$ и для любой формы $\omega\in W^{k+1}$

$$(I(d\omega),\ C)=(d\omega,\ \Pi(C)).$$

Отсюда и по формуле Стокса

$$(I(d\omega),\ C)=(\omega,\ \partial\Pi(C)). \tag{168}$$

Далее

$$(dI(\omega),\ C)=(I(\omega),\ \partial C)=(\omega,\ \Pi(\partial C)). \tag{169}$$

Из (168) и (169)

$$(\omega-dI(\omega)-I(d\omega),\ C)=(\omega,\ C-\Pi(\partial C)-\partial\Pi(C)).$$

Правая часть этого равенства равна нулю в силу (166). Отсюда и вследствие произвольности $C \in S^{k+1}$ получаем (167).

7.3.6. Теорема Пуанкаре и некоторые другие предложения

1. **Теорема**. *Для любой цепи* $C \in S^k$ *граница границы есть нулевая цепь:*

$$\partial \partial C = \theta \in S^{k-2}.$$

Доказательство. Возьмем произвольно $\omega \in W^{k-2}$,

Имеем по формуле Стокса

$$(\omega, \partial \partial C) = (d\omega, \partial C) = (d\, d\omega, C) = 0,$$

поскольку $dd\,\omega = 0$. Теорема доказана.

Замечание. В случае трехмерного стандартного куба эта теорема иллюстрируется рис.2.

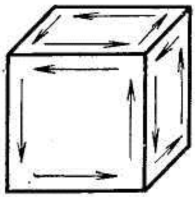


Рис.2.

2. ***Определение***. Цепь $C \in S^k$ называется *замкнутой цепью* или *циклом,* если $\partial C = \theta$ (θ — нулевая цепь в S^{k+1}).

Теорема. *Каждая граница есть цикл.*

Доказательство. Пусть $C' = \partial C$. Тогда

$$\partial C' = \partial \partial C = \theta.$$

3. ***Определение***. Цикл называется *гомологичным нулю,* если он является границей.

Теорема. *В евклидовом пространстве каждый цикл гомологичен нулю.*

Доказательство. Пусть $\partial C = \theta$. Тогда из формулы (166) имеем

$$C = \partial \Pi\,(C),$$

т.е. цикл C гомологичен нулю, поскольку он является границей цепи $\Pi(C)$.

4. ***Определение***. Форма $\omega \in W^k$ называется *замкнутой формой,* или *коциклом*, если $d\omega = 0$.

Определение. Форма $\omega \in W^k$ называется *точной формой,* или *формой, когомологичной нулю,* если она является внешним дифференциалом некоторой формы.

Теорема (Пуанкаре). *В евклидовом пространстве всякая замкнутая форма есть точной.*

Доказательство. Пусть $d\omega=0$. Тогда из формулы (167) имеем $\omega=d\mu$, где $\mu= I(\omega)$.

Замечание. Можно сказать, что в евклидовом пространстве замкнутая форма $\omega \in W^k$ имеет потенциал

$$\mu = I\ (\omega) \in W^{k-1}.$$

Замечание. Мы имеем сопряженные отображения:

$$\partial\colon S^{k+1} \to S^k, \qquad d\colon W^k \to W^{k+1},$$

Можно, однако, написать более содержательные соотношения:

$$\partial\colon S^{k+1} \to C(S^k), \tag{170}$$

$$d\colon W^k \to C(W^{k+1}), \tag{171}$$

где $C(S^k)$ — линейное подпространство циклов в S^k, $C(W^{k+1})$ — линейное подпространство замкнутых форм в W^{k+1}. Теорема Пуанкаре и предыдущая ей теорема утверждают, что оба отображения в евклидовом просторные сюръективны (являются отображениями «на»). Утверждение Пуанкаре и приведенное доказательство остаются в силе, если объекты даны в звездной области евклидова пространства (не обязательно во всем евклидовом пространстве).

5. **Теорема**. *В евклидовом пространстве интеграл от замкнутой формы степени k по любому k-мерному циклу равен нулю.*

Доказательство. Пусть $\omega \in W^k$, $C\in\ (S^k)$ и $d\omega= 0$, $\partial C = \theta$. Так как $\partial C=\theta$, то по теореме п. 3 найдется цепь $\tilde{C}$ такая, что $C=\partial \tilde{C}$. Тогда по теореме Стокса имеем

$$(\omega, C) = (\omega, \partial) \tilde{C} = (d\omega, \tilde{C}) = 0,$$

поскольку $d\omega= 0$.

Замечание. доказательство можно провести также с помощью теоремы Пуанкаре. А именно, так как $d\omega =0$, то найдется форма $\tilde{\omega}$ такая, что $\omega = d\tilde{\omega}$. Тогда

$$(\omega,C) = (d\tilde{\omega}\ ,C) = (\tilde{\omega}\ , \partial C) = 0,$$

поскольку $\partial C = \theta$.

7.3.7. Регулярное погружение. Комбинаторная поверхность

1. *Регулярным погружением k-мерного* стандартного куба h ($h \subset \mathrm{R}^k$) в пространство R^n, $n \geq k$, называется отображение

$$\varphi: h \to \mathrm{R}^n$$

при условии, что производная отображения φ во всех точках куба h имеет ранг $=k$. Регулярное погружение $\varphi(h \to \mathrm{R}^n)$ мы будем называть также *регулярным кубом* в R^n.

Очевидно, что регулярный куб является частным случаем сингулярного. Поэтому к регулярным кубам применимы все понятия и утверждения, высказанные в предыдущих параграфах.

2. Пусть дано гладкое отображение $\varphi: h \to \mathrm{R}^n$. Обозначим через H образ куба h в R^n. Пусть t — произвольная точка куба h, $x = \varphi(t)$ — ее образ. Не исключено, что имеется еще точка $t' \in h$ отличная от точки t, образ которой также есть точка x (не исключено, что имеется даже бесконечное множество таких точек). Иначе говоря, не исключено, что отображение куба h на его образ H не является взаимно-однозначным. Это обстоятельство в равной мере относится и к случаю, когда $\varphi: h \to \mathrm{R}^n$ есть сингулярный куб, и к случаю регулярного куба. Ввиду этого обстоятельства ни сингулярный, ни регулярный куб нельзя определять как образ H ($H = \varphi(h)$); и тот и другой определяются как отображение h на H (с последующим установлением условий, при которых два отображения $h \to H$ и $h_1 \to H$ считаются тем самым сингулярным или тем самым регулярным кубом).

Предположим, что отображение $\varphi: h \to \mathrm{R}^n$ является регулярным кубом. В этом случае любая точка $t \in h$ имеет окрестность U_t, которая отображается на свой образ в R^n взаимно-однозначно. Таким образом, куб h локально гомеоморфно отображается на свой образ H.

3. При $k=1$ регулярный или сингулярный куб называют *ориентированной дугой* (т.е. направленной дугой, соответственно регулярной или сингулярной).

Ориентированная дуга c есть гладкое отображение в R^n ориентированного отрезка h. Если дуга c регулярна, то отображение локально гомеоморфно и образ h не имеет локальных особенностей (точек возврата, угловых точек и т.п.). Однако в целом отображение может не быть гомеоморфным, т.е. для образа возможны самопересечения. Например, возможна картина, которая изображена на рис.3.

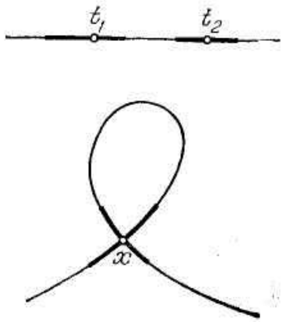


Рис.3.

Здесь t_1 и t_2 — две точки отрезка *h*, которые отображаются в одну точку *x;* вместе с тем каждая из точек t_1, t_2 имеет окрестность, которая отображается на свой образ взаимно однозначно. На рис. 3 эти окраины в их образы изображены жирными линиями.

4. Важный пример сингулярного куба при k=2, n=2 дает отображение

$$x = ar\cos b\theta, \quad y = ar\sin b\theta,$$

считая, что куб (квадрат) *h* задан в осях Or, $O\theta$ неравенствами $0 \leq r \leq 1$, $0 \leq \theta \leq 1$ (рис. 4).

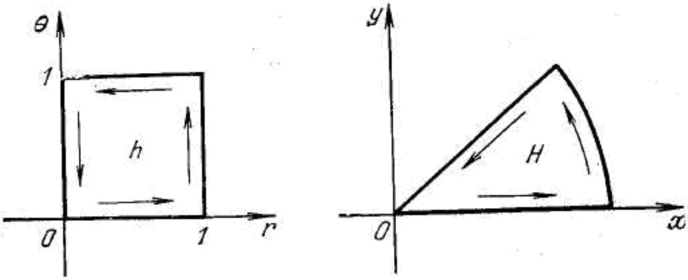


Рис. 4.

Здесь мы имеем именно сингулярный куб, поскольку вся сторона r = 0 отображается в одну точку (0, 0) плоскости (*x, y).* Образом *h* в плоскости *(x, y)* является круговой сектор *H;* внутренние стрелки на *h* и на *H* указывают ориентацию границы рассматриваемого сингулярного куба (квадрата, поскольку k = 2).

Заметим, что получить круговой сектор как образ регулярного куба невозможно. Тем самым уже этот пример поясняет целесообразность рассмотрения сингулярных кубов.

5. Мы подробно изложили интегрирование формы по цепи. В задачах анализа, однако, интегрируют также по кривой или по поверхности (что можно не различать, имея в виду k-мерную поверхность). Для конкретных задач анализа (и его приложений) может быть достаточным понятие, которое мы назовем *комбинаторной поверхностью.* Ее можно рассматривать как частный случай цепи с коэффициентами ±1, сингулярные кубы которой не как угодно набросаны в пространстве, а приложенные друг к другу с соблюдением некоторых условий. Мы не станем перечислять эти условия и ограничимся отсылкой читателя к рис.5.

Рис.5.

На рис. 5 изображена двумерная поверхность с краем; будем ее представлять себе как некоторый сферический сегмент H. Он составлен из шести сферических треугольников, которые мы представим себе в качестве образов шести двумерных сингулярных кубов: $c_1, c_2, c_3, c_4, c_5, c_6$ (см. п. 4). Стрелки показывают ориентацию их границ. Для каждого c_j граница состоит из двух внутренних одномерных граней (к каждой из которых примыкает соседний сингулярный куб) и одной краевой одномерной грани (вдоль которой соседнего куба нет). Рассмотрим цепь

$$C = \lambda_1 c_1 + \ldots + \lambda_6 c_6,$$

считая $\lambda_j=\pm1$. Существенно следующее обстоятельство: судя по рисунку, в данном случае λ_j можно выбрать так, что все внутренние одномерные грани попарно уничтожатся. Именно так будет, например, при $\lambda_j=1$. Тогда граница C будет одномерной цепью

$$\partial C=\lambda_1\partial c_1+ \ldots + \lambda_6\partial c_6,$$

образ которой есть край H (с заданной ориентацией, которая показана стрелками). В таком случае мы скажем, что цепь C есть комбинаторная поверхность, которая представляет сегмент $H.$ Если в пространстве дана 2-форма ω, то в качестве интеграла от ω по поверхности H

примем интеграл от ω по представляющей H комбинаторной поверхности, т.е. по цепи C.

Однако ясно, что рядом с $\lambda_j = +1$ в равной мере пригодны $\lambda_j = -1$ (при всех j). Таким образом, H представляется также комбинаторной поверхностью — C. Таким образом, интеграл от ω по H определен с точностью до знака. Выбор C или — C есть выбор ориентации H; после этого выбора интеграл определен вполне, поскольку при любом другом выборе $\lambda_j=\pm1$ не будет происходить попарного уничтожения всех внутренних одномерных граней кубов c_j. Следовательно, при данном наборе $c_1, \dots ,c_6$ существуют только две комбинаторные поверхности, которые представляют сегмент H: C и $-C$. К сожалению, так обстоит дело лишь для заданного набора сингулярных кубов. Возможно безконечное множество других наборов других сингулярных кубов в любом числе, с помощью которых H также можно представить в виде комбинаторной поверхности. Можно обойтись, например, всего одним сингулярным кубом. Но доказать, что во всех этих случаях мы будем получать с точностью до знака одно и то же значение интеграла формы ω, тяжело даже в частном примере сферического сегмента H. Для общего изложения теории интеграла формы по поверхности проще оказывается другой путь, который идет через понятие многообразия.

6. В заключение еще несколько слов скажем по поводу одного класса в некотором смысле элементарных поверхностей. Обычно именно их имеют в виду в элементарном анализе, когда рассматривают интеграл по поверхности.

Пусть $H=c(h)$ — образ в n-мерном евклидовом простоpанстве E некоторой области $h\subset R^k$, $k\leq n$, при отображении c: $\mathrm{R}^k\to E$. Обозначая здесь и далее той же буквой c сужение c на h, предположим, что отображение c: $h\to H$ взаимнооднозначно. При этом условии мы можем H назвать k -мерной поверхностью в E. Заметим, что на этот раз буква h обозначает не стандартный куб, а произвольную область в R^k. Тому клас поверхностей, о котором идет речь, довольно широкий (но не исчерпывает всего, что называют k -мерными поверхностями в E).

Пусть в E в окрестности H задана внешняя форма ω степени k. Для простоты записи будем считать, что ω имеет в E одночленное координатное представление с коэффициентом G. Тогда мы можем определить интеграл от ω по H так же, как в свое время определяли интеграл по сингулярному кубу. А именно,

$$\int\limits_H \omega = \int\limits_h c^*\omega = \int\limits_h (G\circ c)\det \hat{c}'. \qquad (172)$$

Мы сохранили обозначения, которые применялись в п. 8 п. 7.13.1, (см. формулу (123)); изменен только смысл *h*. Справа в (172) *h* обозначает область *k*-кратного интегрирования в R^k.

Естественно поставить вопрос об инвариантности этого определения: если $H = c_1(h_1)$, где c_1, h_1 - другое отображение и другая область, подчиненные тем же условиям, что *c* и *h*, будет ли

$$\int_{h_1} c_1^* \omega = \int_h c^* \omega \tag{173}$$

верным равенством? На этот вопрос можно ответить утвердительно, если c_1 и c имеют одинаковую ориентацию относительно R^k, а именно, если композиция $c^{-1}{}_1 \circ c$ имеет производную с положительным определителем. (Напомним, что c и c_1 обозначают отображения, суженные на h и h_1) В этом случае

$$\int_{h_1} c_1^* \omega = \int_{h_1} (G \circ c_1) \det \hat{c}_1' =$$

$$= \int_h (((G \circ c_1) \det \hat{c}_1') \circ (c_1^{-1} \circ c)) ((\det \hat{c}_1')^{-1} \circ (c_1^{-1} \circ c)) \det \hat{c}' =$$

$$= \int_h (G \circ c) \det \hat{c}' = \int_h c^* \omega.$$

Тем самым равенство (173) доказано.

Если пренебречь условием относительно ориентации, то интеграл будет определен с точностью до знака.

8. Гильбертово пространство

Ги́льбертово простра́нство — обобщение евклидова пространства, допускающее бесконечную размерность. Названо в честь Давида Гильберта.

Важнейшим объектом исследования в гильбертовом пространстве являются линейные операторы. Само понятие гильбертова пространства сформировалось в работах Гильберта и Шмидта по теории интегральных уравнений, а абстрактное определение было дано в работах фон Неймана, Риса и Стоуна по теории эрмитовых операторов.

Ги́льбертово простра́нство - *векторное пространство H* над полем комплексных (или действительных) чисел вместе с комплексной (действительной) функцией (*x, y*), определенной на $H \times H$ и обладающей следующими свойствами.

1) $(x, x)=0$ в том и только в том случае, если $x=0$;
2) $(x, x)\geqslant 0$ для всех $x \in H$;
3) $(x+y, z)=(x, z)+(y, z)$, $x, y, z \in H$;
4) $(\alpha x, y)=\alpha(x, y)$, $x, y \in H$, α — комплексное (действительное) число;
5) $(x, y)=\overline{(y, x)}$, $x, y \in H$;
6) если $x_n \in H$, $n=1, 2, \ldots$ и если

$$\lim_{n, m \to \infty} (x_n - x_m, x_n - x_m) = 0,$$

то существует такой элемент $x \in H$, что

$$\lim_{n \to \infty} (x - x_n, x - x_n) = 0;$$

элемент *x* называется пределом последовательности (x_n);

7) *H* - бесконечномерное векторное пространство.

Функция (x, y), удовлетворяющая аксиомам 1) - 5), называется скалярным произведением, или внутренним произведением, элементов *x* и *y*. Величина $\|x\|=(x, x)^{1/2}$ называется нормой (или длиной) элемента $x \in H$. Имеет место неравенство $|(x, y)| \leqslant \|x\| \cdot \|y\|$. Если ввести в H расстояние между элементами $x, y, \in H$ при помощи равенства $\rho(x, y)=\|x-y\|$, то H превращается в *метрическое пространство*.

Два ги́льбертовых простра́нства H и H_1 называются изоморфными (или изометрически изоморфными), если существует взаимно однозначное соответствие $x \leftrightarrow x_1$, $x \in H$, $x_1 \in H_1$, между H и H_1, сохраняющее линейные операции и скалярное произведение.

Ги́льбертово простра́нство составляют наиболее распространенный и важный для приложений класс бесконечномерных векторных пространств. Они представляют собой естественное обобщение понятия конечномерного векторного пространства со скалярным произведением (т. е. конечномерного евклидова пространства, или конечномерного унитарного пространства). Именно, если в конечномерном векторном пространстве (над полем действительных или комплексных чисел) задано скалярное произведение, то свойство 6), называеиое полнотой Ги́льбертова простра́нства , выполняется автоматически. Бесконечномерные векторные пространства *H* со скалярным произведением называются предгильбертовыми пространствами; существуют предгильбертовы пространства, в которых свойство 6) не выполняется. Всякое предгильбертово пространство может быть дополнено до Ги́льбертова простра́нства .

Иногда в определение Ги́льбертова простра́нства не включается условие бесконечномерности, т. е. предгильбертовым пространством называется векторное пространство над полем комплексных (или действительных) чисел со скалярным произведением, а Ги́льбертовым простра́нством называется полное предгильбертово пространство.

Примеры Ги́льбертова простра́нства.

1) Комплексное пространство l^2 (или l_2). Элементами этого Ги́льбертова простра́нства являются бесконечные последовательности комплексных чисел $x = \{\xi_1, \ \xi_2, \dots\}$, $y = \{\eta_1, \ \eta_2, \dots\}$ со сходящейся суммой квадратов модулей:

$$\sum_{k=1}^{\infty} |\xi_k|^2 < +\infty, \quad \sum_{k=1}^{\infty} |\eta_k|^2 < +\infty;$$

скалярное произведение определяется равенством

$$(x, \ y) = \sum_{k=1}^{\infty} \xi_k \overline{\eta_k}.$$

2) Пространство $l^2(T)$ (обобщение примера 1)). Пусть *T* - произвольное множество. Элементами Ги́льбертова простра́нства $l^2(T)$ являются комплекснозначные функции *x(t)* на *T,* отличные от нуля не более чем в счетном множестве точек $t \in T$ и такие, что ряд

$$\sum_{t \in T} |x(t)|^2$$

сходится. Скалярное произведение определяется равенством

$$(x,\ y) = \sum_{t \in T} x(t)\overline{y(t)}.$$

Всякое Ги́льбертово простра́нство изоморфно пространству $l^2(T)$ для некоторого соответствующим образом подобранного *T.*

3) Пространство $L^2(S,\ \Sigma,\ \mu)$ (или $L_2(s,\ \Sigma,\ \mu)$) комплекснозначных функций $x(s)$, определенных на множестве Sc вполне аддитивной положительной мерой μ (заданной на σ-алгебре Σ-подмножеств множества S), измеримых и имеющих интегрируемый квадрат модуля:

$$\int_S |x(s)|^2 d\mu(s) < +\infty.$$

В этом Ги́льбертовом простра́нстве скалярное произведение определяется равенством

$$(x(s),\ y(s)) = \int_S x(s)\overline{y(s)}\, d\mu(s).$$

4) *Соболева пространство* $W_l^2(\Omega)$, обозначаемое также $H_{(l)}$ (см. *Вложения теоремы*).

5) Ги́льбертово простра́нство функций со значениями в Ги́льбертовом простра́нстве. Пусть *H*- некоторое Ги́льбертово простра́нство со скалярным произведением (*x, y*), $x,\ y \in H$. Пусть, далее, Ω- произвольная область в R_n, а $f(x)$, $x \in \Omega$ - функция с ее значениями в *H,* измеримая в смысле Бохнера (см. *Бохнера интеграл*).и такая, что

$$\int_\Omega \| f(x) \|_H^2 dx < \infty,$$

где *dx*-мера Лебега на Ω(вместо меры Лебега можно взять любую другую положительную счетно аддитивную меру). Если на этом множестве функций определить скалярное произведение

$$(f(x),\ g(x))_\perp = \int_\Omega (f(x),\ g(x))\, dx,$$

то получится новое Ги́льбертово простра́нство H_1.

6) Множество непрерывных *Бора почти периодических функций* на прямой образует предгильбертово пространство, если скалярное произведение определяется равенством

$$(x(t),\ y(t)) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t)\overline{y(t)}\, dt.$$

Существование предела вытекает из теории почти периодических функций. Это пространство пополняется *Безиковича почти периодическими функциями* класса B^2.

Пространства l^2 и L^2 были введены и изучены Д. Гильбертом в основополагающих работах по теории интегральных уравнений и бесконечных квадратичных форм. Определение Ги́льбертова простра́нства было дано Дж. Нейманом, Ф. Риссом и М. Стоуном, которые положили также начало его систематическому изучению.

Ги́льбертово простра́нство является естественным обобщением обычного трехмерного пространства евклидовой геометрии, и многие геометрические понятия имеют интерпретацию в Ги́льбертовом простра́нстве, что позволяет говорить о геометрии Ги́льбертова простра́нства . Два вектора *x, y* из Ги́льбертова простра́нства *H* называются ортогональными $(x \perp y)$, если $(x,\ y)=0.$ Два линейных многообразия $\mathfrak{M}$ и $\mathfrak{N}$ из *H* называются ортогональными $(\mathfrak{M} \perp \mathfrak{N})$, если каждый элемент из $\mathfrak{M}$ ортогонален каждому элементу из $\mathfrak{N}$.Ортогональным дополнением множества $A \subset H$ называется множество $B=\{x|\ (x,\ A)=0\},$ т. е. множество элементов $x \in H$, ортогональных ко всем элементам из *A*. Оно обозначается $H \ominus A$ или, если *H* подразумевается, $A^{\perp}$. Ортогональное дополнение $\mathfrak{N}$ произвольного множества $\mathfrak{M}$ из *H* есть замкнутое линейное многообразие. Если $\mathfrak{M}$ - замкнутое линейное многообразие в Ги́льбертовом простра́нстве (называемое также подпространством), то всякий элемент $x \in H$ единственным образом может быть представлен в виде суммы $x=y+z,\ y \in \mathfrak{M},\ z \in \mathfrak{N}$. Это разложение называется теоремой об ортогональном дополнении и записывается обычно в виде

$$H=\mathfrak{M} \oplus \mathfrak{N}.$$

При этом теорема справедлива также в случае, если *H* есть предгильбертово пространство, а $\mathfrak{M}$ - замкнутое линейное многообразие в H. В связи с этим уместно отметить, что некоторые другие утверждения теории Ги́льбертова простра́нства справедливы полностью или частично и в предгильбертовых пространствах. Однако практически это обстоятельство не очень существенно, ибо встречающиеся в приложениях пространства либо полны, либо известно, как их пополнить.

Множество $A \subset H$ называется ортонормированным множеством, или ортонормированной системой, если любые различные два вектора из *A* ортогональны и если норма каждого вектора $y \in A$ равна единице.

Ортонормированное множество называется полным ортонормированным множеством, если не существует ненулевого вектора из *H*, ортогонального ко всем векторам этого множества. Если $\{y_i\}$ - ортонормированная последовательность, а $\{\alpha_i\}$ - последовательность $\sum_i \alpha_i y_i$ скаляров, то ряд сходится в том и только в том случае, когда

$$\sum_i |\alpha_i|^2 < \infty;$$

при этом

$$\left\| \sum_i \alpha_i y_i \right\|^2 = \sum_i |\alpha_i|^2$$

(теорема Пифагора в Ги́льбертовом простра́нстве).

Пусть *A* - ортонормированное множество в Ги́льбертовом простра́нстве *H,* а *x* - произвольный вектор из *H.* Тогда $(x, y) = 0$ для всех $y \in A$, за исключением конечного или счетного множества векторов. Ряд

$$Px = \sum_{y \in A} (x, y)\, y$$

сходится, и его сумма не зависит от порядка расположения его ненулевых членов. Оператор *P* является оператором ортогонального проектирования, или проектором, на замкнутое линейное многообразие, порождаемое множеством *A.* Множество $A \subset H$ называется ортонормированным базисом линейного многообразия

$\mathfrak{N} \subseteq H$, если *A* содержится в $\mathfrak{N}$ и если для любого $x \in \mathfrak{N}$ имеет место

$$x = \sum_{y \in A} (x, y)\, y,$$

т. е. любой вектор $x \in \mathfrak{N}$ разлагается по системе *A,* или может быть представлен при помощи векторов системы .4. Набор чисел $\{(x, y) | y \in A\}$ называется набором коэффициентов Фурье элемента *x* по базису *A.* Каждое подпространство Ги́льбертова простра́нства *H* (в частности, само *H*), имеет ортонормированный базис.

В $l^2(T)$ ортонормированным базисом является набор функций $\{x_t, t \in T\}$, определяемых формулой $x_t(s) = 1$ при $s = t$, $x_t(s) = 0$ при $s \neq t$. В пространстве $L^2(s, \Sigma, \mu)$ разложение вектора по базису принимает вид разложения функции по системе ортогональных функций - важный метод решения задач математической физики.

Для ортонормпрованного множества $A \subset H$ следующие утверждения эквивалентны: *A* полно; *A* является ортонормированным базисом для *H*; $\|x\|^2 = \Sigma_{y \in A} |(x, y)|^2$ для любого $x \in H$.

Все ортонормированные базисы данного Ги́льбертова простра́нства имеют одну и ту же мощность. Этот факт позволяет определить размерность Ги́льбертова простра́нства . Именно, размерностью Ги́льбертово простра́нство называется мощность произвольного ортонормированного базиса в нем. Иногда эта размерность называется гильбертовой размерностью (в отличие от линейной размерности Ги́льбертова простра́нства , т. е. мощности базиса Гамеля (Хамеля) - понятия, не учитывающего топологическую структуру Ги́льбертова простра́нства). Два Ги́льбертова простра́нства изоморфны в том и только в том случае, когда они имеют одну и ту же размерность. С понятием размерности связано понятие дефекта, или коразмерности, подпространства. Именно, дефектом подпространства H_1 Ги́льбертова простра́нства *H* называется размерность ортогонального дополнения $H_1^{\perp} = H \ominus H_1$. Подпространство, дефект которого равен 1, т. е. ортогональное дополнение которого одномерно, называется гиперпространством. Параллельное ему плоское множество называется гиперплоскостью.

Некоторые из геометрических понятий требуют использования терминологии линейных операторов в Ги́льбертовом простра́нстве; к ним относится, в частности, понятие раствора линейных многообразий. Раствором многообразий M_1 и M_2 в Ги́льбертовом простра́нстве *H* называется норма $\theta(M_1, M_2)$ разности операторов, проектирующих *H* на замыкание этих линейных многообразий.

Простейшие свойства раствора:

а) $\theta(M_1, M_2) = \theta(\bar{M}_1, \bar{M}_2) = \theta(H \ominus \bar{M}_1, H \ominus \bar{M}_2)$;
б) $\theta(M_1, M_2) \leqslant 1$,

причем в случае строгого неравенства $\dim M_1 = \dim M_2$.

Во многих задачах, относящихся к Ги́льбертовым простра́нствам, участвуют лишь конечные наборы векторов Ги́льбертова простра́нства, т. е. элементы конечных линейных многообразий Ги́льбертова простра́нства. Поэтому понятия и методы линейной алгебры играют в теории Ги́льбертова простра́нства большую роль. Векторы $g_1, g_2, \dots, g_n$ в Ги́льбертовои простра́нстве называются линейно независимыми, если равенство

$$\sum_{k=1}^{n} \alpha_k g_k = 0,$$

где α_k - скаляры, возможно лишь в том случае, когда все α_k равны нулю. Для линейной независимости векторов необходимо и достаточно, чтобы их *Грама определитель* был отличен от нуля.

Счетная последовательность векторов $g_1, \ldots, g_n, \ldots$ называется линейно независимой последовательностью, если линейно независима каждая ее конечная часть. Каждая линейно независимая последовательность может быть ортогонализирована, т. е. может быть построена такая ортонормированная система $e_1, e_2, \ldots,$ что для каждого *n линейные оболочки* множеств $\{g_k\}_{k=1}^n$ и $\{e_k\}_{k=1}^n$ совпадают. Это построение называется процессом ортогонализации (ортонормализации) Грама - Шмидта и осуществляется следующим образом:

$$e_1 = \frac{g_1}{\|g_1\|}, \quad h_2 = g_2 - (g_2, e_1)\, e_1, \quad e_2 = \frac{h_2}{\|h_2\|}, \ \ldots,$$

$$h_n = g_n - \sum_{k=1}^{n1-} (g_n, e_k), \ e_k, \ e_n = \frac{h_n}{\|h_n\|}, \ \ldots .$$

В множестве Ги́льбертова простра́нства определены операции прямой суммы и тензорного произведения Ги́льбертова простра́нства. Прямой суммой Ги́льбертова простра́нства $H_i, \ i = 1, 2, \ldots, n,$ где каждое H_i обладает соответствующим скалярным произведением, называется Ги́льбертовым простра́нством

$$H = H_1 \oplus H_2 \oplus \ldots \oplus H_n,$$

определяемое следующим образом: в векторном пространстве $H_1 \dot{+} H_2 \dot{+} \ldots \dot{+} H_n$ - прямой сумме векторных пространств $H_1, \ldots, H_n$ - задается скалярное произведение равенством

$$([x_1, \ \ldots, \ x_n], \ [y_1, \ \ldots, \ y_n]) = \sum_{i=1}^n (x_i, y_i)_{H_i}.$$

При $i \neq j$ элементы из H_i и H_j в прямой сумме

$$H = \sum_{i=1}^{n} \oplus H_i$$

взаимно ортогональны, и проектирование H на H_i совпадает с ортогональным проектированием H на H_i. Понятие прямой суммы Ги́льбертова простра́нства обобщается на случай бесконечного множества прямых слагаемых. Пусть для каждого *v* из некоторого множества *A* индексов задано Ги́льбертово простра́нство H_ν. Прямой суммой Ги́льбертова простра́нства называется (и обозначается $\Sigma_{\nu \in A} \oplus H_\nu$) совокупность *H* всех определенных на *A* функций $\{x_\nu\}$, обладающих тем свойством, что $x_\nu \in H_\nu$ для каждого $\nu \in A$, и

$$\sum_{\nu \in A} \| x_\nu \|^2 < \infty.$$

При этом в *H* линейные операции определяются равенством

$$\{x_\nu\} + \{y_\nu\} = \{x_\nu + y_\nu\}, \quad \alpha \{x_\nu\} = \{\alpha x_\nu\},$$

и скалярное произведение - равенством

$$(\{x_\nu\}, \{y_\nu\}) = \sum_{\nu \in A} (x_\nu, y_\nu)_{H_\nu}.$$

При таком способе введения линейных операций и скалярного произведения прямая сумма

$$H = \sum_{\nu \in A} \oplus H_\nu$$

становится Ги́льбертовым простра́нством

Другой важной операцией в множестве Ги́льбертова простра́нства является тензорное произведение. Тензорным произведением Ги́льбертова простра́нства $H_i,\ i=1,\ 2,\ \ldots,\ n,$ называется Ги́льбертово простра́нство, определяемое следующим образом. Пусть $H_1 \odot H_2 \odot \ldots \odot H_n$ - тензорное произведение векторных пространств $H_1, \ldots, H_n$. В векторном пространстве $H_1 \odot H_2 \odot \ldots \odot H_n$ существует единственное скалярное произведение такое, что

$$(x_1 \odot x_2 \odot \ldots \odot x_n,\ y_1 \odot y_2 \odot \ldots \odot y_n) = \prod_{i=1}^{n} (x_i,\ y_i)_{H_i}$$

для всех $x_i,\ y_i \in H_i$. Векторное пространство становится, таким образом, предгильбертовым пространством, пополнение которого есть Ги́льбертово простра́нство, обозначаемое $H_1 \otimes H_2 \otimes \ldots \otimes H_n$ или $\prod_{i=1}^{n} \otimes H_i$ и называемое тензорным произведением Ги́льбертова простра́нства H_i.

Ги́льбертовы простра́нства образуют важный класс *банаховых пространств*: любое Ги́льбертово простра́нство *H* есть банахово пространство относительно нормы $\|x\| = (x,\ x)^{1/2}$, причем для любых двух векторов $x,\ y \in H$ имеет место равенство параллелограмма:

$$\| x + y \|^2 + \| x - y \|^2 = 2\,(\| x \|^2 + \| y \|^2)$$

Равенство параллелограмма выделяет класс Ги́льбертова простра́нства среди банаховых пространств, т. е. если в действительном нормированном пространстве *B* для любой пары элементов $x, y \in B$. имеет место равенство параллелограмма, то функция

$$(x,\ y) = \frac{1}{4}\,(\| x + y \|^2 - \| x - y \|^2)$$

удовлетворяет аксиомам скалярного произведения и тем самым превращает *B* в предгильбертово пространство (а если *B*- банахово, то - в Ги́льбертово простра́нство). Из равенства параллелограмма следует, что Ги́льбертово простра́нство есть равномерно выпуклое пространство. Как в любом банаховом пространстве, в Ги́льбертовом простра́нстве можно определить две топологии - сильную (нормированную) и слабую. Эти топологии различны, но Ги́льбертово простра́нство сепарабельно в сильной топологии тогда и только тогда, когда оно сепарабельно в слабой топологии; выпуклое множество (в частности, линейное многообразие) в Ги́льбертовом простра́нстве сильно замкнуто тогда и только тогда, когда оно слабо замкнуто.

Как и в теории общих банаховых пространств, в теории Ги́льбертовых простра́нств важную роль играет понятие сепарабельности. Ги́льбертово простра́нство сепарабельно тогда и только тогда, когда оно имеет счетную размерность. Ги́льбертовы простра́нства l^2 и $H_{(l)}$ сепарабельны. Ги́льбертово простра́нство $l^2(T)$ сепарабельно тогда и только тогда, когда *T* не более чем счетно; Ги́льбертово простра́нство $L^2(s,\ \Sigma,\ \mu)$ сепарабельно, если мера μ имеет счетный базис. Ги́льбертово простра́нство B^2 не сепарабельно.

Любой ортонормированный базис в сепарабельном Ги́льбертовом простра́нстве *H* является одновременно безусловным базисом Шаудера в *H*, рассматриваемом как банахово пространство. Однако в сепарабельных Ги́льбертовых простра́нствах существуют и не ортогональные базисы Шаудера. Так, справедлива теорема: пусть $\{f_k\}$ - полная система векторов в Ги́льбертовом простра́нстве *H* и пусть λ_n и Λ_n- наименьшее и наибольшее собственные значения *Грама матрицы,*

$$\{\alpha_{jk}\}_{j,\ k=1}^{n},\quad \alpha_{jk}=(f_k, f_j).$$

Если

$$\underline{\lim}_{n\to\infty} \lambda_n > 0 \quad \text{и} \quad \overline{\lim}_{n\to\infty} \Lambda_n < \infty,$$

то 1) последовательность $\{f_k\}$ есть базис в H; 2) существует биортогональная к $\{f_k\}$ последовательность $\{g_k\}$, которая также является базисом в *H.* Как и в любом банаховом пространстве, описание множества линейных функционалов на Ги́льбертовом простра́нстве и исследование свойств этих функционалов имеет большое значение. Линейные функционалы в Ги́льбертовом простра́нстве устроены особенно просто. Всякий линейный функционал *f* в Ги́льбертовом простра́нстве *H* однозначно записывается в виде $f(x)=(x,\ x^*)$ для всех $x \in H$, где $x^* \in H$; при этом $\|f\|=\|x^*\|$. Пространство H^* линейных функционалов *f* на *H*, сопряженное к *H*, изометрически антинзоморфно *H* (т. е. соответствие $f \to x^*$ изометрнчно, аддитивно н антиоднородно: $\alpha f \to \bar{\alpha} x^*$). В частности, Ги́льбертово простра́нство рефлексивно; поэтому справедливы следующие утверждения: Ги́льбертово простра́нство слабо секвенциально полно; для того чтобы подмножество Ги́льбертовых простра́нств было относительно слабо компактным, необходимо и достаточно, чтобы оно было ограниченным.

Основным содержанием теории Ги́льбертова простра́нства является теория линейных операторов в Ги́льбертовом простра́нстве. Само понятие Ги́льбертова простра́нства сформировалось в работах Д. Гильберта и Э. Шмидта по теории интегральных уравнений, а абстрактное определение Ги́льбертова простра́нства было дано в работах Дж. Неймана, Ф. Рисса и М. Стоуна по теории эрмитовых операторов. Теория операторов в Ги́льбертовом простра́нстве представляет особый важный раздел общей теории операторов но двум причинам.

Во-первых, теория самосопряженных и унитарных операторов в Ги́льбертовом простра́нстве является не только самой разработанной частью общей теории линейных операторов, но и имеющей исключительно широкие приложения в других областях функционального анализа и в ряде других разделов математики и физики. Теория линейных операторов в Ги́льбертовом простра́нстве позволяет рассмотреть многие задачи математической физики с единой общей точки зрения; прежде всего - вопросы, относящиеся к собственным значениям и собственным функциям. Кроме того, теория самосопряженных операторов в Ги́льбертовом простра́нстве служит математическим аппаратом квантовой механики: при описании квантовомеханической системы наблюдаемые (энергия, импульс, координаты и т. п.) интерпретируются как самосопряженные операторы в некотором Ги́льбертовом простра́нстве, а состояние системы задается элементом этого Ги́льбертова простра́нства. В свою очередь, задачи квантовой механики до настоящего времени оказывают влияние на развитие теории самосопряженных операторов, а также теории алгебр операторов в Ги́льбертовом простра́нстве.

Во-вторых, теория *несамосопряженных операторов* в Ги́льбертовом простра́нстве (в частности, циклических, нильпотентных, одноклеточных, сжимающих, спектральных и скалярных операторов) является важной моделью теории линейных операторов в более общих пространствах.

Важный класс линейных операторов в Ги́льбертовом простра́нстве образуют всюду определенные непрерывные операторы, называемые также ограниченными операторами в Ги́льбертовом простра́нстве. Если ввести в множестве $\mathfrak{B}(H)$ всех ограниченных линейных операторов в *H* операции сложения, умножения на число и умножения операторов, а также норму оператора, по обычным правилам (см. *Линейные операторы*) и определить *инволюцию* в $\mathfrak{B}(H)$ как переход к *сопряженному оператору,* то $\mathfrak{B}(H)$ становится *банаховой алгеброй* с инволюцией. Важнейшими классами ограниченных операторов в Ги́льбертовом простра́нстве являются *самосопряженные операторы, унитарные операторы* и *нормальные операторы,* так как они обладают специальными свойствами по отношению к скалярному произведению. Эти классы операторов хорошо изучены; основным инструментом в их изучении являются простейшие из ограниченных самосопряженных операторов, а именно: операторы ортогонального проектирования, или ортогональные проекторы, часто называемые просто *проекторами.* Способ, позволяющий строить любые ограниченные самосопряженные, унитарные и нормальные операторы в комплексном Ги́льбертовом простра́нстве с помощью проекторов, дается *спектральным разложением* соответствующих операторов, особенно простым в случае сепарабельного Ги́льбертова простра́нства.

Более сложным разделом теории линейных операторов в Ги́льбертовом простра́нстве является теория неограниченных операторов. Важнейшими неограниченными операторами в Ги́льбертовом простра́нстве являются замкнутые линейные операторы с плотной областью определения; в частности, неограниченные самосопряженные и нормальные операторы. Между самосопряженными и унитарными операторами в Ги́льбертовом простра́нстве существует взаимно однозначное соответствие, определяемое *Кэли преобразованием.* Большое значение имеет (в частности, в теории линейных дифференциальных операторов) класс *симметричных операторов* в Ги́льбертовом простра́нстве и теория самосопряженных расширений симметричных операторов.

Неограниченные самосопряженные и нормальные операторы в комплексном Ги́льбертовом простра́нстве *Н* также допускают спектральное разложение. Спектральное разложение является большим достижением теории самосопряженных и нормальных операторов в Ги́льбертовом простра́нстве. Оно соответствует классической теории приведения эрмитовых и нормальных комплексных матриц в *n*-мерном унитарном пространстве. Именно спектральное разложение и связанное с ним операторное исчисление для самосопряженных и нормальных операторов обеспечивают теории операторов в Ги́льбертовом простра́нстве широкую область применения во многих разделах математики. Для ограниченных самосопряженных операторов в l_2 спектральное разложение было найдено Д. Гильбертом, который также ввел понятие *разложения единицы* для самосопряженного оператора. В настоящее время известно несколько подходов к спектральной теории самосопряженных и нормальных операторов. Один из наиболее глубоких дает теория банаховых алгебр. Спектральное разложение для неограниченного самосопряженного оператора было найдено Дж. Нейманом. Его работе предшествовали важные исследования Т. Карлемана, который получил спектральное разложение для случая симметрического интегрального оператора, а также впервые обнаружил, что между симметрическими ограниченными и неограниченными операторами полной аналогии нет. На важность понятия самосопряженного оператора впервые обратил внимание Э. Шмидт.

9. Банахово пространство

B-пространство - полное нормированное *векторное пространство*. Исходными для создания теории Банахова простра́нства послужили введенные (в 1904-18) Д. Гильбертом (D. Hilbert), М. Фреше (M. Frechet) и Ф. Рисом (F. Riesz) функциональные пространства. Именно в этих пространствах были первоначально исследованы фундаментальные понятия сильной и слабой сходимости, компактности, линейного функционала, линейного оператора и др. Банаховы простра́нства, названы по имени С. Банаха, который в 1922 начал систематическое изучение этих пространств на основе введенной им аксиоматики и получил глубокие результаты.

Теория Банаховы простра́нств развивалась параллельно с общей теорией *линейных топологических пространств.* Эти теории взаимно обогащались идеями и фактами. Так, идея полунормы, заимствованная из теории нормированных пространств, стала необходимым инструментом для построения теории локально выпуклых линейных топологических пространств. Понятие слабой сходимости элементов и линейных функционалов в Банаховом простра́нстве обрело законченную форму в понятии слабой топологии. Теория Банахова простра́нства представляет собой хорошо разработанную область функционального анализа, имеющую (непосредственно или через теорию операторов) многочисленные применения в различных разделах математики.

Проблематика Банахового простра́нства складывается из нескольких направлений: геометрия единичной сферы, геометрия подпространств, линейно топологической классификация, ряды и последовательности в Банаховом простра́нстве, наилучшие приближения в Банаховом простра́нстве, функции со значениями в Банаховом простра́нстве и др. Относительно теории операторов в Банаховом простра́нстве следует отметить, что многие ее предложения имеют непосредственное отношение к геометрии и топологии Банахового простра́нства.

Примеры. Встречающиеся в математическом анализе Банаховые простра́нства- это чаще всего множества функций или числовых последовательностей, подчиненные некоторым условиям.

1) $l_p,\ p\geqslant 1$,- пространство числовых последовательностей $x=\{\xi_n\}$, для которых

$$\sum_{n=1}^{\infty} |\ \xi_n\ |^p < \infty,$$

с нормой

$$||\ x\ || = \left(\sum_{n=1}^{\infty} |\ \xi_n\ |^p\right)^{1/p}.$$

2) *m* - пространство ограниченных числовых последовательностей с нормой

$$||\ x\ || = \sup_n |\ \xi_n\ |.$$

3) *c* - пространство сходящихся числовых последовательностей с нормой

$$||\ x\ || = \sup_n |\ \xi_n\ |.$$

4) c_0 - пространство сходящихся к нулю числовых последовательностей с нормой

$$||\ x\ || = \max_n |\ \xi_n\ |.$$

5) $C\ [a,\ \ b]$- пространство непрерывных на $[a,\ \ b]$ функций $x=x\ (t)$ с нормой

$$||\ x\ || = \max_{a\leqslant t\leqslant b} |\ x\ (t)\ |.$$

6) $C\ [K]$ - пространство непрерывных функций на компакте K с нормой

$$\| x \| = \max_{t \in K} | x(t) |.$$

7) $C^n[a,\ b]$ - пространство функций, имеющих непрерывные производные до порядка *n* включительно, с нормой

$$\| x \| = \sum_{k=0}^{n} \max_{a \leqslant t \leqslant b} | x^{(k)}(t) |.$$

8) $C^n[I_m]$ - пространство всех непрерывно дифференцируемых до порядка *n* функций, определенных в *m*- мерном кубе, с равномерной нормой по всем производным порядка не выше *n*.

9) $M\ [a,\ b]$ - пространство ограниченных измеримых функций с нормой

$$\| x \| = \operatorname{vrai} \max_{a \leqslant t \leqslant b} | x(t) |.$$

10) $A\ (D)$ - пространство функций, аналитических в открытом единичном круге D и непрерывных в замкнутом круге $\overline{D}$, с нормой

$$\| x \| = \max_{z \in \overline{D}} | x(z) |.$$

11) $L_p(S;\ \Sigma,\ \mu),\ p \geqslant 1,$ - пространство функций $x(s)$ на множестве *S* с вполне аддитивной мерой *m* и с нормой

$$\| x \| = \left(\int_S | x(s) |^p \mu(ds) \right)^{1/p}.$$

12) $L_p[a, b],\ p \geqslant 1,$ - частный случай пространства $L_p(S; \Sigma, \mu)$ - пространство измеримых по Лебегу функций, суммируемых со степенью *p,* с нормой

$$\| x \| = \left(\int_a^b | x(s) |^p ds \right)^{1/p}.$$

13) *AP* - пространство Бора почти периодических функций с нормой

$$\| x \| = \sup_{-\infty < t < +\infty} | x(t) |.$$

Пространства $C\ [a, b],\ C^n[a, b],\ L_p[a, b],\ c,\ l_p$ сепарабельны; пространства $M\ [a, b],\ m,\ AP$ несепарабельны; $C\ [K]$ сепарабельно в том и только в том случае, если *К* - метрический компакт.

Подпространство *Y* Банахова простра́нства, рассматриваемое отдельно от вмещающего пространства *X,* есть Банахово простра́нство. Факторо-пространство X/Y нормированного пространства по подпространству Y становится нормированным пространством, если норму определить так: пусть $Y_1 = x_1 + Y$ - класс смежности. Тогда

$$\| Y \| = \inf_{y \in Y} \| x_1 + y \|.$$

Если *X* - Банахово простра́нство, то *X/Y*- тоже Банахово простра́нство. Множество всех линейных функционалов, определенных в нормированном пространстве *X,* с нормой

$$\| f \| = \sup_{x \in X} \frac{| f(x) |}{\| x \|}, \quad x \neq 0,$$

называется пространством, сопряженным с .X, и обозначается X^*. Оно является Банаховым простра́нством.

Для Банахового простра́нства справедлива *Хана - Банаха теорема* о продолжении линейных функционалов: если линейный функционал определен на подпространстве Y нормированного пространства *X,* то его можно распространить с сохранением линейности и непрерывности на все пространство *X*. Более того, при этом можно обеспечить сохранение нормы расширенного функционала:

$$\| f \|_X = \sup_{x \in X} \frac{| f(x) |}{\| x \|} = \| f \|_Y = \sup_{y \in Y} \frac{| f(y) |}{\| y \|}.$$

Справедливо и более общее утверждение: пусть действительная функция *p* (*x*), определенная в линейном пространстве, удовлетворяет условиям:

$$p(x+y) \leqslant p(x) + p(y),$$
$$p(\lambda x) = \lambda p(x), \quad \lambda \geqslant 0, \quad x,\ y \in X,$$

и пусть $f(x)$ - действительный линейный функционал, определенный на подпространстве $Y \subset X$ и такой, что

$$f(x) \leqslant p(x), \quad x \in Y.$$

Тогда существует линейный функционал $F(x)$, определенный на всем *X* и такой, что

$$F(x) = f(x) \text{ для } x \in Y; \ F(x) \leqslant p(x) \text{ для } x \in Y.$$

Следствием теоремы Хана - Банаха является "обратная" -формула, связывающая нормы X и X^*:

$$\| x \| = \max_{b \in X^*} \frac{| f(x) |}{\| f \|}, \quad f \neq 0, \ x \in X,$$

причем max в этой формуле достигается на некотором $f = f_x \in X^*$. Другое важное следствие - наличие *достаточного множества* непрерывных линейных функционалов. Говорят, что в Банаховом простра́нстве *X* существует достаточно много непрерывных линейных функционалов, если для любых $x_1 \neq x_2 \in X$ имеется такой определенный на *X* линейный функционал *f*, что $f(x_1) \neq f(x_2)$.

Для многих конкретных Банаховых простра́нств известен общий вид линейного функционала. Так, в $L_p[a, b], (p > 1)$ каждый линейный функционал определяется по формуле:

$$f(x) = \int_a^b x(t)\, y(t)\, dt,$$

где $y \in L_q[a, b], \ \frac{1}{p} + \frac{1}{q} = 1$, а каждая функция $y(t) \in L_q$ определяет по этой формуле линейный функционал *f*, причем

$$\| f \| = \left\{ \int_a^b | y(t) |^q \, dt \right\}^{1/q}.$$

Таким образом, пространством, сопряженным с L_p, является $L_q: L_p^* = L_q$. В $L_1[a, b]$ линейные функционалы задаются той же формулой, но в этом случае $y \in M$, так что $L_1^* = M$.

Пространство X^{**}, сопряженное с X^*, называют вторым сопряженным. Аналогично определяются третье, четвертое и т. д. сопряженные пространства. Каждый элемент из *X* может быть отождествив с некоторым линейным

$$F(f) = f(x), \ \ \forall f \in X^* \ \ (F \in X^{**}, \ x \in X).$$

функционалом, определенным на X^*:

При этом $\| F \| = \| x \|$. Тогда можно считать *X* подпространством пространства X^{**} и $X \subset X^{**} \subset X^{IV} \ldots, \ X^* \subset X^{***} \subset \ldots$. Если при указанном вложении Банаховое простра́нство совпадает со своим вторым сопряженным, то оно называется рефлексивным. В этом случае все включения оказываются равенствами. Если же *X* не рефлексивно, то среди включений нет ни одного равенства. Если факторпространство X^{**}/X имеет конечную размерность и, то X называется квазирефлексивным порядка *n*. Квазирефлексивные пространства существуют для любого *n*.

Критерии рефлексивности Банахово простра́нство: 1) *X* рефлексивно тогда и только тогда, когда для каждого $f \in X^*$ найдется $x \in X$, на котором достигается sup в формуле

$$\| f \| = \sup_{x \in X} \frac{| f(x) |}{\| x \|}, \ \ x \neq 0.$$

2) В рефлексивных Банаховых простра́нствах, и только в них, каждое ограниченное множество компактно относительно слабой сходимости:

любая его бесконечная часть содержит слабо сходящуюся подпоследовательность (теорема Эберлейна- Шмульяна). Пространства L_p и $l_p,\ p>1,$ рефлексивны. Пространства $L_1,\ l_1,\ C,\ M,\ c,\ m,\ AP$ нерефлексивны.

Банахово простра́нство называется слабо полным, если в нем каждая слабо сходящаяся в себе последовательность слабо сходится к элементу пространства. Каждое рефлексивное пространство слабо полно. Кроме того, слабо полны Банаховы простра́нства L_1 и l_1. Еще более широкий класс - Банаховые простра́нства, не содержащие подпространств, изоморфных c_0. Эти пространства во многом подобны слабо полным.

Банахово простра́нство называется строго нормированным, если его единичная сфера S не содержит отрезков. Для количественной оценки выпуклости единичного шара вводятся модули выпуклости: локальный модуль выпуклости

$$\delta(x,\ \varepsilon)=\inf_{\|x-y\|\geqslant\varepsilon}\left\{1-\left\|\frac{x+y}{2}\right\|\right\},\ x,\ y\in S,\ \ 0<\varepsilon\leqslant 2,$$

и равномерный модуль выпуклости

$$\delta(\varepsilon)=\inf_{x\in S}\delta(x,\ \varepsilon).$$

Если $\delta(x,\varepsilon)>0$ для всех $x\in X$ и всех $\varepsilon>0$, то Банахово простра́нство называется локально равномерно выпуклым. Если $\delta(\varepsilon)>0$, то пространство называется равномерно выпуклым. Каждое равномерно выпуклое Банахово простра́нство локально равномерно выпукло; каждое локально равномерно выпуклое Банахово простра́нство строго нормировано. В конечномерных Банаховых простра́нствах верны и обратные включения. Если Банахово простра́нство равномерно выпукло, то оно рефлексивно.

Банахово простра́нство называется гладким, если для любых линейно независимых элементов x и y функция $\psi(t) = \| x + ty \|$ дифференцируема при всех *t*. Банахово простра́нство называется равномерно гладким, если для его модуля гладкости

$$\rho(\tau) = \sup_{x,\, y \in S} \left\{ \frac{\| x + \tau y \| + \| x - \tau y \|}{2} - 1 \right\}, \quad \tau > 0,$$

выполняется условие

$$\lim_{\tau \to 0} \frac{\rho(\tau)}{\tau} = 0.$$

В равномерно гладких Банаховых простра́нствах, и только в них, норма равномерно дифференцируема по Фреше. Равномерно гладкое Банахово простра́нство гладко. Обратное верно, если Банахово простра́нство конечномерно. Банахово простра́нство *X* равномерно выпукло (равномерно гладко) в том и только в том случае, если X^* равномерно гладко (равномерно выпукло). Между модулем выпуклости Банахова простра́нства *X* и модулем гладкости X^* существует связь

$$\rho_{X^*}(\tau) = \sup_{0 < \varepsilon \leqslant 2} \left[\frac{\varepsilon(\tau)}{2} - \delta_X(\varepsilon) \right].$$

Если Банахово простра́нство равномерно выпукло (равномерно гладко), то таковы любое его подпространство и факторпространство. Банаховы простра́нства L_p и $l_p (p > 1)$ равномерно выпуклы и равномерно гладки, причем ,

$$\delta(\varepsilon) \asymp \begin{cases} \varepsilon^2 \ (1 < p \leqslant 2), \\ \varepsilon^p \ (2 \leqslant p < \infty); \end{cases}$$

$$\rho(\tau) \asymp \begin{cases} \tau^p \ (1 < p \leqslant 2), \\ \tau^2 \ (2 \leqslant p < \infty); \end{cases}$$

$$(f(\varepsilon) \asymp \varphi(\varepsilon) \Longleftrightarrow a < f(\varepsilon)/\varphi(\varepsilon) < b).$$

Банаховы простра́нства $M,\ C,\ A,\ L_1,\ AP,\ m,\ c,\ l_1$ не являются строго нормированными и не являются гладкими.

В Банаховом простра́нстве справедливы следующие важные теоремы для линейных операторов:

Теорема Банаха- Штейнхауза. Если семейство линейных операторов $\mathcal{T} = \{T_\alpha\}$ ограничено в каждой точке:

$$\sup_{T \in \mathcal{T}} \| T_\alpha x \| < \infty \quad \text{для всех} \quad x \in X,$$

то оно ограничено по норме:

$$\sup_{T \in \mathcal{T}} \| T_\alpha \| < \infty .$$

Теорема Банаха об обратном операторе. Если линейный непрерывный оператор отображает взаимно однозначно Банахово простра́нство *X* на Банахово простра́нство *Y*, то обратный оператор T^{-1} тоже непрерывен.

Теорема о замкнутом графике. Если замкнутый линейный оператор отображает Банахово простра́нство *X* в Банахово простра́нство *Y*, то он непрерывен.

Изометрия Банахового простра́нства сравнительно редкое явление. Классический пример - Банаховы простра́нства L_2 и l_2. Банаховы простра́нства $C\ [K_1]$ и $C\ [K_2]$ изометричны в том и только в том случае, если K_1 и K_2 гомеоморфны (теорема Банаха- Стоуна). Для изоморфных Банаховых простра́нств мерой близости служит число

$$d\,(X,\ Y) = \ln \inf_T \|\, T \,\| \cdot \|\, T^{-1} \,\|,$$

где *T* пробегает всевозможные операторы, осуществляющие изоморфизм между X и Y. Если *X* изометрично *Y*, то $d\,(X,\ Y) = 0$. Однако существуют и не изометричные пространства, для которых $d\,(X,\ Y) = 0$; их называют почти изометричными. Свойства Банаховых простра́нств, сохраняющиеся при изоморфизме, называются линейно топологическими. К ним относятся сепарабельность, рефлексивность, слабая полнота. Изоморфная классификация Банахового простра́нства содержит, в частности, следующие утверждения:

$$L_r \neq L_s;\ l_r \neq l_s,\ r \neq s;$$
$$L_r \neq l_s,\ r \neq s \ \text{или}\ r = s = 2;$$
$$M = m;\ C\,[0,\ 1] \neq A\,(D);$$

$C\ [K] = C\ [0,\ 1]$, если *К* - метрический компакт мощности континуума; $C^n\,[I_m] \neq C\,[0,\ 1]$.

Каждое сепарабельное Банахово простра́нство изоморфно локально равномерно выпуклому.

Отвлекаясь от линейной природы нормированных пространств, можно рассматривать их топологию, классификацию. Два пространства гомеоморфны, если между их элементами может быть установлено взаимно однозначное и взаимно непрерывное (не обязательно линейное) соответствие. Неполное нормированное пространство не гомеоморфно никакому Банаховому простра́нству. Все бесконечномерные сепарабельные Банаховы простра́нства гомеоморфны.

Универсальными (см. *Универсальное пространство*),в классе сепарабельных Банаховых простра́нств являются C[0, 1] и A(D). В классе рефлексивных сепарабельных Банаховых простра́нств нет даже изоморфно универсального. Банахово простра́нство l_1 универсально в несколько ином смысле: каждое сепарабельное Банахово простра́нство изометрично некотоому его факторпространству.

В каждом из перечисленных выше Банаховых простра́нствах, кроме L_2 и l_2, существуют подпространства без дополнения. В частности, в т и M не дополняемо каждое бесконечномерное сепарабельное подпространство, в C[0, 1] не дополняемо каждое бесконечномерное рефлексивное подпространство. Если в Банаховом простра́нстве все подпространства дополняемы, то оно изоморфно гильбертову пространству. Подпространство Yдополняемо в том и только в том случае, если существует проектор, отображающий X на *Y*.

Относительной проекционной константой $\lambda(Y,\ X)$ подпространства Yв X называется нижняя грань норм проекторов на *Y*. Каждое *n*-мерное подпространство Банахова простра́нства дополняемо и $\lambda\,(Y_n,\,X)\leqslant\sqrt{n}.$

Абсолютной проекционной константой $\lambda(Y)$ Банахового простра́нства Y называется

$$\lambda\,(Y)=\sup_X \lambda\,(Y,\ X),$$

где *X* пробегает все Банаховые простра́нства, содержащие Y в качестве подпространства. Для любого бесконечномерного сепарабельного Банахового простра́нства Y имеем $\lambda(Y)=\infty$. Банаховые простра́нства, для которых $\lambda(Y)\leqslant\lambda<\infty$, образуют класс $\mathcal{P}_\lambda\ (\lambda\geqslant 1)$. Класс $\mathcal{P}_1$ совпадает с классом пространств

$C(Q)$, где Q - экстремально несвязные компакты (см. *Экстремально несвязные пространства*).

Основные теоремы о конечномерных Банаховых простра́нствах:
1) Конечномерное нормированное пространство (*Минковского пространство*) полно, т. е. является Банаховым простра́нством.
2) Каждый линейный оператор в конечномерном Банаховом простра́нстве непрерывен.
3) Конечномерное Банахово простра́нство рефлексивно (размерность X* равна размерности X).
4) Банахово простра́нство конечномерно в том и только в том случае, если его единичный шар компактен.
5) Все *n*-мерные Банаховы простра́нства попарно изоморфны; их множество становится компактом, если ввести расстояние

$$d(X,\ Y) = \ln \inf \| T \| \cdot \| T^{-1} \|.$$

Ряд

$$\sum_{k=1}^{\infty} x_k, \quad x \in X, \qquad (*)$$

называется сходящимся, если существует предел S последовательности частных сумм:

$$\lim_{n \to \infty} \left\| S - \sum_{k=1}^{n} x_k \right\| = 0.$$

Если

$$\sum_{k=1}^{\infty} \| x_k \| < \infty,$$

то ряд (*) сходится; в этом случае он называется абсолютно сходящимся. Ряд называется безусловно сходящимся, если он сходится при любой перестановке его членов. Сумма безусловно сходящегося ряда не зависит от порядка его членов. Для рядов в конечномерном пространстве (и, в частности, для числовых рядов) безусловная сходимость эквивалентна абсолютной. В бесконечномерных Банаховых простра́нствах из абсолютной сходимости следует безусловная, но обратное неверно ни в одном бесконечномерном Банаховом простра́нстве. Последнее есть следствие теоремы Дворецкого- Роджерса: каковы бы ни были числа $a_k \geqslant 0$, подчиненные условию $\sum a_k^2 < \infty$, в любом бесконечномерном Банаховом простра́нстве существует такой безусловно сходящийся ряд $\sum x_k$, что $\| x_k \| = a_k, k = 1, 2, \ldots$.

В пространстве c_0 (а значит, и в любом Банаховом простра́нстве, содержащем подпространство, изоморфное c_0) для любой сходящейся к нулю последовательности $a_k \geqslant 0$ существует безусловно сходящийся ряд $\sum x_k, \|x_k\| = a_k$. В $L_p(S, \Sigma, \mu)$ из безусловной сходимости ряда $\sum x_k$ следует, что

$$\sum_{k=1}^{\infty} \| x_k \|^s < \infty,$$

где

$$s = \begin{cases} 2 \ \ (1 \leqslant p \leqslant 2), \\ p \qquad (p \geqslant 2). \end{cases}$$

В равномерно выпуклом Банаховом простра́нстве с модулем выпуклости $\delta(\varepsilon)$ из безусловной сходимости ряда $\sum x_k$ следует, что

$$\sum_{k=1}^{\infty} \delta (\| x_k \|) < \infty .$$

Ряд $\sum x_k$ называется слабо абсолютно сходящимся, если для каждого $f \in X^*$ сходится числовой ряд $e| f(x_k)|$. Каждый слабо абсолютно сходящийся ряд в *X* сходится в том и только в том случае, если X не содержит подпространства, изоморфного c_0.

Последовательность элементов $\{e_k\}^\infty$ Банахова простра́нства называется минимальной, если любой ее член лежит вне замыкания $X^{(n)}=[e_k]_{k\neq n}$ линейной оболочки остальных членов. Последовательность называется равномерно минимальной, если

$$\rho(e_n;\ X^{(n)})\geqslant\gamma\,\|e_n\|,\ 0<\gamma\leqslant 1,\ n=1,\ 2,\ \ldots\ .$$

Если $\gamma=1,$ то последовательность называется системой Ауэрбаха. В каждом re-мерном Банаховом простра́нстве существует полная система Ауэрбаха $\{e_n\}_1^n.$ Для каждой минимальной системы существует сопряженная система линейных функционалов $\{f_n\},$ связанная с $\{e_k\}$ соотношениями биортогональности: $f_i(e_j)=\delta_{ij}.$ В этом случае система $\{e_k,\ f_k\}$ называется биортогональной. Множество линейных функционалов называется тотальным, если оно аннулирует только нулевой элемент пространства. В каждом сепарабельном Банаховом простра́нстве существует полная минимальная система с тотальной сопряженной. Каждый элемент $x\in X$ может быть формально разложен в ряд по биортогональной системе:

$$x\sim\sum_{k=1}^{\infty}f_k(x)\,e_k,$$

однако в общем случае этот ряд расходится.

Система элементов $\{e_k\}_1^\infty$ называется базисом в *X*, если каждый элемент $x\in X$ может быть единственным образом представлен в виде сходящегося ряда

$$x = \sum_{k=1}^{\infty} a_k e_k, \quad a_k = a_k(x).$$

Каждый базис в Б. п.- полная равномерная минимальная система с тотальной сопряженной. Обратное неверно, как показывает пример системы $\{e^{int}\}_{-\infty}^{\infty}$ в $C\ [0,\ 2\pi]$ и $L_1\ \{0,\ 2\pi]$.

Базис называется безусловным, если каждая его перестановка - также базис, в противном случае базис называется условным. Система $\{e^{int}\}_{-\infty}^{\infty}$ в $L_p[0,\ 2\ \pi], p>1,\ p\neq 2,$ - условный базис. Безусловным базисом в $L_p,\ p>1,$ является система Хаара. В пространствах С и L_1 нет безусловного базиса. Каждое Банахово простра́нство с безусловным базисом либо рефлексивно, либо содержит подпространство, изоморфное l_1 или c_0.

Нормированные базисы $\{e'_k\}$ и $\{e''_k\}$ в Банаховых простра́нствах X_1 и X_2 называются эквивалентными, если соответствие $e'_k \leftrightarrow e''_k$ ($k=1,\ 2,\ \ldots$) может быть продолжено до изоморфизма между X_1 и X_2. В каждом из пространств $l_2,\ l_1,\ c_0$ каждый нормированный безусловный базис эквивалентен естественному базису. Базисы, построенные в важных для приложений Банаховых простра́нств, не всегда хорошо приспособлены к решению задач, например, теории операторов. В связи с этим введены *T* - базисы, или базисы суммирования. Пусть $\{t_{ij}\}_1^{\infty}$ - матрица *регулярного метода суммирования.* Система элементов $\{e_n\}\subset X$ называется *T*- базисом, соответствующим данному методу суммирования, если каждый $x\in X$ единственным образом представляется в виде ряда

$$x \sim \sum_{k=1}^{\infty} a_k e_k,$$

суммируемого этим методом к x. Тригонометрическая система $\{e^{int}\}_{-\infty}^{\infty}$ в $C\ [0,\ 2\ \pi]$ - базис суммирования для методов Чезаро и Абеля. Каждый Т-базис - полная минимальная (не обязательно равномерно) система с тотальной сопряженной. Обратное неверно.

10. Топологические пространства.

Рассмотрим еще один тип пространств - топологическое пространство.

Топологическое пространство — это пары $(X,\ \tau)$, где τ — система подмножеств G множества M, называемый *топологией* в X, которое содержит:

1) само множество X и пустое множество Ø, т.е. $X \in \tau$ и $\varnothing \in \tau$;

2) пересечение любой пары своих подмножеств, т.е. $G_i \bigcap G_j \in \tau$;

3) объединение любого (конечного или бесконечного) множества своих подмножеств, т.е. $\bigcup G_i \in \tau$.

Из приведенного определения следует, что v содержит также пересечение любого *конечного* множества своих подмножеств.

Множества, которые принадлежат системе τ, называются *открытыми* множествами пространства $(X,\ \tau)$. Одно и то же множество X может допускать несколько топологий и при этом получаются разлияные пространства. Всякое множество X допускает *тривиальную топологию,* при которой открытыми множествами считаются только X и Ø *(пространство слипшихся точек),* а также *дискретную топологию,* когда открыто любое подмножество X.

Множества $X \backslash G$, дополнительные к открытым, называются *замкнутыми* множествами топологического пространства. Из определения топологического пространства следует, что замкнутыми множествами есть:

1) X и Ø;

2) объединение конечного числа замкнутых множеств;

3) пересечение любого (конечного или бесконечного) числа замкнутых множеств.

Как видно, имеет место дуальность в определении открытых и замкнутых множеств топологического пространства.

10.1. Структура топологического пространства

Структура топологического пространства, вводимая на произвольном множестве *X*, снабжает элементы множества *X* свойством "притяжения", или "близости".

Напомним, что топологическим пространством (X,τ) называется множество *X*, в котором зафиксирован класс τ подмножеств, называемых *открытыми*, удовлетворяющий следующей системе аксиом (*аксиомы топологии для открытых подмножеств*):
1^0. $X \in \tau$;
2^0. $\varnothing \in \tau$;
3^0. Для любого набора $\{U_2\}_{2\in I}, U_2 \in \tau$, выполнено условие $\bigcup_{2\in I} U_2 \in \tau$.
4^0. Для любого *конечного* набора $\{U_i\}^n_{i=1}, U_i \in \tau$, выполнено условие $\bigcap^n_{i=1}, U_i \in \tau$.

Элементы множества *X* называются *точками*, класс τ - *топологией* на множестве *X*.

Требование конечности в 4^0 является существенным. В самом деле, рассмотрим множество вещественных чисел R, класс открытых множеств которого определен привычным из курса математического анализа способом. Открытыми считаются подмножества следующих типов: пустое подмножество $\varnothing$,открытые интервалы вида $(a, b)=\{x \in \mathrm{R} | a<a<\mathrm{b}\}$ и их объединения. Очевидно, подмножество (-∞, ∞) может быть представлено в виде объединения открытых интервалов. Такая топология на числовой прямой называется *классической*.

Теперь рассмотрим систему вложенных интервалов $(-1/i,\ 1/i\ )^{\infty}_{i=1}$, которая обладает единственной общей точкой {0}. Одноточечное множество на числовой прямой не является открытым в классической топологии. На одном и том же множестве могут быть введены различные топологии. Например, считая открытыми на множестве вещественных чисел только множества видов $\varnothing$, (-*a*, *a*) и (-∞, ∞), получим *концентрическую топологию*. Объявляя открытыми любые объединения одноточечных подмножеств любого множества *X*, а также пустое подмножество и само X, получим *дискретную топологию*, превращающую множество X в *пространство изолированных точек*. *Подпространством* $(Y,\tau|_Y)\subset(X,\tau)$ топологического пространства (X, τ) называется подмножество $Y\in X$ со следующей (индуцированной) топологией $\tau|Y$: открытыми в Y являются подмножества вида $U|_Y:= U\bigcap Y$ (рис.).

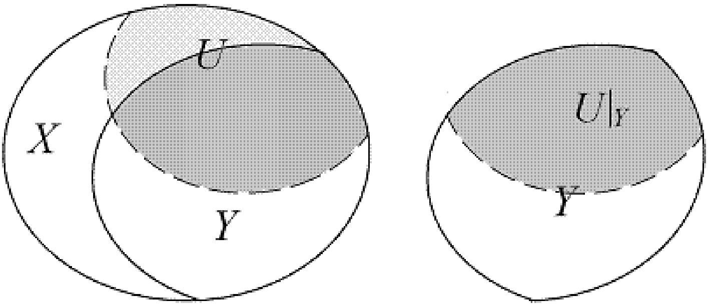


Рис.1. Индуцированная топология подпространства

Упражнение. Проверьте, что множества $U|_Y$ действительно образуют топологию на множестве Y, то есть удовлетворяют аксиомам топологии.

В качестве примера построим топологию подпространства на отрезке $[a, b]$, если на содержащей его прямой введена классическая топология. Открытыми на отрезке $[a, b]$ будут подмножества следующего вида (и только они): $\varnothing$, $(c, d)\subset [a, b]$, а также $[a, d)$ и $(c, b]$. *Базой топологии пространства* (X, τ) называется подкласс $\tau_B\subset \tau$, такой, что для всякого открытого подмножества $U\in\tau$ и для всякой точки $p\in U$ найдется такое подмножество $V\in\tau_B$, что $p\in V\subset U$. Иными словами, подкласс $\tau_B \subset \tau$ является базой топологии τ, если любое открытое множество $U\in \tau$ представимо в виде объединения некоторого набора множеств, принадлежащих подклассу τ_B. Очевидно, одна и та же топология на множестве X может быть задана различными базами. При этом база полностью определяет топологию: открытыми являются в точности те множества, которые представимы в виде объединения подмножеств базы. Если пространство (X, τ) таково, что подкласс τ_B может состоять из счетного набора подмножеств, то говорят, что (X, τ) - *пространство со счетной базой*. Пример топологического пространства со счетной базой легко построить. Таковым является числовая прямая с классической топологией. База топологии представлена интервалами с концами в

рациональных точках. Множество таких интервалов, как нетрудно видеть, счетно. Читателю предлагается построить счетную базу классической топологии на плоскости.
Назовем *замкнутым* любое множество, дополнение к которому открыто. *Замыкание* $\overline{Y}$ подмножества Y в топологическом пространстве X - это пересечение всех открытых в X подмножеств, содержащих подмножество Y. Можно доказать, что замыкание $\overline{Y}$ - это наименьшее (по включению) замкнутое подмножество пространства X, содержащее подмножество Y. Очевидно, замкнутое подмножество совпадает со своим замыканием, то есть $\overline{Y}=Y$.
Упражнение. Используя систему вложенных интервалов, докажите, что в классической топологии точка на числовой прямой является замкнутым подмножеством.
Пусть $p\in(X, \tau)$ - точка топологического пространства (X, τ). *Окрестностью* точки p называется любое открытое в X подмножество, содержащее точку p.

Аксиомы отделимости - это дополнительные ограничения, накладываемые на топологическую структуру и приближающие свойства топологического пространства к привычным свойствам пространства R^n. Наиболее часто используются следующие аксиомы.
T1. *Для любых двух различных точек p и qсуществует окрестность точки p, не содержащая точку q.* Это означает, что точка есть замкнутое множество; конечные множества замкнуты. Действительно, если выполнена аксиома T1, то дополнение к точке p равно объединению всех не содержащих p окрестностей всех отличных от p точек q. Объединение любого набора открытых множеств открыто; дополнение к открытому множеству замкнуто. Обратно, если точка - замкнутое множество, то она совпадает со своим замыканием. Это значит, что точка p обладает хотя бы одной окрестностью, не содержащей точку q (иначе бы точка $q\neq p$ принадлежала замыканию точки p, что противоречило бы замкнутости точки p).
T2. *Для любых двух различных точек p и q существуют непересекающиеся окрестности,* то есть множества $U_p\in\tau$ и $U_q\in\tau$, такие, что $U_p\bigcap U_q=\varnothing$.
T3. *Любая точка и любое не содержащее ее замкнутое множество обладают непересекающимися окрестностями.* Равносильная формулировка: любая окрестность любой точки содержит замыкание некоторой окрестности этой точки.

Упражнение. Докажите, что приведенные формулировки аксиомы Т3 равносильны.

Т4. *Любые два непересекающиеся замкнутые множества обладают непересекающимися окрестностями.* Равносильная формулировка: каждая окрестность замкнутого множества содержит замыкание некоторой окрестности этого множества.

Упражнение. Докажите, что приведенные формулировки аксиомы Т4 равносильны.

Топологическое пространство (X, τ) называется *хаусдорфовым*, если оно удовлетворяет аксиоме Т2.

Упражнение. Покажите, что хаусдорфово пространство удовлетворяет аксиоме Т1.

Хаусдорфовыми пространствами являются, например, числовая прямая с классической топологией, а также любое пространство изолированных точек. Числовая прямая с концентрической топологией нехаусдорфова.

Упражнение. Докажите, что подпространство хаусдорфова пространства хаусдорфово.

Именно выполнение аксиомы Т2 на числовой прямой обеспечивает единственность предела числовой последовательности. Более того, нетрудно доказать, что во всяком хаусдорфовом пространстве предел сходящейся последовательности точек единствен. В нехаусдорфовом пространстве предел может быть не один!

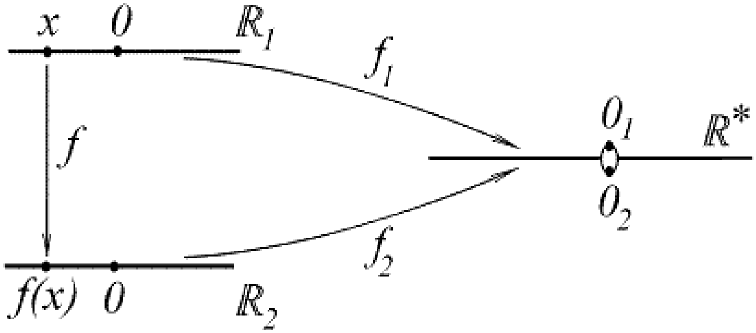


Рис.2. Пример нехаусдорфова топологического пространства

Простейший пример нехаусдорфова пространства может быть построен следующим образом (рис.2). Возьмем два экземпляра R_1 и R_2 числовой прямой с классической топологией. Рассмотрим отображение f:$R_1\backslash\{0\}\to R_2\backslash\{0\}$, такое, что $f(x)=x$. Топологическое пространство R^+ получается отождествлением открытых интервалов $(-\infty, 0)$ и $(\infty, 0)$ на прямых R_1 и R_2 посредством отображения f. Имеют место отображения числовых прямых R_1 и R_2 в пространство f_i: $R_i\to R^+$, такие, что каждой точке с ненулевой координатой x ставится в соответствие точка, отвечающая паре (x, $f(x)$) при построении пространства R^+. При этом пространство R^+ содержит две "точки с нулевой координатой"; это $0_1:=f_1(0)$ и $0_2:=f_2(0)$. Открытыми в пространстве R^+ будем считать образы открытых подмножеств числовых прямых R_1 и R_2. Тогда точки 0_1 и 0_2 не обладают непересекающимися окрестностями. Рассмотрим теперь числовую последовательность $a_1,\ldots,a_k,\ldots$ с ненулевыми членами, сходящуюся к нулю в R_1, и такую же последовательность в R_2. Тогда отображениями f_1 и f_2 поставляется последовательность, имеющая два предела $0_1\neq 0_2$.

Метрическим пространством называется множество X, снабженное функцией d:$X\times X\to R$, удовлетворяющей следующим аксиомам (*аксиомы расстояния*):

1^0. *Знакоопределенность* $d(p, q)\geq 0$ для любых точек $p\in X$ и $q\in X$.

2^0. *Невырожденность*: $d(p, q)=0$ тогда и только тогда, когда $p=q$.

3^0. *Симметричность*: $d(p, q)=d(q, p)$ для любых точек $p\in X$ и $q\in X$.

4^0. *Неравенство треугольника*: $d(p, q)+d(q, r)\geq d(r, q)$ для любых точек $p\in X$, $q\in X$, и $r\in X$.

Функция d называется *функцией расстояния, или метрикой.* Метрическое пространство является топологическим. Индуцированная метрикой d топология на множестве X определяется следующим образом: объявим открытыми подмножества $\varnothing$, X и всевозможные объединения подмножеств (*открытых шаров*) вида

$$B_a(p):=\{q\in X \mid d(p, q) < a\}$$

для всех положительных чисел a и точек $p\in X$. Выполнение аксиом 1^0- 4^0 топологического пространства легко проверить самостоятельно. Любое метрическое пространство с топологией, индуцированной метрикой, хаусдорфово. Действительно, пусть p и q - две различные точки метрического пространства (X, d), и пусть расстояние $d(p, q)$ между ними равно R. Тогда в качестве искомых непересекающихся окрестностей можно взять открытые шары $B_{R/3}(p)$ и $B_{R/3}(q)$ радиуса $R/3$ с центрами в точках p и q. Далеко не всякое топологическое пространство допускает метрику, такую, чтобы топология была индуцирована этой метрикой. Например, любое нехаусдорфово пространство не является метризуемым. Это и понятно: образно говоря, оно содержит "слипшиеся" точки. Расстояние между такой парой точек, если бы оно было определено, должно быть отлично от нуля, но эти точки не обладают непересекающимися окрестностями.

10.2.Непрерывные отображения топологических пространств и свойства непрерывных отображений. Гомеоморфизм топологических пространств.

Введем понятие непрерывного отображения, установим его связь с понятием непрерывной числовой функции, дадим строгие доказательства привычных свойств непрерывных отображений, приведем пример категории, исследуем понятие гомеоморфизма и дадим наглядные примеры гомеоморфных и локально гомеоморфных пространств.

Непрерывным называется отображение $f:(X,\tau)\rightarrow(Y,\sigma)$ топологических пространств, при котором выполнено следующее свойство: прообраз $f^{-1}(U)$ каждого открытого в пространстве Y подмножества $U\in\sigma$ является открытым подмножеством в пространстве X. Равносильное определение: непрерывным называется отображение, при котором прообразы замкнутых множеств замкнуты.

Упражнение. Проверьте эквивалентность двух определений непрерывного отображения.

Также нам потребуется определение отображения, непрерывного в точке.

Отображение $f:(X,\tau)\rightarrow(Y,\sigma)$ топологических пространств называется *непрерывным в точке* $x_0\in X$, если для всякой окрестности V точки $f(x_0)$ существует такая окрестность U точки x_0, что $f(U)\subset V$. Сравним это общее определение с привычным определением непрерывной в точке числовой функции из курса математического анализа. Функция $f:D\rightarrow E$ непрерывна в точке $x_0\in D$, если выполнено условие:

$$\forall\varepsilon>0\exists\delta>0 \,||\, x-x_0|<\delta\Rightarrow| f(x)-f(x_0)|<\varepsilon.$$

Числовая прямая является метрическим пространством с метрикой вида $d(x, x_0)=|x-x_0|$, поэтому, выбрав окрестности точек в виде открытых шаров (которые в данном случае являются открытыми интервалами), получим определение функции, непрерывной в точке. доказать, что отображение топологических пространств $f:(X,\tau) \rightarrow (Y, \sigma)$ непрерывно тогда и только тогда, когда оно непрерывно в каждой точке. В самом деле, если отображение $f:X\rightarrow Y$ непрерывно, а V- окрестность точки $f(x)$, то ее прообраз $f^{-1}(V)$ является окрестностью точки x, причем $f(f^{-1}(V))\subset V$. Обратно, если отображение f непрерывно в каждой точке и V- открытое множество в пространстве Y, то каждая точка множества $f^{-1}(V)$ обладает окрестностью, образ которой принадлежит множеству V. Таким образом, прообраз $f^{-1}(V)$ является объединением открытых подмножеств, следовательно, он открыт в пространстве X. Теперь обратимся к простейшим свойствам непрерывных тображений топологических пространств.

Теорема 1. *Тождественное отображение топологического пространства в себя непрерывно.*

Для доказательства рассмотрим тождественное отображение $id_X:X\rightarrow X$, переводящее каждую точку пространства X в себя. Пусть $U\in\tau$ - открытое подмножество, тогда, очевидно, $id_X^{-1}(V)=V$ и, следовательно, $id_X^{-1}(V)\in\tau$, что и требовалось.

Теорема 2. *Для любого пространства* (X,τ) *постоянное отображение*

$c:X\to\{\bullet\}$, где $\{\bullet\}$- одноточечное пространство, непрерывно. Доказать это свойство читателю предлагается в качестве *упражнения*.

Теорема 3. *Отображение вложения подпространства в топологическое пространство непрерывно.*

Действительно, рассмотрим вложение подпространства $i:Y\to X$, и пусть $V\in\tau$ - произвольное открытое в X подмножество. Тогда $i^{-1}(V)=V\bigcap i(Y)$ - открытое в Y подмножество, согласно определению топологии подпространства.

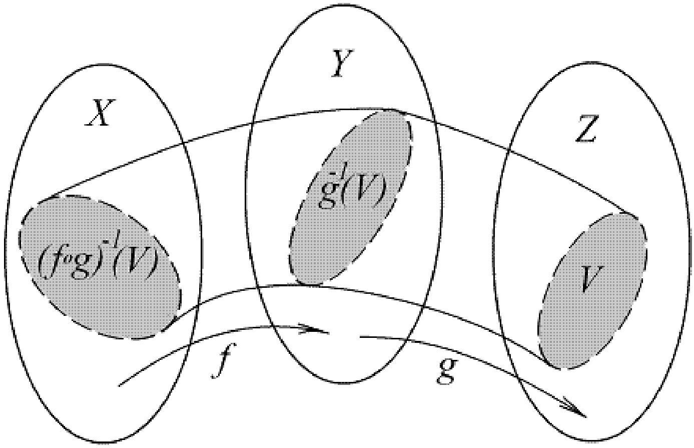


Рис.3. К доказательству теоремы о композиции

Теорема 4. *Композиция непрерывных отображений топологических пространств непрерывна* (рис.3).

Для доказательства рассмотрим непрерывные отображения топологических пространств $f:X\to Y$ и $g:Y\to Z$ и их композицию

$$X \xrightarrow{f} Y \xrightarrow{g} Z.$$

Пусть V- произвольное открытое подмножество пространства Z. Его прообраз относительно композиции $g\circ f$ равен

$$(g\circ f)^{-1}(V)=f^{-1}(g^{-1}(V)).$$

По определению непрерывного отображения, подмножество $g^{-1}(V)$ открыто в пространстве Y, а подмножество $f^{-1}(g^{-1}(V))$ открыто в пространстве X.

Совокупность подмножеств $\{V_\alpha|\alpha\in\mathrm{A}\}$ называется *покрытием множества* X, если $\bigcup_{\alpha\in\mathrm{A}}V_\alpha=X$. Здесь символом А обозначено множество индексов, нумерующих элементы покрытия V_α, при этом мощность покрытия может быть счетной (конечной или бесконечной) или несчетной. Покрытие топологического пространства X называют *открытым*, если все составляющие его подмножества V_α открыты в X. Топологическое пространство X называется *компактным*, если всякое его открытое покрытие содержит конечное покрытие.

Ясно, что конечное множество, наделенное дискретной топологией, компактно, а бесконечное множество с дискретной топологией некомпактно. Менее тривиальный пример некомпактного топологического пространства - евклидово вещественное пространство R^n, топология которого задана с помощью метрики

$$d(\overline{x},\overline{y})=\sqrt{(x_1-y_1)^2+...+(x_n-y_n)^2}$$

Для доказательства некомпактности пространства R^n достаточно указать одно его открытое покрытие, которое не содержит конечного подпокрытия. Такое покрытие образуют открытые шары радиуса $\sqrt{n}$, центры которых расположены во всех точках с целочисленными координатами. Действительно, такие шары образуют покрытие, причем любой конечный набор таких шаров покрытия не образует.

Теорема 5. *Пусть $f{:}X\to Y$ - сюръективное непрерывное отображение топологических пространств. Если X - компактное топологическое пространство, то и Y компактно.*

Действительно, пусть Δ - открытое покрытие пространства Y. Множества $f^{-1}(V)$ для $V\in\Delta$ составляют открытое покрытие пространства X, и если $U_1,...,U_s$ - его конечное подпокрытие, то f- $f(U_1),..., f(U_s)$ - конечное подпокрытие покрытия Δ пространства Y.

Теперь обратимся к понятию *категории.*

Категория состоит из двух классов - *класса объектов* и *класса морфизмов,* - удовлетворяющих требованиям :

1. Для каждой пары объектов X,Y определено множество $Mor(X,Y)$

морфизмов из X *в* Y;

2. Для каждой тройки объектов X,Y,Z определено отображение

$$\mu{:}Mor(X,Y)\times Mor(Y,Z)\rightarrow Mor(X,Z)$$

ставящее в соответствие двум морфизмам $f{:}X\rightarrow Y$ и $g{:}Y\rightarrow Z$ их *композицию* $\mu(f,g)=g\circ f{:}X\rightarrow Z$.

3. Множества $Mor(X,Y)$ и композиция морфизмов удовлетворяют аксиомам:

а) композиция ассоциативна: для каждой тройки морфизмов

$$T\xrightarrow{f}X\xrightarrow{g}Y\xrightarrow{h}Z$$

имеет место равенство $h\circ(g\circ f)=(h\circ g)\circ f$;

б) для каждого объекта X существует *тождественный морфизм* $id_X{:}X\rightarrow X$, такой, что любых объектов T и R и для любых морфизмов $f{:}T\rightarrow X$ и $g{:}X\rightarrow R$ имеют место равенства $id_X\circ f=f$, $g\circ id_X=g$;

в) если пары (X,Y) и (X',Y') различны, то пересечение множеств $Mor(X,Y)$ и $Mor(X',Y')$ пусто.

Рассматривая в качестве объектов все топологические пространства (то есть все множества со всевозможными топологиями на них) и в качестве морфизмов все возможные непрерывные отображения между топологическими пространствами, получим *категорию* $\mathcal{T}op$ *топологических пространств (и непрерывных отображений)*. Нам также знакомы категории множеств (и их отображений), векторных пространств (и их линейных отображений), групп (и их гомоморфизмов) и колец (и их гомоморфизмов).

Отображение $f{:}(X,\tau)\rightarrow(Y,\sigma)$ топологических пространств называется *гомеоморфизмом*, если оно

1) непрерывно,

2) взаимно однозначно,

3) обладает непрерывным двусторонним обратным отображением f^{-1}, то есть таким непрерывным отображением $f^{-1}{:}\ Y\rightarrow X$, что

$$f\circ f^{-1}=id_Y,\quad f^{-1}\circ f=id_X$$

Для любого неодноточечного пространства (X,τ) постоянное отображение $c{:}X\rightarrow\{\bullet\}$ не является гомеоморфизмом (оно не взаимно однозначно и тем более не обратимо).

Из непрерывности обратимого отображения $f:X \to Y$ не следует непрерывность обратного отображения f^{-1}. Пусть $f:X \to X$ - тождественное отображение непустого множества X, наделенного дискретной топологией, в то же множество X, наделенное любой другой топологией. Такое отображение непрерывно (*проверьте!*). Обратное отображение f^{-1} не является непрерывным.

Приведем примеры гомеоморфных топологических пространств, известных из других курсов геометрии.

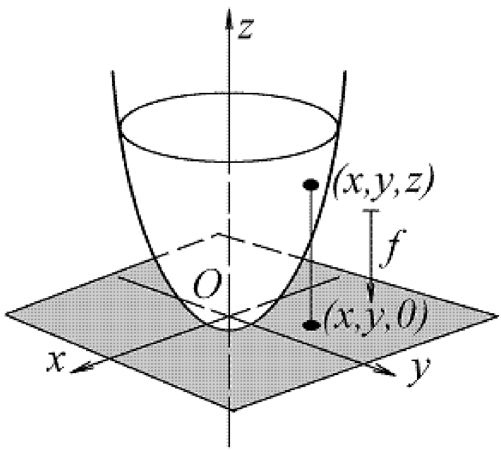


Рис.4. Пример гомеоморфизма

1. Параболоид вращения (рис. 4), заданный в трехмерном евклидовом пространстве уравнением $2pz=x^2+y^2$, гомеоморфен евклидовой плоскости.

Для построения отображения, осуществляющего гомеоморфизм, будем считать, что плоскость задана уравнением z=0 (рис. 4). Отображение f - проектирование параболоида на плоскость xOy параллельно координатной оси Oz- задается в координатах следующим образом:

$$f:(x,y,z) \to (x,y,0).$$

Обратное отображение плоскости на параболоид имеет вид:

$$f^{-1}:(x,y,0)\to(x,y,(x^2+y^2)/2p).$$

Оба отображения заданы непрерывными функциями и, следовательно, непрерывны, что и доказывает гомеоморфизм.

2. Трехосный эллипсоид гомеоморфен сфере.

Каноническое уравнение трехосного эллипсоида имеет вид:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1.$$

Уравнение сферы с центром в начале координат:

$$x^2+y^2+z^2=R^2.$$

Отображение эллипсоида на сферу, осуществляющее гомеоморфизм, задается в координатах следующим образом:

$$(x,y,z)\to\left(\frac{R}{a}x,\frac{R}{b}y,\frac{R}{c}z\right).$$

Гомеоморфизм устанавливает взаимно однозначное соответствие между открытыми подмножествами гомеоморфных топологических пространств. В самом деле, в силу непрерывности отображения $f:(X,\tau_X)\to(Y,\tau_Y)$, осуществляющего гомеоморфизм, прообраз любого открытого в Y множества открыт, то есть f непрерывно

$$\Rightarrow \forall\, U\in\tau_Y\, f^{-1}(U)\in\tau_X.$$

Аналогично, в силу непрерывности обратного отображения прообраз при отображении f^{-1} любого открытого в X множества также открыт:

$$f^{-1}\ \text{непрерывно} \Rightarrow \forall\, V\in\tau_X\ (f^{-1})^{-1}(V)=f(V)\in\tau_Y.$$

Доказанное означает, что если некоторое топологическое утверждение верно для одного из двух гомеоморфных топологических пространств, то такое же утверждение верно для другого. Гомеоморфные пространства топологически неразличимы, поэтому о них говорят, что они имеют один топологический тип. Отношение гомеоморфности в классе топологических пространств рефлексивно, симметрично и транзитивно, то есть является отношением эквивалентности (докажите!) Отображение $f:X \to Y$ топологических пространств называется *локальным гомеоморфизмом,* если у каждой точки p пространства X имеется такая окрестность U_p, что ограничение $f|U_p:U_p \to Y$отображения f осуществляет гомеоморфизм окрестности U_p на ее образ $f(U_p) \subset Y$. При этом пространства X и Y называются *локально гомеоморфными*. Топологические пространства могут быть негомеоморфными, но локально гомеоморфными. Таковы, например, тор (компактное пространство) и евклидова плоскость (некомпактное пространство.)

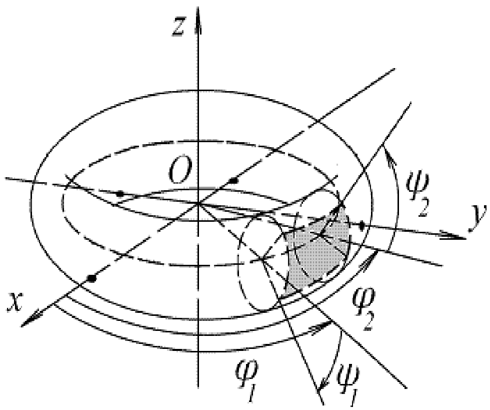


Рис.5. Пример локального гомеоморфизма

Действительно, представим тор как поверхность, образованную вращением окружности около оси, лежащей в плоскости окружности и не пересекающей ее. Тогда точки тора снабжаются угловыми координатами, например, так, как показано на рис.5. Положительные углы φ отсчитываются от положительного направления оси Ox в сторону положительного направления оси Oy, положительные углы ψ отсчитываются от плоскости xOy в сторону положительного направления оси Oz. Тогда область, соответствующая значениям угловых координат $\varphi_1<\varphi < \varphi_2$ и $\psi_1<\psi<\psi_2$, гомеоморфна прямоугольнику (рис.6).

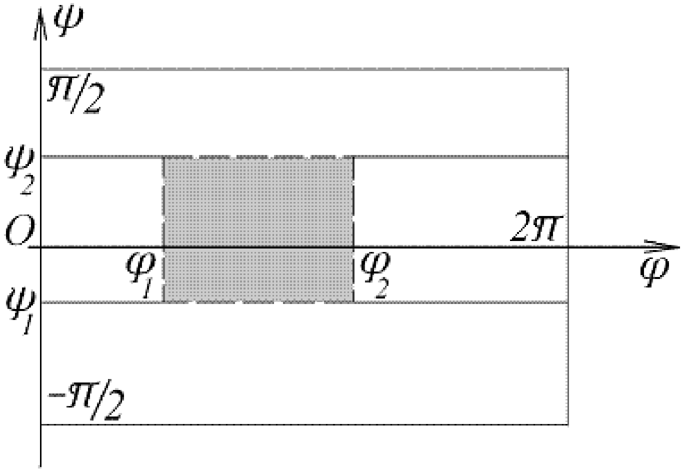


Рис.6. Пример локального гомеоморфизма

Локальные гомеоморфизмы чрезвычайно важны в дифференциальной геометрии, а точнее - при построении дифференциального и интегрального исчислений на геометрических объектах, отличных от евклидова пространства R^n (например, прямой R или плоскости R^2).

10.3. Топологические и дифференцируемые многообразия

(Вещественным) топологическим многообразием (без края) называется хаусдорфово топологическое пространство (X,τ) со счетной базой, каждая точка которого обладает окрестностью, гомеоморфной евклидову пространству R^n. Число n называется *размерностью* топологического многообразия X. Иногда удобнее использовать равносильную форму определения многообразия.

(Вещественным) топологическим многообразием (без края) называется хаусдорфово топологическое пространство (X,τ) со счетной базой, локально гомеоморфное евклидову пространству R^n. Важно, что топологические многообразия локально устроены как евклидовы пространства подходящей размерности. Однако можно построить топологические пространства, удовлетворяющие требованиям хаусдорфовости, локально гомеоморфные евклидову пространству, но

не являющиеся многообразиями. Простейший пример такого топологического пространства поставляет объединение всех параллельных прямых на евклидовой плоскости, заданных уравнениями $y=\alpha$, где α - иррациональное число. Каждая из этих прямых удовлетворяет всем требованиям для топологического многообразия. Объединение всех прямых с иррациональными ординатами является хаусдорфовым пространством, локально гомеоморфным числовой прямой $R=R^1$. Однако на таком объединении невозможно ввести счетную базу топологии. Чтобы убедиться в этом, зафиксируем на каждой из прямых какую-нибудь счетную бесконечную базу. Очевидно, объединение баз по всем прямым рассматриваемого набора несчетно, так как его мощность не меньше мощности множества иррациональных чисел. Простейшими примерами топологических многообразий являются числовая прямая с классической топологией, окружность и евклидова плоскость. Более содержательный пример - двумерный тор $T^2=S^1\times S^1$. Ранее был рассмотрен пример локального гомеоморфизма тора и евклидовой плоскости. Тор - хаусдорфово пространство (рис.7).

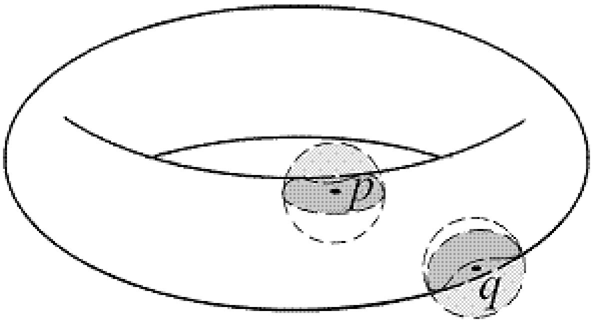


Рис.7. Тор как хаусдорфово пространство

Для поверки этого факта рассмотрим две различные точки p и q тора, вложенного в трехмерное евклидово пространство R^3 способом, описанным ранее. Пространство R^3, как известно, является метрическим. Поэтому в качестве непересекающихся окрестностей точек p и q в пространстве R^3 можно выбрать открытые шары достаточно малого радиуса r. Тогда искомые непересекающиеся окрестности точек p и q на торе представлены его пересечениями с открытыми шарами.

Тор - топологическое пространство, допускающее счетную базу. Для ее построения зафиксируем счетную базу классической топологии в трехмерном евклидовом пространстве. Затем в качестве множеств, образующих базу на торе, возьмем пересечения тора и множеств базы топологии трехмерного пространства.

Упражнение. Пользуясь определением, убедитесь в том, что такие пересечения действительно образуют базу топологии тора.

Построенная таким способом база топологии тора имеет мощность, не превосходящую мощность базы топологии трехмерного евклидова пространства, то есть является счетной. Таким образом, двумерный тор - хаусдорфово пространство со счетной базой, локально гомеоморфное евклидовой плоскости, т. е. он является топологическим многообразием размерности два.

(Вещественным) топологическим многообразием с краем называется хаусдорфово топологическое пространство (X,τ) со счетной базой, каждая точка которого принадлежит одному из двух следующих классов:
1) класс точек, каждая из которых обладает окрестностью, гомеоморфной евклидову пространству R^n;
2) класс точек, каждая из которых обладает окрестностью, гомеоморфной полупространству

$$\mathbb{R}^n_+ = \{(x_1, x_2, \dots x_n) \in \mathbb{R}^n | x_1 \geq 0\}$$

евклидова пространства R^n.

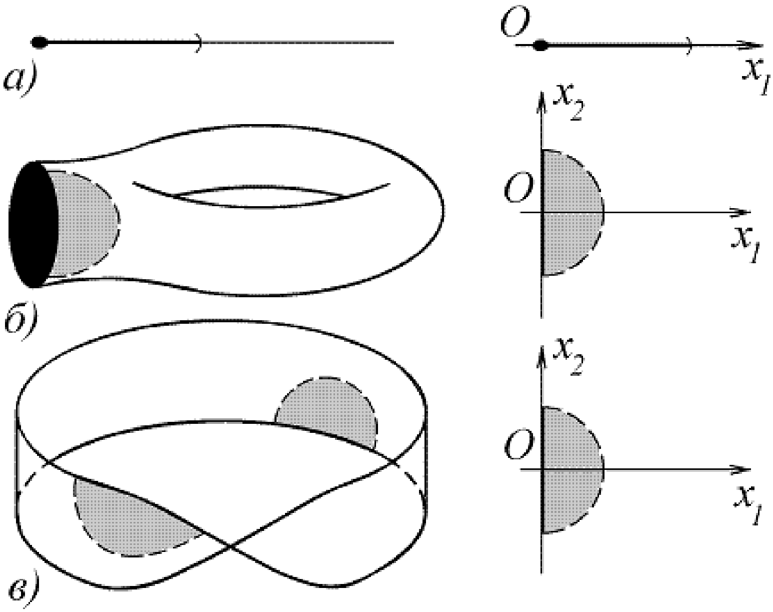


Рис.8. Примеры многообразий с краем: а) полупрямая, б) тор с удаленным диском, в) лента Мебиуса

Точки первого класса называются *внутренними точками* многообразия *X*, точки второго класса - *точками границы* или *точками края*. Число *n* называется *размерностью* топологического многообразия *X*.

Примерами одномерных многообразий с краем служат отрезок и полупрямая, примерами двумерных - круг, тор с удаленным диском и лента Мебиуса (рис.8). Границы трех последних многообразий гомеоморфны окружности, в то время как сами многообразия негомеоморфны.

Для конструирования топологических многообразий часто используется операция *приклеивания*, заключающаяся в следующем. Пусть X и Y- топологические многообразия, $Z_X \subset X$, $Z_Y \subset Y$ - одмножества, причем Z_X гомеоморфно Z_Y. Пусть $f: Z_X \rightarrow Z_Y$ - гомеоморфизм. Отождествляя каждую точку $x \in Z_X$ с ее образом

$f(x) \in Z_Y$, получим новое многообразие $X \coprod_f Y$, называемое *склейкой многообразий X и Y вдоль гомеоморфизма f*.

Построим с помощью этой операции двумерное многообразие, играющее важную роль в различных областях математики - вещественную проективную плоскость RP^2 (рис.9).

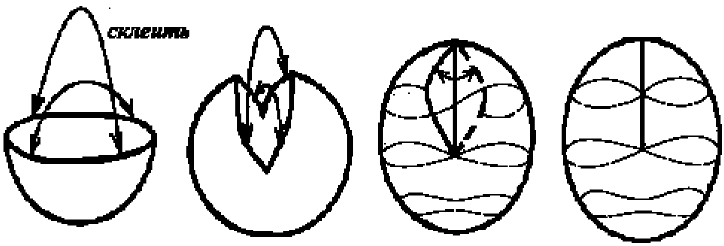


Рис.9. Вещественная проективная плоскость

Построение начинается с полусферы, например, заданной условиями

$$x^2 + y^2 + z^2 = 1, \quad z \leq 0.$$

Ее границей является окружность

$$x^2 + y^2 = 1, \quad z = 0,$$

получаемая в пересечении сферы плоскостью *xOy* (экватор). Рассмотрим гомеоморфизм границы полусферы на себя, задаваемый центральной симметрией относительно начала координат

$$f : (x, y, 0) \mapsto (-x, -y, 0).$$

Вещественная проективная плоскость получается склеиванием экватора полусферы вдоль гомеоморфизма f (рис.9). Для получения другой интерпретации вещественной проективной плоскости рассмотрим трехмерное аффинное пространство. На множестве всех прямых этого пространства определено отношение параллельности, которое, как известно, является отношением эквивалентности. Тогда имеется разбиение множества всех прямых трехмерного пространства на классы параллельности. Каждый такой класс состоит из (бесконечного!) набора всех прямых пространства, имеющих данное направление. Класс параллельных прямых может быть задан направляющим вектором прямой этого класса. Выбор направляющего вектора не однозначен; все направляющие векторы данной прямой (и данного класса параллельных прямых) коллинеарны и имеют ненулевую длину. Тогда каждый класс параллельных прямых задается *направлением*, то есть *классом ненулевых коллинеарных векторов*.

Проективная плоскость - это многообразие, точки которого соответствуют классам параллельных прямых трехмерного аффинного пространства или, равносильно, классам коллинеарности ненулевых векторов трехмерного векторного пространства. Зафиксируем в нем базис; тогда каждый вектор $\vec{a} \neq \vec{0}$ задается координатами $\vec{a} = (a_0, a_1, a_2)$, причем хотя бы одна из координат отлична от нуля.

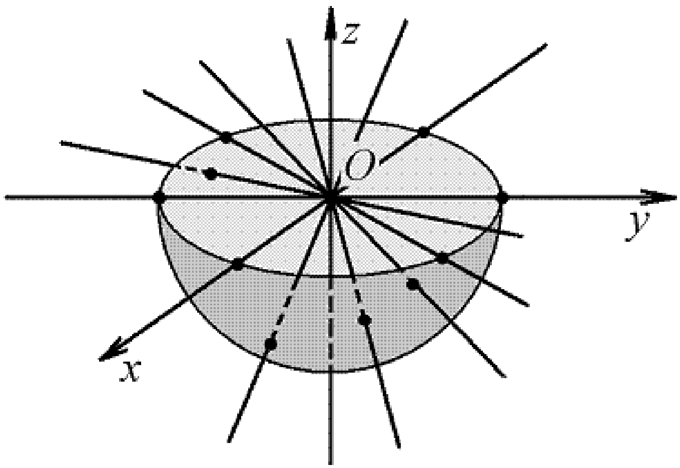


Рис.10. Соответствие между направлениями и точками проективной плоскости

Векторы, принадлежащие одному направлению, имеют пропорциональные координаты, поэтому направление может быть задано способом, известным из курса проективной геометрии как *однородные координаты*. Однородные координаты a_0: a_1: a_2 - это класс пропорциональных упорядоченных троек чисел, хотя бы одно из которых отлично от нуля. Каждый из таких классов определяет направление и, следовательно, точку проективной плоскости. Это соответствие представлено на рис.10. Направления изображены в виде прямых, проходящих через начало координат и пересекающих полусферу, из которой проективная плоскость получалась путем склейки границы. Каждая из прямых, не принадлежащих плоскости экваториального сечения (плоскости *xOy*), пересекает полусферу в единственной точке и таким образом выделяет точку проективной плоскости, отвечающую данному направлению. Каждая прямая, принадлежащая плоскости экваториального сечения, пересекает полусферу в двух точках границы, отождествляемых склейкой, и, тем самым, также выделяет единственную точку проективной плоскости.

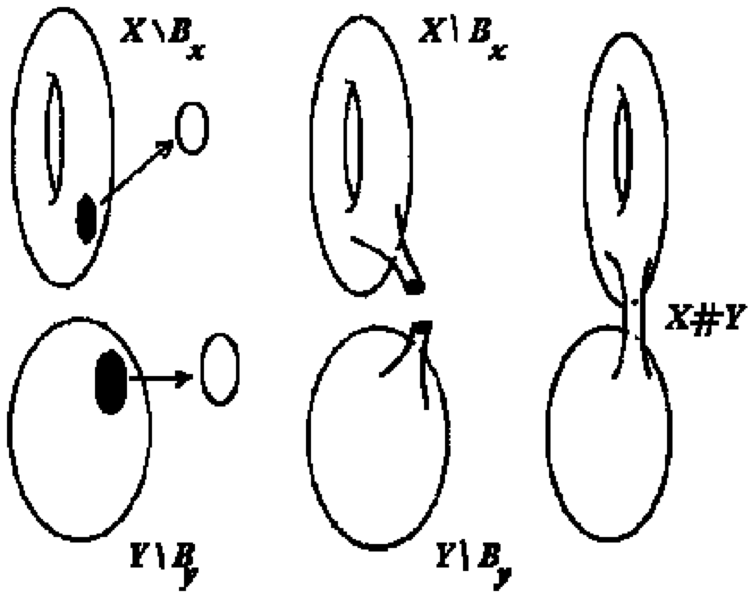


Рис.11. Связная сумма топологических многообразий

Для построения *связной суммы многообразий X и Y* одинаковой размерности n выберем окрестности B_x и B_y точек $x \in X$ и $y \in Y$ соответственно, гомеоморфные открытому n-мерному шару. Очевидно, границы многообразий $X \setminus B_x$ и $Y \setminus B_y$ гомеоморфны (n-1-мерной сфере). Пусть f - какой-нибудь гомеоморфизм границ. Тогда связная сумма

$X\sharp Y$ многообразий X и Y определяется как их склейка вдоль гомеоморфизма f (рис.11).

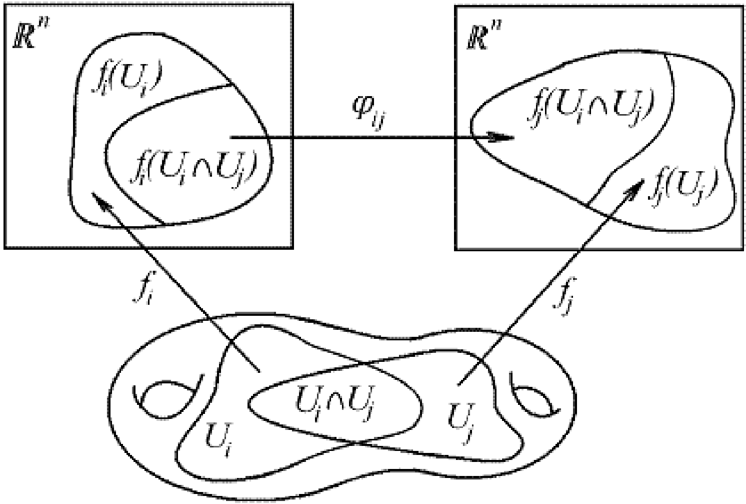


Рис.12. Координатные отображения и отображения склейки

Топологическое многообразие X называется *дифференцируемым класса* C^k (или C^∞), если гомеоморфные отображения f_i*:*$U_i \to R^n$ подчинены следующему условию: для любых i, j ограничения отображений f_i и f_j *согласованы* в следующем смысле: определено отображение

$$\varphi_{ij} := f_j \circ (f_i|_{U_i \cap U_j})^{-1} : f_i(U_i \cap U_j) \to f_j(U_i \cap U_j),$$

задаваемое функциями, имеющими непрерывные частные производные порядка k (соответственно, имеющими непрерывные частные производные любого порядка), такое, что

$$f_j|(U_i \cap U_j) = \varphi_{ij} \circ (f_i|(U_i \cap U_j))$$

(рис.12). Отображения f_i называются *координатными отображениями*, пара (U_i, f_i) - *картой*, а соответствующий покрытию

$\{U_i\}_{i \in I}$ набор $\{(U_i, f_i)\}$ - *атласом* дифференцируемого многообразия X. Отображения φ_{ij} иногда называют *отображениями склейки*.

Многообразия класса C^∞ называют также *аналитическими*.

В качестве примера снабдим двумерный тор структурой дифференцируемого многообразия. Выше было доказано локальный гомеоморфизм тора и евклидовой плоскости. Атлас на торе может быть построен выбором различных начал отсчета угловых координат φ и ψ. Рассмотрим построение атласа на торе подробнее. Назовем *параллелью* геометрическое место точек тора, имеющих одно и то же значение угловой координаты ψ; *меридианом* - геометрическое место точек тора, имеющих одно и то же значение угловой координаты φ, т. е. любую его образующую.

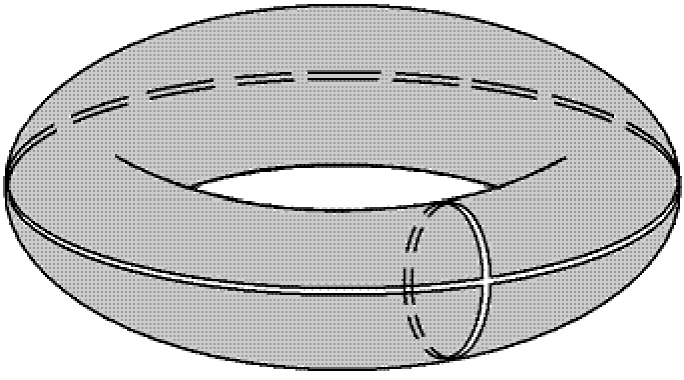


Рис.13. Выбор координатной карты на торе

Разрезав тор вдоль выбранных параллели и меридиана (рис.13), получим множество точек, гомеоморфное прямоугольнику, называемому разверткой тора. Введем (новые) координаты на развертке так, что $\psi = \pm \pi$ соответствует параллели разреза, $\varphi = 0$ и $\varphi = 2\pi$ - меридиану разреза. Прямоугольник $0 < \varphi < 2\pi$, $-\pi < \psi < \pi$ с гомеоморфизмом в тор образует координатную карту.

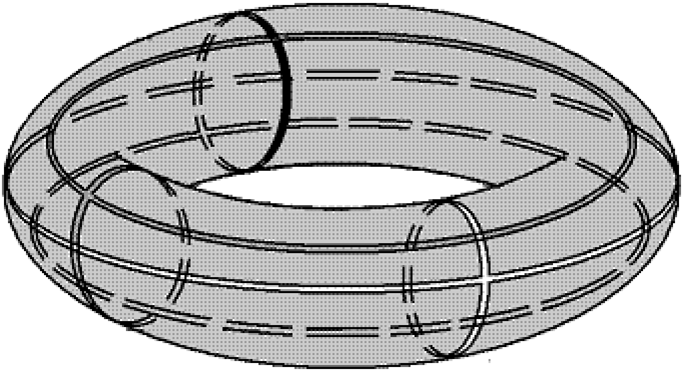


Рис.14. Атлас на торе

Выберем три пары параллель-меридиан, причем так, что все три параллели и все три меридиана различны (рис.14). Каждая такая пара задает способ отсчета угловых координат на торе только что описанным образом, причем каждая точка тора покрыта хотя бы одной из трех карт. Пусть U_1 и U_2 - две построенные таким способом координатные карты, и точка с координатами ${}_1\varphi=0$, ${}_1\psi=0$ в карте U_1 имеет в карте U_2 координаты ${}_2\varphi=\varphi_0$, ${}_2\psi=\psi_0$. Тогда имеют место уравнения перехода от координат ${}_1\varphi$, ${}_1\psi$ в карте U_1 к координатам ${}_2\varphi$, ${}_2\psi$ в карте U_2:

$${}_2\varphi = {}_1\varphi + \varphi_0, \quad {}_2\psi = {}_1\psi + \psi_0.$$

Очевидно, что отображение склейки, заданное этими уравнениями, имеет производные любого порядка, и построенная нами структура вещественного многообразия на двумерном торе является, согласно определению, структурой аналитического многообразия.

Структуры дифференцируемых многообразий нетрудно ввести, например, на сфере, параболоиде вращения и других известных поверхностях второго порядка.

Упражнение. С помощью однородных координат построим локальные карты на проективной плоскости. Карта ${}^0U$ соответствует множеству точек с ненулевой координатой a_0; координаты в этой карте имеют вид

$$^0x = a_1/a_0,\ ^0y = a_2/a_0.$$

Аналогично,

$$^1U:\ a_1 \neq 0,\ ^1x = a_0/a_1,\ ^1y = a_2/a_1.$$

Также,

$$^2U:\ a_2 \neq 0,\ ^1x = a_0/a_2,\ ^1y = a_1/a_2.$$

Выпишите функции перехода и убедитесь в том, что вещественная проективная плоскость RP^2- аналитическое многообразие.

11. Топология оболочек

11.1. Описание геометрических объектов

Геометрические объекты. Мы будем интересоваться формой окружающих предметов, их размерами и взаимным расположением, не вдаваясь в подробности физических свойств. Другими словами, мы будем изучать и моделировать геометрические свойства реальных или воображаемых объектов. Нашей конечной целью является построение математических моделей геометрии этих объектов. Эти модели нужны для принятия решений, для проведения исследований, для производства материальных ценностей.

Геометрическое моделирование изучает методы построения математической модели, описывающей геометрические свойства предметов окружающего мира. Оно базируется на аналитической и дифференциальной геометрии, вычислительной математике, вариационном исчислении, топологии и разрабатывает собственные математические методы моделирования.

Инструментом для геометрического моделирования служат математические методы решения тех или иных задач. Используемые методы позволят описывать геометрические свойства предметов, создавать их математические модели и **исследовать их путем проведения различных расчетов и численных экспериментов**. При необходимости мы сможем редактировать моделируемые объекты и строить их графические отображения.

Для описания геометрических свойств окружающих предметов мы будем строить твердые тела или просто тела. **Тела в свою очередь мы будем описывать точками, линиями и поверхностями.** Все они обладают определенными общими свойствами, поэтому ими можно оперировать как объектами. Точки, линии, поверхности и тела будем называть геометрическими объектами (рис. 1).

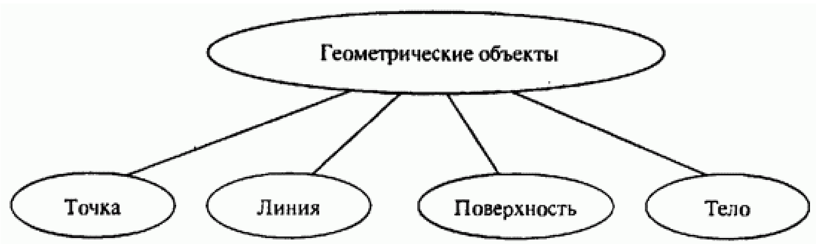


Рис. 1

Геометрические объекты будут служить основными элементами математической модели геометрии реальных или воображаемых объектов. Мы будем строить их в трехмерном евклидовом пространству считая их неизменными во времени.

В большинстве случаев мы будем использовать декартовы прямоугольные системы координат. **В декартовой системе координат базисные векторы имеют одинаковую длину и постоянное направление в любом месте пространства.** Это упрощает описание объектов, так как базисные векторы при дифференцировании выступают в роли констант. Мы рассмотрим также описание геометрических объектов в криволинейных системах координат.

Обозначения. Для количественных характеристик геометрических объектов мы будем использовать скалярные величины, векторы, а также тензоры. Скалярные величины будем обозначать строчными буквами латинского или греческого алфавита. Векторы в пространстве будем обозначать строчными буквами латинского алфавита, выделенными жирным шрифтом. Двухмерные векторы будем обозначать строчными буквами латинского алфавита, выделенными жирным курсивом. Точки будем обозначать прописными латинскими буквами. Тензоры и матрицы будем обозначать прописными буквами латинского алфавита, выделенными жирным шрифтом.

Систему координат с началом в точке О и базисными векторами $\mathbf{e}_1, \mathbf{e}_2$, $\mathbf{e}_3$ будем обозначать через $O\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$.

Для векторов в пространстве мы будем использовать записи типа

$$\mathbf{r} = r_1\mathbf{e}_1 + r_2\mathbf{e}_2 + r_3\mathbf{e}_3 = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = [r_1\ r_2\ r_3]^{\top},$$

где r_1, r_2, r_3— компоненты вектора $\mathbf{r}$. В данной главе нам удобно использовать обозначения компонент векторов с индексами, равными ее номеру. В других главах мы будем также использовать обозначения компонент векторов через x, y, z .

Для описания сложных геометрических объектов нам потребуются векторы как в трехмерном, так и в двухмерном пространстве, например, на области параметров поверхности. Для двухмерных векторов мы будем использовать запись типа

$$\boldsymbol{p} = p_1\boldsymbol{e}_1 + p_2\boldsymbol{e}_2 = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = [p_1\ p_2]^{\top},$$

где p_1, p_2 — компоненты вектора $\mathbf{p}$.

Операцию скалярного произведения векторов будем обозначать точкой: например,

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3.$$

Символом × будем обозначать операцию векторного произведения векторов: например,

$$\mathbf{a} \times \mathbf{b} = (a_2 b_3 - a_3 b_2)\mathbf{e}_1 + (a_3 b_1 - a_1 b_3)\mathbf{e}_2 + (a_1 b_2 - a_2 b_1)\mathbf{e}_3.$$

Запись двух векторов рядом **ab** будет означать операцию диадного произведения векторов **a** и **b**, например,

$$\mathbf{ab} = \begin{bmatrix} a_1 b_1 & a_1 b_2 & a_1 b_3 \\ a_2 b_1 & a_2 b_2 & a_2 b_3 \\ a_3 b_1 & a_3 b_2 & a_3 b_3 \end{bmatrix}.$$

Точка. Геометрическое моделирование мы начнем изучать с простых объектов, переходя постепенно к более сложным. Точка R пространства в общем случае описывается координатами u^1, u^2, u^3 некоторой системы координат. В декартовой прямоугольной системе координат точку можно описать с помощью радиус-вектора $\mathbf{r} = [\, r_1, r_2, r_3 \,]^T$. Радиус-вектор определяет преобразование переноса, переводящее начальную точку декартовой системы координат в заданную точку пространства. Компоненты радиус-вектора точки равны ее координатам. Радиус-вектор в отличие от просто вектора связан с началом координат. Эта разница сказывается на формулах преобразования координат и на формулах изменения положения в пространстве.

11.2. Математическая модель геометрии объектов

Моделирование реального или воображаемого объекта представляет собой совокупность действий, которые позволяют создавать его математическую модель, редактировать ее, изменять ее положение и ориентацию в пространстве и обеспечивают взаимодействие с другими моделями. Взаимодействием мы называем выполнение различных операций над моделями: установление зависимости параметров одной модели от параметров других моделей, определение взаимного положения

моделей. Для выполнения этих действий необходима информация об объекте. Геометрическая информация об объекте может храниться в виде структуры данных или может вычисляться. Определим **математическую модель реального или воображаемого объекта как совокупность данных и функций, позволяющих получить необходимую информацию об объекте и изменять его модель требуемым образом** (рис.2).

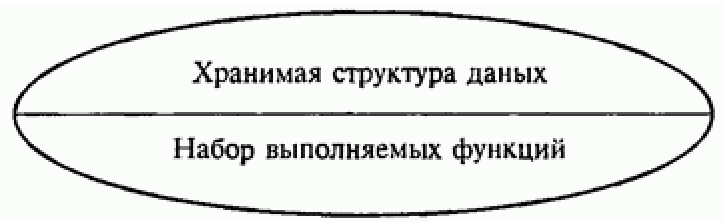


Рис. 2. Математическая модель

Программную реализацию структуры данных и функций называют классом.

Геометрические объекты будут иметь свои данные и свои функции.

Для построения точки в структуре ее данных достаточно хранить три координаты и иметь функции выполнения операций над радиус-векторами.

Для построения произвольной линии нужно знать зависимость ее радиус-вектора $\mathbf{r}(t)$ от параметра, область изменения параметра и иметь функции вычисления производных радиус-вектора.

Для построения поверхности нужно знать зависимость ее радиус-вектора $\mathbf{r}(u, v)$ от параметров, область изменения параметров и иметь функции вычисления частных производных радиус-вектора.

Тело мы будем описывать совокупностью ограничивающих его поверхностей, дополненной информацией об их связях друг с другом.

С математической точки зрения все геометрические объекты равноправны. Для них существует ряд общих выполняемых функций.

Все геометрические объекты могут быть подвержены модификациям сдвига, поворота, масштабирования, симметрии, поэтому они должны иметь функции, соответствующие этим действиям. Кроме того, для работы с геометрическим объектом нужны функции создания объекта (конструкторы), удаления объекта (деструктор), функция создания копии объекта, функции доступа к данным объекта, функции редактирования данных объекта. Математическая модель должна быть дополнена функциями, обеспечивающими взаимодействие объектов и выполнение над ними различных операций.

Мы рассмотрели теоретические основы геометрического моделирования. Дальнейшее изложение будет посвящено конкретным моделям и методам их построения.

11.3. Топологические объекты

Моделирование окружающих нас предметов требует привлечения более сложных геометрических объектов, чем точки, кривые и поверхности. Одной поверхностью в общем случае невозможно описать геометрическую форму некоторого заданного предмета, но это можно сделать с помощью нескольких поверхностей, связав их определенным образом. То, каким образом поверхности будут связаны друг с другом, составляет дополнительную информацию, которой мы снабдим новые объекты, построенные на основе уже рассмотренных геометрических объектов. **Нашей целью является получение геометрической модели окружающих предметов**. Все эти предметы занимают некоторую часть пространства, или, другими словами, занимают конечный объем пространства. Для их моделирования нужно описать совокупность поверхностей, отделяющих внутренний объем предмета от остальной части пространства. Для этого потребуется набор определенным образом построенных и обрезанных поверхностей и информация о взаимной связи этих поверхностей — как одна поверхность переходит в другую.

Оболочки.

Поверхности могут быть замкнутыми по одному или двум параметрическим направлениям или незамкнутыми. Незамкнутые поверхности имеют границу. **Границей** будем назвать линию на поверхности, соответствующую движению ее параметров по границе их области определения. Линию на замкнутой поверхности, по которой она замыкается сама на себя, будем называть **швом**. Поверхности могут стыковаться друг с другом по границам. Можно сказать, что по шву замкнутая поверхность стыкуется сама с собой. Совокупность стыкующихся по границам поверхностей будем называть **оболочкой**. Оболочка может состоять из одной поверхности или нескольких поверхностей. Также как и отдельная поверхность, оболочка может быть замкнутой и незамкнутой. Замкнутая оболочка не имеет границы. Незамкнутая оболочка имеет одну или несколько границ.

В данной главе нас будут интересовать свойства геометрических объектов, не зависящие от количественных характеристик. Мы будем рассматривать непрерывную связь между точками геометрических объектов. Предположим, что оболочка выполнена из эластичного неразрываемого и не склеиваемого материала. Исследуем свойства этой оболочки, которые сохраняются при всевозможных ее деформациях. **Деформацией** будем называть изменение формы оболочки путем растяжения, сжатия, сдвига или изгиба ее поверхности, не приводящее к разрывам и не требующее склеивания поверхностей оболочки. Эластичная оболочка в виде куба может быть деформирована в сферу, или эллипсоид, или оболочку в виде тетраэдра, но не может быть деформирована в тороидальную оболочку.

Сфера, эллипсоид, оболочка в виде тетраэдра или куба могут быть преобразованы друг в друга путем непрерывных и обратимых отображений. Свойства геометрических объектов, сохраняющиеся при непрерывных и обратимых отображениях одного пространства в другое, изучает **<u>топология</u>**. **С топологической точки зрения сфера, эллипсоид, оболочка в виде тетраэдра или куба <u>эквивалентны</u>.** Свойства, характеризующие непрерывность точек некоторой оболочки, являются топологическими свойствами. Несмотря на кажущуюся неопределенность, топологические свойства геометрических объектов связаны с фундаментальными математическими понятиями.

Топология изучает общий случай оболочек, которые могут самопересекаться, иметь или не иметь границы, уходить в бесконечность. Топология оперирует своими объектами, которые несут информацию о их взаимной связи друг с другом, и устанавливает между ними соотношения. При моделировании окружающих нас объектов мы будем строить оболочки из топологических объектов. Они будут нести и количественную геометрическую информацию и топологическую информацию. Количественная геометрическая информация топологического объекта содержится в его геометрическом носителе, которым может являться точка, кривая или поверхность. В данной главе мы сосредоточим внимание на топологических свойствах моделируемых объектов.

Вершины, ребра, циклы, грани.

Рассмотрим оболочки, построенные на основе поверхностей в трехмерном евклидовом пространстве. Для отслеживания связей составляющих оболочку поверхностей дополним поверхности информацией об этих связях и введем топологические объекты. Топологические объекты будут нести одновременно метрическую и топологическую информацию.

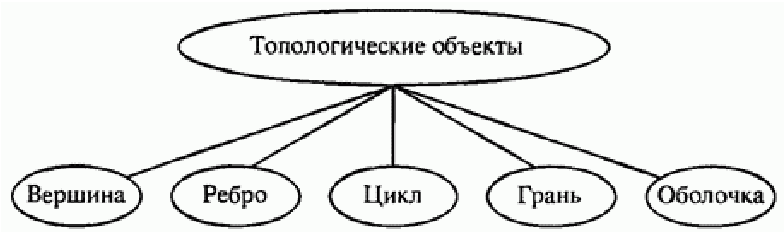


Рис.3. Топологические объекты

Одним из топологических объектов является оболочка. При построении оболочки будем использовать такие топологические объекты, как грани, ребра, вершины и циклы (рис.3). Все топологические объекты имеют общие принципы построения.

Гранью будем называть топологический объект, построенный на основе поверхности. Фактически грань представляет собой поверхность плюс информация о том, какая сторона поверхности является наружной стороной грани, и информация об ее положении в оболочке, т. е. информация об ее соседях. Информация о соседних гранях оформляется в виде циклов.

Цикл — это топологический объект, который описывает одну из границ грани, и содержит информацию о том, где и как к данной грани примыкают соседние грани. Так как вдоль одного цикла к данной грани могут примыкать несколько соседних граней, то цикл состоит из нескольких участков. Каждый участок цикла опирается на некоторое ребро.

Ребром будем называть топологический объект, построенный на основе линии стыковки соседних граней или на основе граничной линии оболочки. Грани стыкуются только по ребрам. Таким образом, каждая грань со всех сторон окружена ребрами. **Вершиной** будем называть топологический объект, построенный на основе точки, в которой стыкуются ребра. Вершины могут лежать только на краях ребер. Каждое ребро начинается и оканчивается в вершине. Если ребро замкнуто, то оно начинается и оканчивается в одной и той же вершине. Цикл состоит из ребер, образующих замкнутую линию вдоль одной из границ грани. Цикл всегда замкнут и ему приписывается определенное направление. Грань может содержать несколько циклов, причем один из них является внешним, а остальные — внутренними и целиком лежащими внутри внешнего цикла. За положительное направление цикла примем направление движения вдоль цикла, при котором грань всегда находится слева, если смотреть с наружной стороны грани.

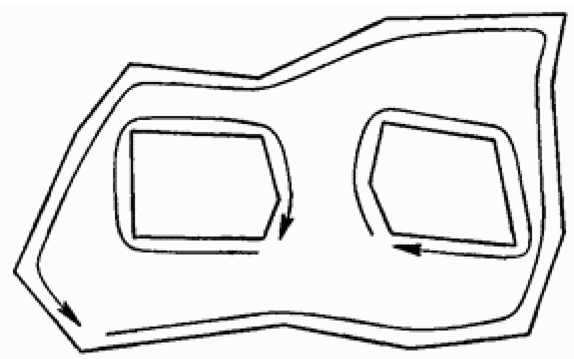


Рис.4. Ориентация внешнего и внутренних циклов грани

Таким образом, внешний цикл грани ориентирован против часовой стрелки, а внутренние циклы ориентированы по часовой стрелке, если смотреть с наружной стороны грани. Каждый цикл проходит по одной из границ поверхности. На рис. 4 приведен пример грани с ее циклами.

Грани, ребра и вершины строятся на базе известных геометрических объектов (точек, кривых и поверхностей) добавлением к ним информации о своих соседях и о взаимной ориентации. В результате геометрические объекты приобретают новое качество, чем и обусловлено введение топологических объектов.

11.4. Эйлерова характеристика оболочек

Топологические свойства оболочки могут быть выражены через количество ее граней, ребер, вершин и циклов. Пусть оболочка содержит F граней, E ребер, V вершин и L циклов. Число вершин, ребер, граней и циклов оболочки связаны между собой соотношением

$$F - E + V + (F - L) = H, \tag{1}$$

где величина H называется эйлеровой характеристикой оболочки. Формула (5.2.1) носит имя формулы Эйлера. Если каждая грань оболочки имеет один цикл, то F-L=0 и слагаемое в скобках в левой части (1) может быть опущено. Одну и ту же оболочку можно построить из различного набора граней. Например, сферическую оболочку можно построить из двух полусфер или из нескольких сферических сегментов подобно футбольному мячу. Как будет видно далее, эйлерова характеристика оболочки не зависит от числа и формы составляющих ее граней, но зависит от природных характеристик оболочки, которые изучает топология. Рассмотрим, как изменяется эйлерова характеристика оболочки при изменении составляющих ее элементов. Для этого будем изменять состав граней, ребер и вершин некоторого фрагмента оболочки, не изменяя состава остальной части оболочки.

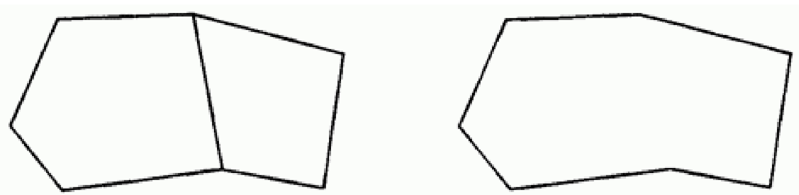


Рис.5. Эйлерова характеристика не изменяется при ликвидации ребра оболочки

При желании оболочкой можно считать только показанный на рисунках фрагмент. На рис. 5 показан фрагмент оболочки, у которого ликвидируется одно ребро. Из рисунка видно, что при ликвидации одного ребра число граней, число ребер и число циклов уменьшается на единицу, а эйлерова характеристика оболочки не изменяется.

Если объединить два ребра, ликвидировав общую для них вершину, то число ребер и вершин оболочки уменьшатся на единицу. Если разрезать ребро на две части, вставив вершину, то число ребер и вершин оболочки увеличится на единицу. Эйлерова характеристика оболочки в обоих случаях не изменится. При введении одного дополнительного ребра между существующими вершинами число граней, число циклов и число ребер увеличится на единицу.

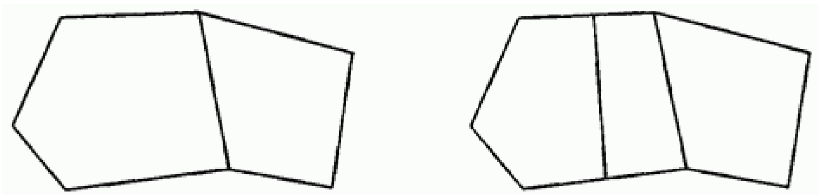


Рис.6. Эйлерова характеристика не изменяется при добавлении ребра в оболочку

При введении дополнительного ребра с добавлением двух новых вершин на его концах число граней и циклов увеличивается на единицу, число ребер увеличивается на три из-за деления двух существующих ребер на две части и добавления нового ребра, а эйлерова характеристика оболочки не изменяется, что показано на рис.6.

Эйлерова характеристика оболочки не изменяется и при введении одного дополнительного ребра, соединяющего существующую вершину и новую вершину, делящую существующее ребро на два ребра.

В приведенных примерах все грани не изменяли число ограничивающих их циклов, а все новые грани имели один цикл. В общем случае грань может иметь вырезы внутри. Грань с вырезами с топологической точки зрения отличается от грани без вырезов, так как первую нельзя преобразовать во вторую путем деформирования.

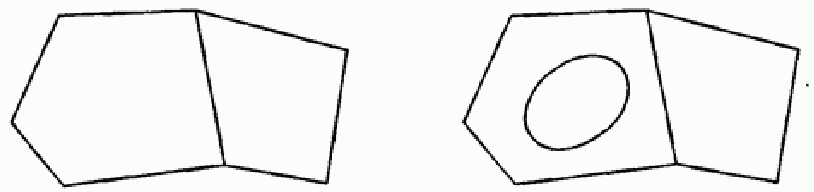


Рис.7. Добавление грани и двух циклов в оболочку

Число вырезов в грани также играет существенную роль. Топологически эквивалентными являются грани, которые путем деформирования могут быть преобразованы одна в другую. Для этого грани должны иметь одинаковое число вырезов или одинаковое число циклов. На рис. 7 показано добавление новой грани в оболочку путем введения замкнутого ребра, целиком лежащего внутри существующей грани.

Число граней, ребер и вершин при этом увеличится на единицу, а число циклов увеличится на два (цикл на добавленном замкнутом ребре должен быть посчитан дважды: один раз в одной грани, второй раз — в другой грани). Эйлерова характеристика оболочки при этом также не изменится. Эйлерова характеристика оболочки не изменится, если мы преобразуем грань с двумя циклами в грань с одним циклом, что приведено на рис.8.

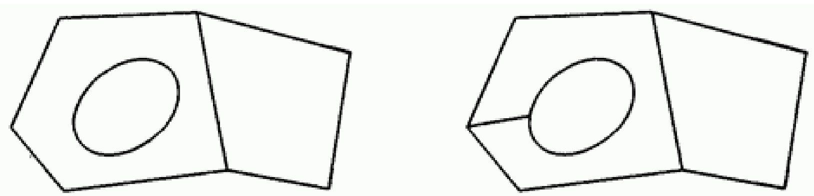


Рис.8. Добавление ребра и ликвидация внутреннего цикла грани

Покажем это. Пусть новое ребро начинается и оканчивается в уже существующих вершинах, тогда число ребер увеличится на единицу, число циклов уменьшится на единицу, а эйлерова характеристика не изменится.

Все перечисленные модификации оболочки не изменяют ее эйлеровой характеристики. Это иллюстрирует то, что эйлерова характеристика не зависит от способа разбиения оболочки на грани, а зависит только от природы оболочки.

Рассмотрим еще один пример, показывающий, что эйлерова характеристика зависит от топологии оболочки, и не зависит от способа раскроя ее на грани.

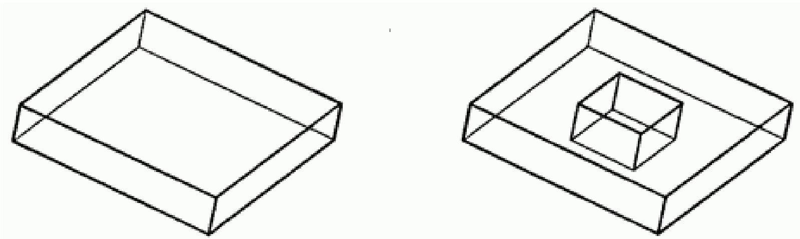


Рис.9. Эйлерова характеристика оболочек различна

Возьмем оболочку в форме четырех угольной призмы и превратим ее в оболочку в форме четырехугольной призмы с четырехугольным отверстием, что показано на рис.9.

Исходная оболочки имела следующие числа граней, циклов, ребер и вершин: F=6, L=12, V=8 и ее эйлерова характеристика равна H=2. Результирующая оболочки имеет: граней F=10, циклов L=12, ребер E=24, вершин V=16, а ее эйлерова характеристика равна H=0. В двух гранях увеличилось число циклов (было по одному циклу, а стало по два).

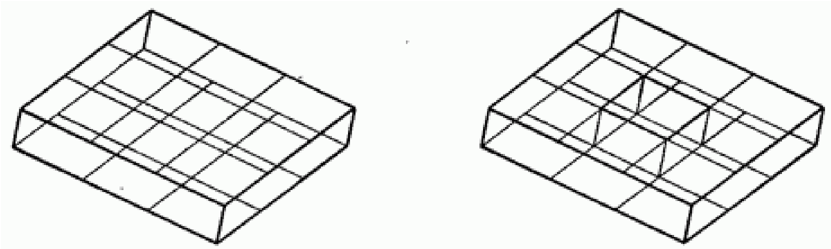


Рис. 10

Эйлерова характеристика новой оболочки уменьшилась на две единицы. Повторим переход от той же призматической оболочки к призматической оболочке с вырезом, только верхнюю и нижнюю грань исходной оболочки представим в виде совокупности девяти граней, как показано на рис.10.

Исходная оболочки имела следующие числа граней, циклов, ребер и вершин: F=22, L=22, V=32, и ее эйлерова характеристика равна H=2. Результирующая оболочки имеет: граней F=24, циклов L=24, ребер E=56, вершин V=32, а ее эйлерова характеристика равна H=0.. Все грани в исходной и результирующей оболочках имеют по одному циклу. Эйлерова характеристика результирующей оболочки, как и предыдущем примере, уменьшилась на две единицы. Это произошло в результате того, что мы получили оболочку, которая топологически неэквивалентна исходной оболочке. Действительно, представив, что обе оболочки, показанные на рис.9, выполнены из легко деформируемого материала, скруглив углы, из первой оболочки получим сферу, а из второй — тор (рис.11).

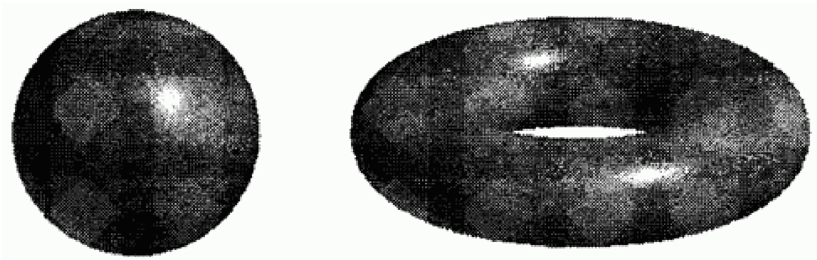


Рис.11. Сфера и тор имеют различную топологию

Никакими деформациями невозможно из сферы получить тор — они имеют разную топологию.

Эйлерова характеристика оболочки отражает ее природные свойства, связанные с возможностью деформировать одну оболочку в другую или, другими словами, установить между ними взаимно однозначное соответствие. Для описания топологических свойств оболочек и их граней используется понятие связности. Оболочки одинаковой связности могут быть деформированы одна в другую при условии, что они имеют равное число границ и одинаковую ориентируемость.

11.5. Связность оболочек

Исследуем топологические свойства оболочек. Эти свойства не зависят от количественных характеристик (объема, площади поверхности и т.п.). Не будем обращать внимания на ребра и вершины, а будем воспринимать оболочку как единое целое. Представим, что оболочка выполнена из легко деформируемого неразрываемого и не склеиваемого материала. Мы можем изменить форму оболочки, но до определенных пределов. Например, если оболочка имеет форму шара, то ей можно придать ему форму куба или тарелки, но нельзя придать ему форму тора или кувшина с ручкой. Данный факт является отражением того, что упомянутые шар, куб и тарелка имеют одинаковую топологию, которая отличается от топологии тора и кувшина с ручкой. Топологической характеристикой описанного свойства оболочек является связность. В приведенном примере оболочки с формой шара, куба или тарелки, имеют одну и ту же связность, которая отлична от связности оболочки с формой тора или кувшина с ручкой. Так как топология характеризует непрерывную связь точек объектов, то для определения топологических свойств следует изучить поведение топологических объектов при разделении их на части.

Связность — это топологическое понятие, характеризующее целостность некоторого объекта. С другой стороны связность отражает возможность разделить некоторый топологический объект на отдельные части. Простейшей является незамкнутая оболочка с одной границей (одним циклом) топологически эквивалентная кругу. Если представить, что оболочки выполнены из легко деформируемого материала, то простейшей является оболочка, которой путем деформирования можно придать плоскую форму с границей в виде окружности.

Прямоугольная, эллиптическая, любая плоская оболочка с одной границей является простейшей. Рассмотрим вопрос: какое минимальное число линий можно провести на поверхности оболочки, чтобы по этим линиям ее можно было бы разрезать на две простейшие? Простейшая оболочка делится на две отдельные части любой линией, проведенной от одной точки границы до другой. Простейшей оболочке приписывается связность равная единице. Ее называют односвязной.

Рассмотрим незамкнутые оболочки, имеющие конечное число циклов (границ), которые путем деформирования без склеивания и наложений можно сделать плоскими. На рис. 12 приведены примеры плоских оболочек с различным числом циклов.

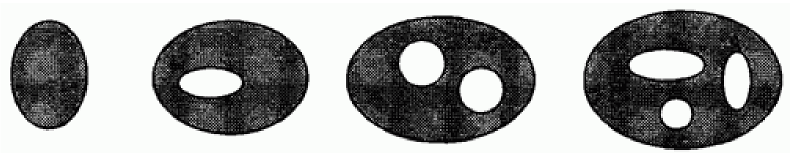


Рис.12. Плоские оболочки различной связности

Если плоская оболочка имеет L циклов (один внешний и L-1 внутренних), то на ней можно провести L-1линий, не разрезающих оболочку на две отдельные части (например, от внешнего цикла к каждому внутреннему циклу). Любая следующая линия, начинающаяся и оканчивающаяся на границе оболочки, разрежет ее на две части. Плоская оболочка с вырезами обладает дополнительными связями, которые мы режем, а оболочка не теряет свою целостность. Целостность оболочки заключается в том, что из некоторой ее точки, двигаясь по ее поверхности, можно попасть в любую другую ее точку. Плоской оболочке с L циклами припишем связность равную L.

Связность оболочки определим минимальным числом линий, по которым ее можно разрезать на две простейшие оболочки (на две отдельные части). Если связность оболочки равна h, то h-1 разрезов достаточно, чтобы ими раскроить оболочку, превратив ее в простейшую. Связность обозначим через h.

Рассмотрим замкнутые оболочки. На оболочке как на поверхности можно свободно строить кривые линии. Предположим, что мы можем резать оболочку по кривым на ней. Если оболочка имеет топологию сферы, то по любой замкнутой кривой на ее поверхности оболочку можно разрезать на две отдельные части, представляющие собой односвязные оболочки. Если оболочка имеет топологию тора, то для раскроя ее в односвязную грань, потребуется, как минимум, две замкнутые кривые на ней. Способ раскроя тороидальной оболочки показан на рис. 12.

Рис. 12. Раскрой тороидальной оболочки

Используя определение, получим, что связность сферы равна единице, а связность тора равна трем.

На замкнутой оболочке связности h можно построить h-1замкнутых кривых, которые не нарушают ее целостности (которые не разрезают ее на отдельные части), но нельзя построить h таких кривых.

Для того, чтобы замкнутую оболочку связности h раскроить в односвязную оболочку, ее нужно разрезать по h-1замкнутым линиям. Эти линии не разрезают оболочку на две отдельные части, но любая совокупность, состоящая из h замкнутых кривых на оболочке, обязательно разрезает оболочку на две части.

Рассмотренные замкнутые оболочки имеют нечетную связность. Существуют оболочки четной связности. Четную связность могут иметь незамкнутые оболочки. Например, цилиндрическая оболочка конечной длины имеет связность равную двум. Действительно, линия, проходящая от одной границы до другой, не разрезает цилиндрическую оболочку на отдельные части, а только раскраивает ее в односвязную оболочку (рис. 13).

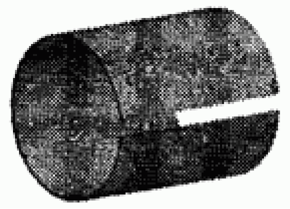


Рис. 13. Раскрой цилиндрической оболочки

Любая другая линия, проходящая от одной границы до другой, разрежет цилиндрическую оболочку на части.

На рис. 14 показаны незамкнутые оболочки, полученные из замкнутых оболочек, если в последних сделать отверстия. Если в замкнутой оболочке связности h выполнить одно отверстие с одной замкнутой границей, то связность полученной незамкнутой оболочки будет равна связности исходной замкнутой оболочки. Каждое последующее отверстие с одной замкнутой границей будет увеличивать связность полученной оболочки на единицу. Связность приведенных на рис.14 незамкнутых оболочек соответственно равна 1, 4, 3, 6.

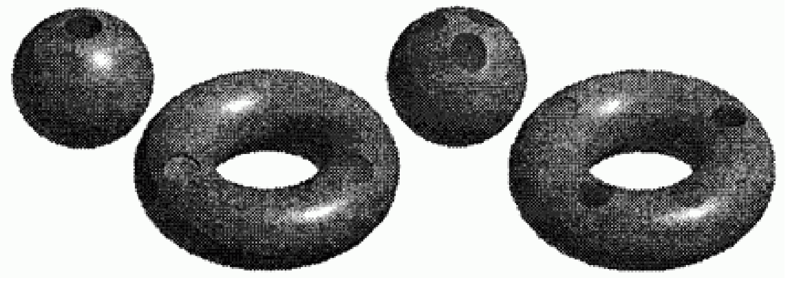


Рис.14. Незамкнутые оболочки

Система разрезов на этих оболочках строится аналогично системе разрезов на соответствующей замкнутой оболочке, только часть линий проводится от одной границы к другой границе.

Связность не является единственной характеристикой оболочки. Две оболочки могут иметь одинаковую связность, но быть топологически различными. Кроме связности оболочка характеризуются ориентируемостью.

11.6. Ориентируемость оболочек

Для многих замкнутых оболочек одну из сторон можно определить как внутреннюю, а другую — как наружную. Для точек оболочек вводится такое топологическое понятие как ориентируемость. Представим, что вокруг всякой точки оболочки проведена окружность достаточно малого радиуса, расположенная на поверхности оболочки. Для каждой окружности определим такое направление обхода, что достаточно близкие точки всегда будут обходиться в одном и том же направлении. Если для некоторой оболочки это можно сделать, то такая оболочка называется ориентируемой. Существуют оболочки, для которых нельзя ввести единое направление обхода для окружностей близких точек. Такие оболочки называются неориентируемыми.

Лист Мёбиуса.

Примером неориентируемой оболочки является лист Мёбиуса, показанный на рис.15.

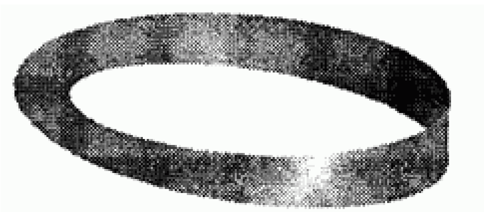


Рис.15. Лист Мёбиуса — односторонняя незамкнутая оболочка

Лист Мёбиуса можно получить, взяв бумажную полосу и склеив дальние края, повернув предварительно их друг относительно друга на 180°. До склеивания краев полосы ее стороны можно покрасить двумя разными цветами. Если покраску проводить после склеивания, то окажется, что мы окрасим одним цветом обе стороны. При движении по листу Мёбиуса мы пройдем по обеим его сторонам.

Для точек листа Мёбиуса нельзя определить ориентацию. Действительно, задав для малой окружности некоторой точки ориентацию и двигаясь по оболочке, мы попадем в исходную точку, но с противоположным направлением. Результатом этого является невозможность окрасить разные стороны оболочки в разные цвета. У оболочки всего одна сторона. Лист Мёбиуса является односторонней оболочкой. Если оболочка является односторонней, то она не ориентируема. Справедливо и утверждение, что если оболочка является двусторонней, то она ориентируема.

Оболочка тогда и только тогда неориентируема, когда на ней можно построить такую замкнутую кривую s, что при движении вдоль этой кривой достаточно малой ориентированной окружности она придет в исходную точку ориентированной в противоположном направлении.

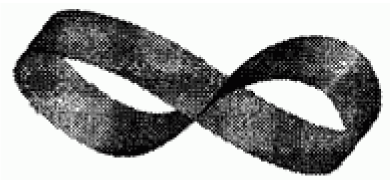


Рис.16. Ориентируемая самопересекающаяся незамкнутая оболочка

Если двигаться вдоль кривой s на односторонней оболочке по одну сторону от этой кривой, то можно оказаться по другую сторону кривой, хотя при движении кривая не пересекалась.

Лист Мёбиуса является незамкнутой оболочкой. Существуют замкнутые односторонние оболочки. Односторонняя оболочка не может разбить пространство на внутреннюю и внешнюю части, поэтому односторонняя замкнутая оболочка всегда пересекает сама себя. Однако не всякая самопересекающаяся оболочка является односторонней. Оболочка, показанная на рис.16, самопересекающаяся, но не односторонняя.

Бутылка Клейна.

Примером замкнутой односторонней оболочки является бутылка Клейна, которая показана на рис.17. Бутылка Клейна имеет одну замкнутую линию самопересечения.

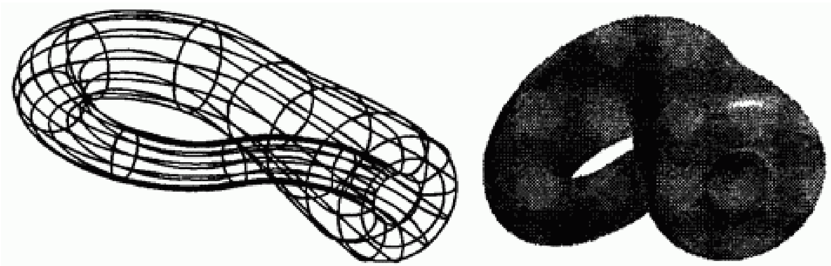


Рис.17. Бутылка Клейна — односторонняя замкнутая оболочка

Она не может служить сосудом. Связность бутылки Клейна равна трем. Система линии, разрезающих бутылку Клейна на две односвязные части, аналогична системе линий тора.

Если бутылку Клейна разрезать плоскостью ее симметрии, то получим две незамкнутые самопересекающиеся оболочки, из которых путем деформирования можно получить два листа Мёбиуса.

Гептаэдр.

Еще одной односторонней оболочкой является гептаэдр. Его можно получить из октаэдра. Для этого удалим четыре грани из восьми: на верхней части октаэдра — левую переднюю и правую заднюю, а на нижней части — правую переднюю и левую заднюю грани, и добавим три квадратные взаимно ортогональные грани, построенные на диагоналях октаэдра. Грани гептаэдра по отдельности приведены на рис. 18.

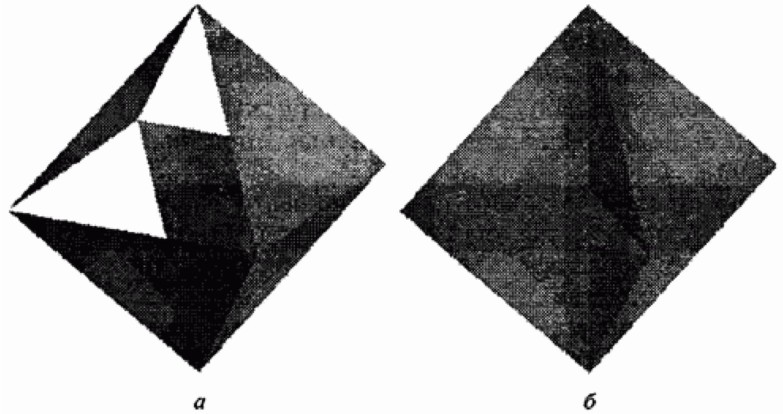


Рис.18. Четыре грани октаэдра (а), грани на диагоналях октаэдра (б)

Оболочка гептаэдра состоит из семи граней. Ребра и вершины гептаэдра совпадают с ребрами и вершинами октаэдра (диагонали не считаются ребрами, а являются линиями самопересечения оболочки). В каждом ребре гептаэдра стыкуются только две грани. Гептаэдр является замкнутой односторонней оболочкой четной связности. Он показан на рис.19.

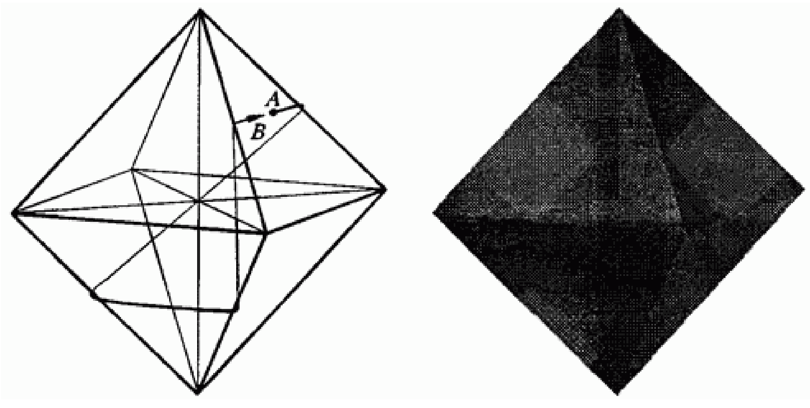


Рис.19. Гептаэдр (семигранник) — самопересекающаяся односторонняя оболочка

Если мы начнем движение по поверхности гептаэдра из точки А, то, двигаясь по указанной на рис. 19 траектории, попадем в точку В, расположенную на другой стороне поверхности.

Таким образом, кроме связности оболочки характеризуются еще и ориентируемостью. Тор и бутылка Клейна обладают одинаковой связностью, обе замкнутые оболочки, но тор является ориентируемой оболочкой, а бутылка Клейна — неориентируемой.

Известно еще несколько неориентируемых оболочек. Замкнутые неориентируемые оболочки пересекают сами себя, и поэтому не могут быть использованы для моделирования деталей.

Оболочки, с которыми мы сталкиваемся в реальности, являются ориентируемыми. Тем не менее, неориентируемые оболочки с математической точки зрения равноправны с ориентируемыми оболочками.

11.7. Оболочки для моделирования тел

Рассмотрим взаимно однозначное и непрерывное отображение одной оболочки на другую. При этом отображении соседние точки остаются соседними. Такое отображение может искажать оболочку, но при этом связанные части останутся связанными. Одним из видов отображения является деформация. При деформации топологический объект как целое непрерывно переходит сам в себя. Движение оболочки в пространстве является частным случаем деформации, тогда как зеркальное отражение оболочки относительно плоскости не является деформацией. При зеркальном отражении изменяется на обратное направление обхода всякой замкнутой кривой на оболочке, тогда как деформация сохраняет направление обхода неизменным.

Все многообразие оболочек можно классифицировать с топологической точки зрения. К одному и тому же типу относятся оболочки, которые топологически могут быть непрерывно и взаимно однозначно отображены одна на другую. Для этого должны быть выполнены следующие три условия:

1) оболочки должны иметь одинаковую связность;

2) оболочки должны быть либо ориентируемы, либо неориентируемъц

3) оболочки должны быть либо замкнуты, либо должны иметь одинаковое число границ.

Эти три необходимых условия непрерывного и взаимно однозначного отображения двух оболочек также являются и достаточными условиями.

Естественно, что оболочки, имеющие разное число границ, не могут быть отображены одна на другую.

Связность обусловливает существование системы разрезов оболочки, которая при топологическом отображении переходит в систему разрезов такой же структуры на другой оболочке. Следовательно, оболочки различной связности не могут быть отображены одна на другую.

Можно доказать, что всякая оболочка, которая может быть отображена на ориентируемую оболочку, также является ориентируемой. Это определяет третье условие принадлежности оболочек к одному типу.

Оболочки, имеющие одинаковую связность, ориентируемость и число границ, являются топологически эквивалентными.

Формула Эйлера-Пуанкаре.

Замкнутые, ориентируемые и не пересекающие сами себя оболочки имеют нечетную связность. Для таких оболочек эйлерова характеристика H связана с ее связностью h соотношением

$$H = 3 - h. \tag{1}$$

Используя это соотношение, получим формулу, связывающую число граней F, число циклов L, число ребер E и число вершин V оболочки с ее связностью h:

$$F - E + V + (F - L) = 3 - h. \tag{2}$$

Данная запись формулы Эйлера справедлива для замкнутых оболочек.

Связность не достаточно удобна для характеристики замкнутой оболочки. Введем еще одно понятие, которым можно заменить связность. Из тора путем его деформирования можно получить оболочку, по форме напоминающую гирю, которую будем называть сферой с ручкой. В общем случае любой замкнутой оболочке путем деформирования можно придать форму сферы с G ручками. Так, если взять толстую плиту, пробить в ней G отверстий и скруглить все ребра, то получим объект, оболочка которого топологически эквивалентна сфере с G ручками. На оболочке, топологически эквивалентной сфере с G ручками, можно провести 2G замкнутых кривых линий, по которым она раскраивается в простейшую оболочку. Любая следующая замкнутая линия разрежет оболочку на две простейшие оболочки. Сфера с G ручками имеет связность h=2G+1. Оболочка реальной детали топологически эквивалентна сфере с некоторым числом ручек. На рис. 20 приведена сфера с четырьмя ручками.

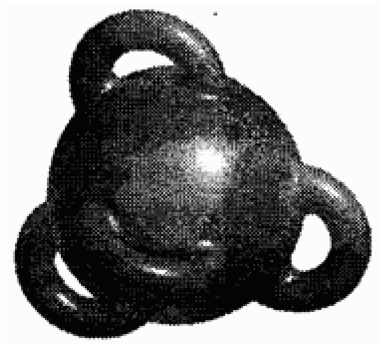


Рис.20. Сфера с четырьмя ручками

Более наглядной, чем связность, характеристикой топологии оболочки может служить число ручек G сферы, к которой путем деформирования можно привести замкнутую оболочку. Примем во внимание, что сфера с G ручками имеет связность h=2G+1 и получим формулу, связывающую число граней F, ребер E, вершин V и циклов L с характерной величиной G.

$$F - E + V + (F - L) - 2(1 - G) = 0. \quad (3)$$

Данная формула называется формулой Эйлера-Пуанкаре. Величина G (genus) характеризует топологический тип оболочки. Формула Эйлера-Пуанкаре позволяет определить топологический тип оболочки, если известно число ее граней, ребер, вершин и циклов:

$$G = 1 - F + \frac{E + L - V}{2}. \quad (4)$$

Реальные объекты могут иметь внутри пустоты. В этом случае объекты будут описываться несколькими оболочками. Одна из этих оболочек является внешней, остальные оболочки лежат внутри нее. Если моделируемый объект имеет m пустот, то он будет описываться $S=m+1$ оболочками. Потребуем, чтобы внутренние оболочки не пересекали друг друга и внешнюю оболочку. Для каждой оболочки справедлива формула (3), а для объекта с S замкнутыми оболочками формула Эйлера-Пуанкаре (3) примет вид

$$F - E + V + (F - L) - 2(S - G) = 0, \quad (5)$$

где F — общее число граней модели, E — общее число ребер модели, V — общее число вершин модели, L — общее число циклов модели, S — общее число оболочек моделируемого объекта, G — топологический тип моделируемого объекта, равный общему числу ручек всех описывающих его оболочек. Таким образом, топологически эквивалентными объектами будут являться два объекта, у которых равно число описывающих их оболочек и соответствующие внешние и внутренние оболочки имеют одинаковый топологический тип.

Если оболочка не является замкнутой, то говорить о ее топологическом типе G не имеет смысла. Для незамкнутой ориентируемой оболочки формула Эйлера имеет вид

$$F - E + V + (F - L) + h = 2. \quad (6)$$

Однородные оболочки.

Для моделирования деталей подходят замкнутые двусторонние не пересекающие сами себя оболочки. К ним относятся оболочки, топологически эквивалентные сфере с G ручками. Именно такие замкнутые оболочки мы и будем использовать для геометрического моделирования. К вершинам, ребрам, циклам и граням предъявим следующие требования. Грани не должны пересекать сами себя. Грани стыкуются только по ребрам, причем в каждом ребре стыкуются только две грани. Оболочка, приведенная на рис.21, некорректна, так как в ребре AB стыкуются четыре грани.

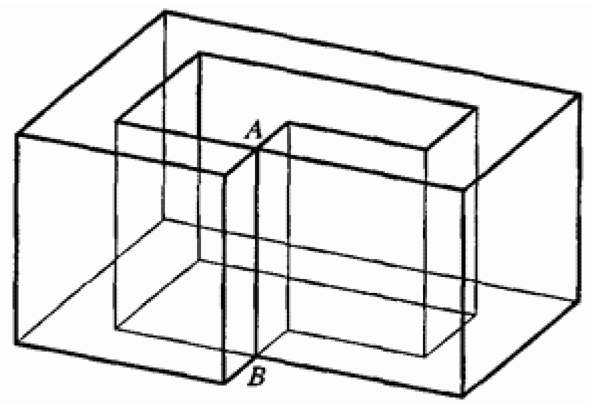


Рис.21. В каждом ребре должны пересекаться только две грани

Такая оболочка топологически должна быть представлена одним из способов, приведенных на рис. 22 и 23.

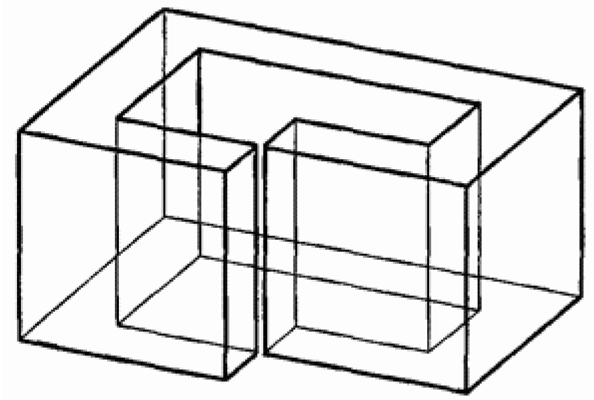


Рис.22. Корректная оболочка

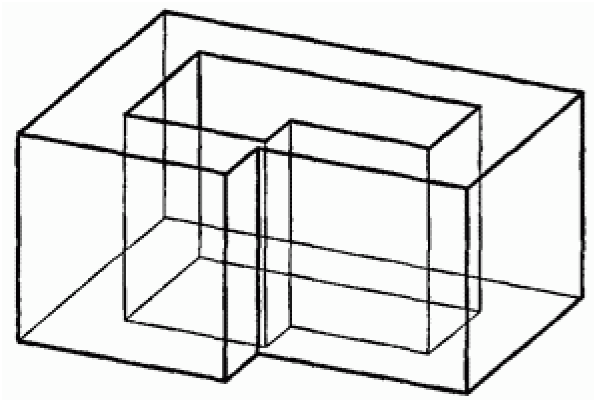


Рис.23. Корректная оболочка

Ребра стыкуются только в вершинах. В каждой вершине может стыковаться любое конечное число ребер.

Каждую вершину можно обойти по поверхности оболочки, но при этом мы должны пересечь все стыкующиеся в вершине ребра и посетить все примыкающие к вершине грани. Такие вершины будем называть простыми.

Оболочка, приведенная на рис.24, также некорректна, так как вершину А можно обойти по поверхностям граней, не пересекая всех стыкующихся в ней ребер.

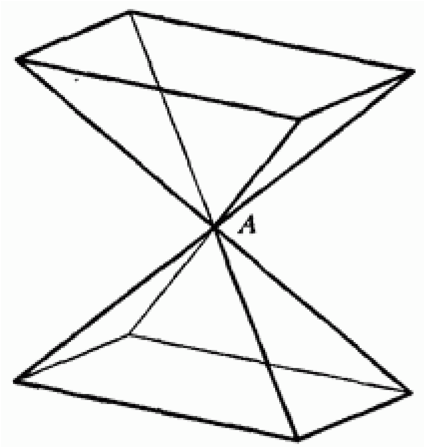


Рис.24. Некорректная оболочка

Такая оболочка топологически должна быть представлена способом, приведенным на рис.25.

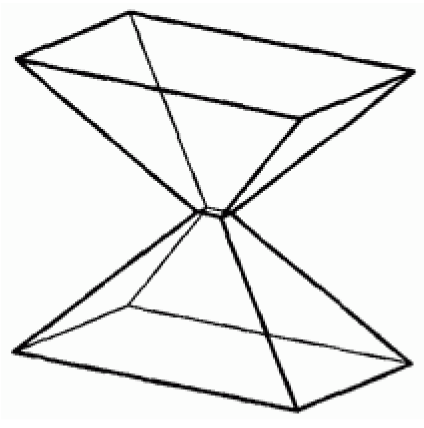


Рис.25. Корректная оболочка

Таким образом, для моделирования мы будем использовать оболочки, вершины которых являются простыми, грани пересекаются по ребрам, причем в каждом ребре стыкуются только две грани.

Такие оболочки будем называть однородными. У однородных оболочек все топологические элементы построены по единым правилам.

11.8. Поверхностное и твердотельное моделирование

В геометрическом моделировании используются термины «поверхностное моделирование» (моделирование поверхностей) и «твердотельное моделирование» (моделирование твердых тел). В обоих случаях результатом моделирования является некоторая оболочка (или несколько оболочек), описывающая поверхность моделируемого объекта. Но процесс моделирования в первом случае отличается от процесса моделирования во втором случае.

В поверхностном моделировании сначала создаются и модифицируются требуемым образом поверхности, описывающие отдельные элементы моделируемого объекта. Эти поверхности обрезают по линиям пересечения, сопрягают друг с другом поверхностями скругления или перехода, а также выполняют над ними другие операции. Затем из полученных поверхностей собирают оболочку. В поверхностном моделировании результирующая оболочка не обязательно должна быть замкнутой. Она может отражать лишь часть (главную часть) моделируемого объекта. Поверхностное моделирование позволяет сосредоточить усилия на сложных формах объекта и широко применяется для проектирования кузовов автомобилей и планеров самолетов.

В твердотельном моделировании с самого начала работа идет с оболочками тел, а не с отдельными поверхностями. Оболочки полностью описывают поверхности моделируемых объектов, отделяющие их внутренний объем от остальной части пространства. Процесс построения оболочки тела в данном случае аналогичен процессу изготовления моделируемого объекта. Сначала создается оболочка некоторой заготовки простой формы. Далее оболочка заготовки изменяется необходимым образом. Для этого используются булевы операции над телами, операция построения тонкостенного тела из заготовки, операция скругления ребер, операция построения ребер жесткости и другие операции. С помощью операций оболочке тела придается требуемая форма.

Два подхода к моделированию имеют много общего и отличаются технологией создания модели. В обоих случаях выполняются аналогичные действия, но в разной последовательности. В следующей главе мы рассмотрим моделирование твердых тел, так как оно включает в себя все элементы поверхностного моделирования.

12. МОДЕЛИРОВАНИЕ ТЕЛ

12.1. Математическая модель тел

Точки, линии и поверхности являются математическими абстракциями. У них хотя бы один размер полагается равным нулю. **Реальные объекты имеют все размеры, отличные от нуля, и занимают некоторый конечный объем в пространстве.** Для геометрического моделирования предметов, занимающих конечный объем, в математике используются объекты, называемые твердыми телами или просто телами. Способ их описания отличается от способа описания кривых и поверхностей. При моделировании тел строятся поверхности, отделяющие занимаемую ими часть пространства от остальной части пространства. Существует несколько подходов к описанию тел.

Многие предметы можно смоделировать, используя только плоские поверхности. Такое представление тел называется плоскогранным (Faceted representation или Faceted). Для описания криволинейных поверхностей плоскогранное представление может аппроксимировать их некоторым количеством пластин треугольной или четырехугольной формы. Использование плоских поверхностей значительно упрощает выполнение операций над телами. Плоскогранное представление широко применяется в строительстве и компьютерной графике для получения тоновых изображений.

Некоторые поверхности можно описать уравнениями в координатной форме (представить поверхности неявно). К ним относятся поверхности второго порядка, поверхность тора и другие. Используя для моделирования тел такие поверхности, мы придем к **конструктивной твердотельной геометрии** (Constructive Solid

Geometry или CSG). Конструктивная твердотельная геометрия оперирует **примитивами,** к которым, как правило, относят **прямоугольную призму, треугольную призму, сферу, цилиндр, конус и тор.** Над примитивами и полученными из них телами можно выполнять различные операции (в первую очередь булевы операции). Используемые конструктивной твердотельной геометрией поверхности (сферическая, цилиндрическая, коническая, поверхность тора и плоскость) делят пространство на две части и для них можно указать, с какой стороны поверхности находится внутренний объем тела. Неявное представление поверхностей дает возможность получить линии их пересечения в аналитической форме. Конструктивная твердотельная геометрия позволяет моделировать большинство промышленных деталей.

Наиболее общий подход к описанию тел состоит в представлении тела совокупностью ограничивающих его объем оболочек, грани и ребра которых заданы параметрически. Каждая оболочка строится из набора стыкующихся друг с другом поверхностей произвольной формы, содержащих полную информацию о своих границах и связях с соседями. Такое описание тел называется представлением с помощью границ (Bounded representation или B-rep).

Оно дает возможность выполнять над телами множество операций, сохраняя при этом единый способ их «внутреннего устройства». **Представление тел с помощью границ позволяет моделировать объекты произвольной формы и сложности**.

Все перечисленные подходы к описанию тел используют топологические объекты и удовлетворяют условиям связности, ориентируемости и замкнутости. Мы будем рассматривать представление тел с помощью границ, опираясь на такие топологические объекты, как вершина, ребро, грань и оболочка. Оболочки тела должны быть однородными, т. е. должны быть описаны по единым правилам. Оболочки состоят из набора граней. Каждая грань базируется на некоторой поверхности. Грань отличается от поверхности тем, что кроме поверхности она в структуре своих данных несет информацию о связях с соседними гранями и об ориентации по отношению к внутреннему объему тела. Там, где это не будет создавать путаницы, грани будем обозначать так же, как и поверхности, на которых они базируются.

Если взять любую деталь, то можно заметить, что ограничивающие ее поверхности делят пространство на две части: одну часть пространства занимает деталь и она находится внутри (вне) ограничивающих оболочек, а другая часть пространства лежит вне (внутри) детали. Перейти из одной части пространства в другую часть пространства, не пересекая ограничивающие оболочки, нельзя. Ограничивающие оболочки как бы изолируют одну часть пространства от другой. Тело может иметь одну или несколько оболочек. Если тело имеет пустоты, то его объем ограничен несколькими оболочками. Одна из этих оболочек является внешней, а остальные оболочки — внутренними. Внутренние оболочки ограничивают пустоты и целиком лежат внутри внешней оболочки. Оболочки тела не должны пересекать друг друга и сами себя. Для описания тел подходят двусторонние (ориентируемые) оболочки. Одной своей стороной каждая оболочка обращена внутрь тела, а другой — наружу. Для того чтобы отличать сторону оболочки, направленную наружу тела, от стороны, направленной внутрь тела, каждой точке оболочки приписывается **нормаль,** которая считается направленной наружу тела. Нормаль внешней оболочки тела направлена вне ограничиваемой ею части пространства, а нормаль внутренней оболочки тела направлена внутрь ограничиваемой ею части пространства. Таким образом, внутренние оболочки тела как бы вывернуты наизнанку.

С математической точки зрения одна внешняя оболочка и, возможно, несколько внутренних оболочек описывают тело. Другими словами, для создания математической модели тела достаточно смоделировать совокупность оболочек, ограничивающих его объем.

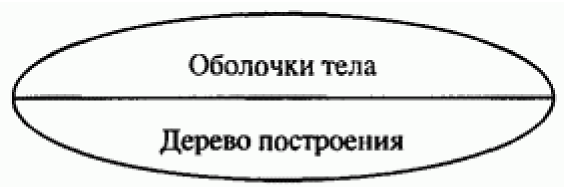


Рис. 1.1. Математическая модель тела

Но для редактирования тела необходима информация о последовательности и способах его построения, поэтому в модель тела включают еще и дерево построения (или протокол построения) тела (рис. 1.1).

В большинстве случаев оболочки, описывающие тело, являются замкнутыми, но тело можно описать и незамкнутой оболочкой, если ее края стянуты в точку. Привлечение топологии при моделировании тел необходимо для корректного выполнения операций над ними.

Топологические объекты были введены в предыдущей главе. Здесь мы остановимся на структуре данных этих объектов. Структура данных должна быть такой, чтобы при желании можно было найти всех соседей любого топологического объекта. Для этого они должны нести информацию о взаимной связи друг с другом. Будем строить оболочки тел из топологических объектов, которые будут нести и количественную геометрическую информацию, и топологическую информацию.

Оболочка тела состоит из набора граней. Грань строится на основе поверхности, которая входит в структуру данных грани. Поверхность является геометрическим носителем грани. Одна сторона грани направлена наружу оболочки, другая внутрь. Грань должна содержать признак ориентации нормали поверхности наружу или внутрь грани. Пусть признак принимает положительное значение, если нормаль поверхности направлена наружу грани, отрицательное значение — в противном случае.

Грани стыкуются между собой по ребрам, лежащим на линиях пересечения граней. Топологический объект ребро строится на основе линии пересечения поверхностей, стыкующихся в ребре граней. Пусть в ребре стыкуются грани, построенные на поверхностях $\mathbf{r}(u,v)$ и $\mathbf{s}(a,b)$. Линия пересечения граней описывается двумя поверхностями и двумя двухмерными линиями — каждая в пространстве параметров соответствующей поверхности. На грани, базирующейся на поверхности $\mathbf{r}(u,v)$ двухмерную линию обозначим векторной функцией $l_{uv}(\mathrm{t})=[\mathrm{a}(t)\mathrm{b}(\mathrm{t})]^{\mathrm{T}}$, а на грани, базирующейся на поверхности $\mathbf{s}(a,b)$ двухмерную линию обозначим векторной функцией $l_{ab}(\mathrm{t})=[\mathrm{u}(t)\mathrm{v}(\mathrm{t})]^{\mathrm{T}}$. Таким образом, линию пересечения граней, на которой базируется ребро, будем записывать в виде

$$\boldsymbol{l}_{uv}(t) = \begin{bmatrix} u(t) \\ v(t) \end{bmatrix} \equiv \begin{bmatrix} u(t) & v(t) \end{bmatrix}^{\mathsf{T}}, \qquad \boldsymbol{l}_{uv}(t) \in \mathbf{r}(u, v),$$

$$\boldsymbol{l}_{ab}(t) = \begin{bmatrix} a(t) \\ b(t) \end{bmatrix} \equiv \begin{bmatrix} a(t) & b(t) \end{bmatrix}^{\mathsf{T}}, \qquad \boldsymbol{l}_{ab}(t) \in \mathbf{s}(a, b),$$

$$t_{\min} \leqslant t \leqslant t_{\max}. \tag{1.1}$$

Ребро может иметь совпадающее со своей линией направление или ей противоположное. Для этого в структуре данных ребра должен быть соответствующий признак. Пусть этот признак принимает положительное значение, если ребро направлено по кривой, и отрицательное значение — в противном случае.

Ребра, ограничивающие грань, входят в структуру данных грани в виде циклов. Цикл — это топологический объект, характеризующий границу грани. Цикл всегда замкнут и имеет определенное направление. Он состоит из списка ребер и их ориентации в цикле. Ориентацию ребра в цикле грани будем описывать признаком, который назовем флагом. Пусть ребру, направление которого совпадает с направлением цикла, приписывается положительный флаг, а ребру, направление которого не совпадает с направлением цикла, приписывается отрицательный флаг. Таким образом, цикл состоит из списка ребер в порядке их следования и списка соответствующих им флагов.

Грань может содержать несколько циклов, причем один из циклов является внешним, а остальные циклы — внутренними. Внутренние циклы должны целиком лежать внутри внешнего цикла.

Цикл будем считать направленным так, чтобы при движении вдоль него грань всегда находилась бы слева, если смотреть с наружной стороны грани. Таким образом, внешний цикл грани ориентирован против часовой стрелки, а внутренние циклы ориентированы по часовой стрелке, если смотреть навстречу нормали грани.

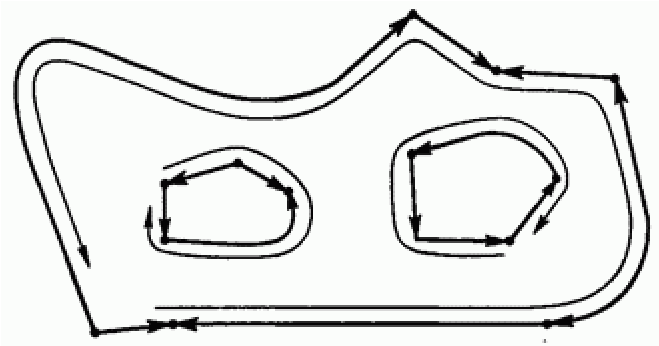


Рис. 1.2. Грань и ее структура данных: поверхность, ее ориентация и список циклов

По каждому циклу можно построить контур на поверхности, который описывает границу грани. На рис. 1.2 приведен пример плоской грани с указанием направлений ее циклов.

Ребро, разделяющее две грани, входит в два цикла: в одном цикле направление ребра совпадает с направлением цикла, а в другом — противоположно направлению цикла. Ребро входит в цикл грани слева от него с положительным флагом, а в цикл грани справа от него с отрицательным флагом.

Ребро начинается и оканчивается в вершинах. Каждая вершина базируется на точке в пространстве и содержит информацию о ребрах, стыкующихся в ней. В каждой вершине могут стыковаться несколько ребер. Если ребро замкнутое, то оно начинается и оканчивается в одной и той же вершине.

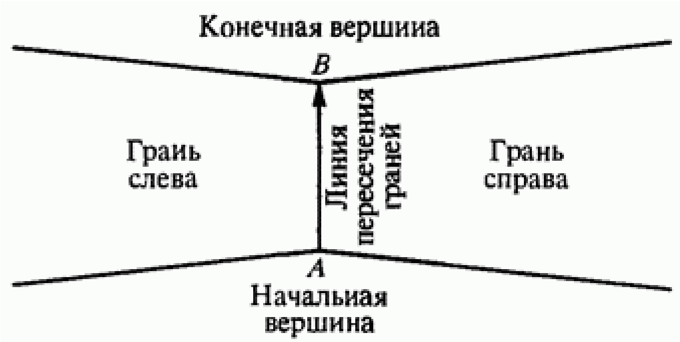


Рис. 1.3. Ребро АВ и его структура данных

Ребро должно содержать информацию о том, какая грань находится слева, а какая — справа от ребра, если смотреть с внешней стороны оболочки вдоль направления ребра, и какая вершина находится в начале ребра и какая — в конце. На рис. 1.3 показано ребро, начинающееся в вершине А и оканчивающееся в вершине В.

Описанная структура данных топологических объектов содержит двунаправленную связь между ними. Оказавшись на одном из топологических объектов, можно последовательно пройти всю оболочку, в которую входит этот объект.

Таким образом, математическая модель тела содержит количественную геометрическую информацию в виде поверхностей, линий и точек их стыковки, топологическую информацию в виде связей точек поверхности тела между собой и информацию о последовательности и способах построения в виде дерева построения.

12.2. Простейшие тела

Способы моделирования деталей часто повторяют технологический процесс их производства. Один из способов моделирования тел заключается в том, что берется некоторая заготовка тела и затем путем удаления и добавления в определенных местах дополнительного объема (материала) получают тело требуемой формы. В качестве заготовок могут браться тела простейшей формы: прямоугольная призма, цилиндр, конус, шар, тор и другие. Простейшие тела состоят из одной оболочки, построенной по общим правилам. Рассмотрим построение оболочки некоторых простейших тел.

Для каждого простейшего тела нам потребуется местная декартова прямоугольная система координат, радиус-вектор начала которой обозначим через **p**, а базисные орты обозначим через i_x i_y i_z

Прямоугольная призма.

Начало местной системы координат поместим в одну из вершин призмы, а ее орты направим по ребрам, стыкующимся в этой вершине. Пусть в направлении орта i_x тело имеет длину, равную x в направлении орта i_y — длину, равную y, а в направлении орта i_z — длину, равную z. Прямоугольная призма состоит из шести граней. Каждая грань представляет собой часть плоскости, ограниченную прямоугольным контуром на ней, с признаком ориентации нормали плоскости наружу тела и одним циклом. Контуры состоят из отрезков прямых (2.2.3). Прямоугольная призма имеет 12 ребер. Каждое ребро состоит из линии пересечения поверхностей соседних граней и признака совпадения направления ребра с направлением линии пересечения. Каждая линия пересечения состоит из двух линий на поверхности: одна на поверхности одной грани, другая на поверхности второй грани. Обе линии на поверхности имеют одинаковую геометрическую и параметрическую длину и полностью совпадают в пространстве. Каждая линия на поверхности представляет собой совокупность поверхности и двухмерной линии на ней. Прямоугольная призма с ориентацией циклов граней показана на рис. 2.1.

Грани тела будут описываться поверхностями:

$$
\begin{array}{lll}
\mathbf{r}_1(u_1,\, v_1) = \mathbf{p} + u_1\mathbf{i}_x + v_1\mathbf{i}_y, & 0 \leqslant u_1 \leqslant x, & 0 \leqslant v_1 \leqslant y, \\
\mathbf{r}_2(u_2,\, v_2) = \mathbf{p} + u_2\mathbf{i}_y + v_2\mathbf{i}_z, & 0 \leqslant u_2 \leqslant y, & 0 \leqslant v_2 \leqslant z, \\
\mathbf{r}_3(u_3,\, v_3) = \mathbf{p} + u_3\mathbf{i}_x + v_3\mathbf{i}_z, & 0 \leqslant u_3 \leqslant x, & 0 \leqslant v_3 \leqslant z, \\
\mathbf{r}_4(u_4,\, v_4) = \mathbf{p} + z\mathbf{i}_z + u_4\mathbf{i}_x + v_4\mathbf{i}_y, & 0 \leqslant u_4 \leqslant x, & 0 \leqslant v_4 \leqslant y, \\
\mathbf{r}_5(u_5,\, v_5) = \mathbf{p} + x\mathbf{i}_x + u_5\mathbf{i}_y + v_5\mathbf{i}_z, & 0 \leqslant u_5 \leqslant y, & 0 \leqslant v_5 \leqslant z, \\
\mathbf{r}_6(u_6,\, v_6) = \mathbf{p} + y\mathbf{i}_y + u_6\mathbf{i}_x + v_6\mathbf{i}_z, & 0 \leqslant u_6 \leqslant x, & 0 \leqslant v_6 \leqslant z.
\end{array} \quad (2.1)
$$

У третьей, четвертой и пятой граней нормаль поверхности направлена наружу тела, а у первой, второй и шестой — внутрь тела. Эта информация содержится в грани в виде признака совпадения нормалей.

Приведем описание одного из ребер.

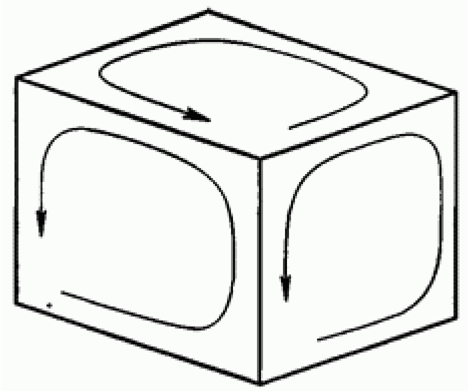


Рис. 2.1. Ориентация циклов граней призматического тела

Например, ребро между первой и второй гранями описывается линией пересечения этих граней, состоящей из двух двухмерных отрезков:

$$\boldsymbol{l}_{uv1}(t) = \left[u_1(t) \quad v_1(t)\right]^\top = [0 \quad yt]^\top, \qquad \boldsymbol{l}_{uv1}(t) \in \mathbf{r}_1(u_1,\, v_1),$$
$$\boldsymbol{l}_{uv2}(t) = \left[u_2(t) \quad v_2(t)\right]^\top = [yt \quad 0]^\top, \qquad \boldsymbol{l}_{uv2}(t) \in \mathbf{r}_2(u_2,\, v_2),$$
$$0 \leqslant t \leqslant 1, \tag{2.2}$$

и двух поверхностей. Выражение l_{uv1} (t) $\in$ $\mathbf{r_1}(u_1,v_1)$ означает, что отрезок l_{uv1} (t) лежит на поверхности $\mathbf{r_1}(u_1,v_1)$, а выражение l_{uv2} (t) $\in$ $\mathbf{r_2}(u_2,v_2)$ означает, что отрезок l_{uv2} (t) лежит на поверхности $\mathbf{r_2}(u_2,v_2)$. Пусть направление ребра совпадает с направлением кривой (2.2), что зафиксируем в признаке совпадения направлений. Признак совпадения направления ребра с направлением цикла грани мы назвали **флагом**. Флаг может принимать два значения: положительное и отрицательное. Ребро входит в цикл первой грани с положительным флагом, а в цикл второй грани — с отрицательным флагом. Если смотреть вдоль направления ребра снаружи тела, то слева от ребра лежит первая грань, а справа от ребра лежит вторая грань.

Для построения граней тела достаточно знать местную систему координат **p**, $\mathbf{i}_x$, $\mathbf{i}_y$,$\mathbf{i}_z$ и стороны *x, y, z* призмы.

Цилиндрическое тело.

Другой заготовкой может служить цилиндрическое тело (рис. 2.2). Начало местной системы координат **p** поместим в центр одного из торцев цилиндра, а орт $\mathbf{i}_z$ направим вдоль его оси. Пусть в цилиндр имеет радиус *r* и длину h.

Цилиндрическое тело имеет три грани. Торцевые грани состоят из частей плоскости, ограниченных окружностями на них, граничного цикла и признака ориентации нормали плоскости грани наружу тела. Цикл каждой торцевой грани состоит из одного замкнутого ребра. Геометрическим носителем такого ребра является линия пересечения, состоящая из двух кривых на поверхности. Одна кривая является окружностью на плоскости, а вторая — линией $v=v_{\min}$ или $v=v_{\max}$ на цилиндрической поверхности. Напомним, что обе линии на поверхности, составляющие линию пересечения, должны иметь одинаковую параметрическую длину.

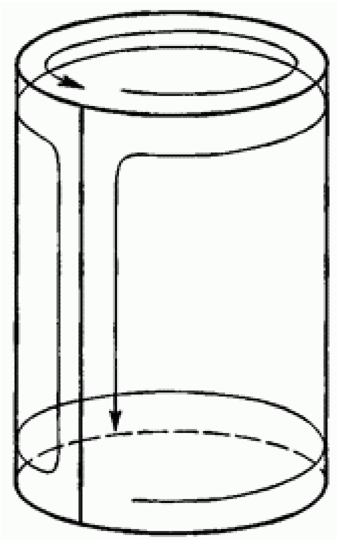


Рис. 2.2. Ориентация циклов граней цилиндрического тела

Боковая грань тела базируется на цилиндрической поверхности

$$\mathbf{r}_1(u_1,\, v_1) = \mathbf{p} + r\cos u_1 \mathbf{i}_x + r \sin u_1 \mathbf{i}_y + h v_1\, \mathbf{i}_z,$$
$$0 \leqslant u_1 \leqslant 2\pi, \quad 0 \leqslant v_1 \leqslant 1. \tag{2.3}$$

Нормали поверхности (2.3) и ее грани совпадают по направлению. Эта грань имеет один цикл. Цилиндрическая поверхность боковой грани является замкнутой по одному из параметров. Грани основания базируются на ограниченных окружностями плоскостях

$$\mathbf{r}_2(u_2,\, v_2) = \mathbf{p} + u_2 \mathbf{i}_x + v_2 \mathbf{i}_y,$$
$$\mathbf{r}_3(u_3,\, v_3) = \mathbf{p} + u_3 \mathbf{i}_x + v_3 \mathbf{i}_y + h \mathbf{i}_z, \tag{2.4-2.5}$$

где параметры u_2, v_2 лежат внутри окружности

l_{uv2}(t) $=[r \cos t \; r \sin t]^{T}$, $0 \le t \le 2\pi$ на плоскости $\mathbf{r_2}(u_2, v_2)$, а параметры u_3, v_3 лежат внутри окружности l_{uv3}(t) $=[r \cos t \; r \sin t]^{T}$, $0 \le t \le 2\pi$, на плоскости $\mathbf{r_3}(u_3, v_3)$.

Нормали поверхности (2.4) и ее грани противоположны по направлению, а нормали поверхности (2.5) и ее грани совпадают.

Ребро, построенное на линии замыкания оболочки, является швом. Шов, так же как и любое другое ребро, базируется на линии пересечения. В данном случае линия пересечения описывается двухмерными кривыми

$$\boldsymbol{l}_{uv01}(t) = \left[u_1(t) \;\; v_1(t)\right]^{\top} = [0 \;\; ht]^{\top}, \qquad \boldsymbol{l}_{uv01}(t) \in \mathbf{r}_1(u_1,\, v_1),$$
$$\boldsymbol{l}_{uv02}(t) = \left[u_1(t) \;\; v_1(t)\right]^{\top} = [2\pi \;\; ht]^{\top}, \qquad \boldsymbol{l}_{uv02}(t) \in \mathbf{r}_1(u_1,\, v_1),$$
$$0 \leqslant t \leqslant 1, \tag{2.6}$$

являющимися линиями $u=u_{\min}$ и $u=u_{\max}$ на цилиндрической поверхности.

Ребро между боковой гранью и основанием (2.4) описывается линией пересечения этих граней, состоящей из двух двухмерных кривых (отрезка и окружности)

$$\begin{aligned} \boldsymbol{l}_{uv1}(t) &= \left[u_1(t) \quad v_1(t)\right]^{\top} = [t \quad 0]^{\top}, & \boldsymbol{l}_{uv1}(t) &\in \mathbf{r}_1(u, v_1), \\ \boldsymbol{l}_{uv2}(t) &= \left[u_2(t) \quad v_2(t)\right]^{\top} = [r\cos t \quad r\sin t]^{\top}, & \boldsymbol{l}_{uv2}(t) &\in \mathbf{r}_2(u_2,\, v_2), \\ & 0 \leqslant t \leqslant 2\pi. \end{aligned} \tag{2.7}$$

Отрезок l_{uv1} (t) лежит на поверхности $\mathbf{r_1}(u_1,v_1)$ окружность l_{uv2} (t) лежит на поверхности $\mathbf{r_2}(u_2,v_2)$. Пусть направление ребра совпадает с направлением кривой (2.7), что отметим соответствующим признаком совпадения направлений. Это ребро входит в цикл первой грани с положительным флагом, а в цикл второй грани — с отрицательным флагом. Если смотреть вдоль направления ребра снаружи тела, то слева от. ребра лежит первая грань, а справа от ребра лежит вторая грань.

Ребро между боковой гранью и основанием (2.5) описывается линией пересечения этих граней, состоящей из двух двухмерных кривых (отрезка и окружности)

$$\begin{aligned} \boldsymbol{l}_{uv1}(t) &= \left[u_1(t) \quad v_1(t)\right]^{\top} = [t \quad h]^{\top}, & \boldsymbol{l}_{uv1}(t) &\in \mathbf{r}_1(u_1,\, v_1), \\ \boldsymbol{l}_{uv3}(t) &= \left[u_3(t) \quad v_3(t)\right]^{\top} = [r\cos t \quad r\sin t]^{\top}, & \boldsymbol{l}_{uv3}(t) &\in \mathbf{r}_3(u_3,\, v_3), \\ & 0 \leqslant t \leqslant 2\pi. \end{aligned} \tag{2.8}$$

Отрезок l_{uv1} (t) лежит на поверхности $\mathbf{r_1}(u_1,v_1)$ окружность l_{uv3} (t) лежит на поверхности $\mathbf{r_3}(u_3,v_3)$ Пусть направление ребра совпадает с направлением кривой (2.8), что отметим соответствующим признаком совпадения направлений. Это ребро входит в цикл первой грани с отрицательным флагом, а в цикл третьей грани — с положительным флагом. Если смотреть вдоль направления ребра снаружи тела, то слева от ребра лежит третья грань, а справа — первая грань.

Циклы граней основания содержат всего одно ребро. Цикл боковой грани состоит из списка ребер с соответствующими флагами:

ребро на базе кривой (2.7) — флаг положительный,

ребро на базе кривой (2.6) — флаг положительный,

ребро на базе кривой (2.8) — флаг отрицательный,

ребро на базе кривой (2.6) — флаг отрицательный.

Цилиндрическое тело и ориентация циклов его граней показаны на рис. 2.2.

Для построения граней тела достаточно знать местную систему координат **p**, $\mathbf{i}_x$, $\mathbf{i}_y$, $\mathbf{i}_z$ радиус r и высоту h цилиндра.

Коническое тело.

Тело в форме усеченного конуса строится аналогично цилиндрическому телу с той лишь разницей, что в качестве поверхности боковой грани вместо цилиндрической используется коническая поверхность

$$\mathbf{r}_1(u_1,\, v_1) = \mathbf{p} + (r + hv_1 \operatorname{tg}\gamma)(\cos u_1 \mathbf{i}_x + \sin u_1 \mathbf{i}_y) + hv_1 \mathbf{i}_z, \quad 0 \leqslant u_1 \leqslant 2\pi, \quad 0 \leqslant v_1 \leqslant 1. \tag{2.9}$$

Грани основания базируются на ограниченных окружностями поверхностях

$$\mathbf{r}_2(u_2,\, v_2) = \mathbf{p} + u_2 \mathbf{i}_x + v_2 \mathbf{i}_y,$$
$$\mathbf{r}_3(u_3,\, v_3) = \mathbf{p} + u_3 \mathbf{i}_x + v_3 \mathbf{i}_y + h \mathbf{i}_z, \quad \text{(2.10-2.11)}$$

где параметры u_2,v_2 лежат внутри окружности

$$l_{uv2}(t) = [r\cos t \quad r\sin t]^{\top}, \qquad 0 \leqslant t \leqslant 2\pi,$$

а параметры u_3, v_3 лежат внутри окружности

$$l_{uv3}(t) = [(r + h\operatorname{tg}\gamma)\cos t \quad (r + h\operatorname{tg}\gamma)\sin t]^{\top}, \qquad 0 \leqslant t \leqslant 2\pi.$$

Ребро, построенное на боковой грани, будет являться швом. В общем случае коническое тело имеет три грани. Если конус не усеченный, то одна из торцевых граней стянута в точку. Стянутую в точку грань можно исключить из модели тела, тогда оболочка тела с топологической точки зрения будет незамкнутой, хотя диаметр отверстия в ней равен нулю.

Для построения граней тела достаточно знать местную систему координат $\mathbf{p}$, $\mathbf{i}_x$, $\mathbf{i}_y$,$\mathbf{i}_z$ радиус r одного из оснований конуса, высоту h и угол конусности γ.

Сферическое тело.

Сферическая поверхность может служить оболочкой сферического тела, но эта оболочка не является замкнутой, так как имеет два отверстия нулевого радиуса в полюсах. Начало местной системы координат $\mathbf{p}$ для сферического тела поместим в центр сферы. Оболочка, описываемая сферической поверхностью

$$\mathbf{r}(u,\ v) = \mathbf{p} + r\cos v\cos u\,\mathbf{i}_x + r\cos v\sin u\,\mathbf{i}_y + r\sin v\,\mathbf{i}_z,$$
$$0 \leqslant u \leqslant 2\pi, \quad -\frac{\pi}{2} \leqslant v \leqslant \frac{\pi}{2},$$

топологически эквивалентна цилиндрической поверхности.

Ребра в полюсах стянуты в точку, но двухмерные кривые на сферической поверхности в полюсах имеют ненулевую длину. Цикл грани сферического тела составляют три ребра, одно из которых является швом

$$
\begin{aligned}
&\boldsymbol{l}_{uv1}(t) = \left[u(t) \quad v(t)\right]^{\top} = [0 \quad t]^{\top}, \qquad \boldsymbol{l}_{uv1}(t) \in \mathbf{r}(u,\, v),\\
&\boldsymbol{l}_{uv2}(t) = \left[u(t) \quad v(t)\right]^{\top} = [2\pi \quad t]^{\top}, \qquad \boldsymbol{l}_{uv2}(t) \in \mathbf{r}(u,\, v),\\
&-\frac{\pi}{2} \leqslant t \leqslant \frac{\pi}{2},
\end{aligned}
$$

а два других описываются линиями на полюсах сферы:

$$
\begin{aligned}
&\boldsymbol{l}_{uv01}(t) = [u(t) \quad v(t)]^{\top} = \left[t \quad -\frac{\pi}{2}\right]^{\top}, \qquad \boldsymbol{l}_{uv01}(t) \in \mathbf{r}(u,\, v),\\
&\boldsymbol{l}_{uv02}(t) = [u(t) \quad v(t)]^{\top} = \left[t \quad \frac{\pi}{2}\right]^{\top}, \qquad \boldsymbol{l}_{uv02}(t) \in \mathbf{r}(u,\, v),\\
&0 \leqslant t \leqslant 2\pi.
\end{aligned}
$$

Линии пересечения в полюсах состоят из двух одинаковых линий на сфере. Шов входит в список ребер цикла дважды, один раз с положительным флагом, второй раз — с отрицательным флагом. Сферическое тело показано на рис. 2.3.

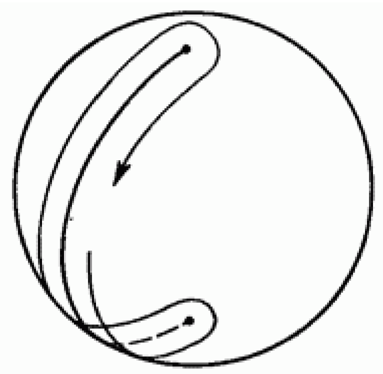


Рис. 2.3. Ориентация цикла грани сферического тела

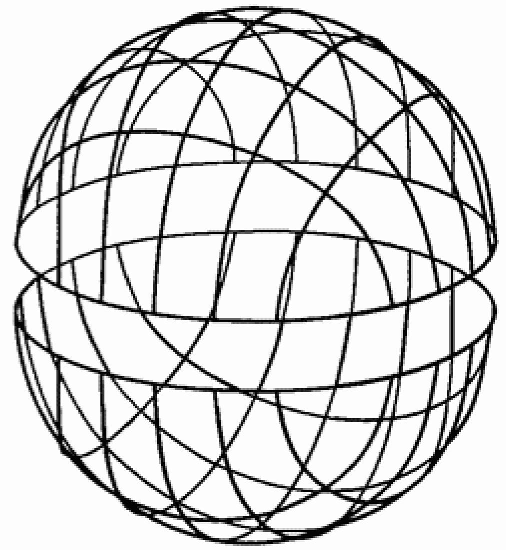


Рис. 2.4. Построение сферического тела по двум полусферам

Сферическое тело можно построить из двух полусфер, описываемых векторными функциями

$$\mathbf{r}_1(u,\ v) = \mathbf{p} + x\mathbf{i}_x + y\mathbf{i}_y + \sqrt{r^2 - x^2 - y^2}\,\mathbf{i}_z,$$

$$\mathbf{r}_2(u,\ v) = \mathbf{p} + x\mathbf{i}_x + y\mathbf{i}_y - \sqrt{r^2 - x^2 - y^2}\,\mathbf{i}_z,$$

(2.12-2.13)

где

$$x = x(u,\ v) = r\left(\cos u + \cos v - 1 + \frac{2u(1-\cos v - \sin v)}{\pi} + \frac{2v(1-\cos u - \sin u)}{\pi}\right),$$

$$y = y(u,\ v) = r\left(\sin u - \sin v - \frac{2u(1-\cos v - \sin v)}{\pi} + \frac{2v(1-\cos u - \sin u)}{\pi}\right),$$

$$0 \leqslant u \leqslant \frac{\pi}{2}, \quad 0 \leqslant v \leqslant \frac{\pi}{2},$$

r - радиус сферы, $\mathbf{p}$— радиус-вектор центра, $\mathbf{i}_x$, $\mathbf{i}_y$,$\mathbf{i}_z$ — орты, определяющие ориентацию сферы. Поверхность (2.12) описывает верхнюю полусферу, а поверхность (6.2.13) описывает нижнюю полусферу. Нормали верхней поверхности и ее грани совпадают по направлению, а нормали нижней поверхности и ее грани противоположны по направлению. Оболочка сферического тела, состоящая из двух полусфер, является замкнутой. Ее полусферы стыкуются по четырем ребрам:

$$l_{uv11}(t) = [u(t) \;\; v(t)]^{\top} = \left[\frac{\pi t}{2} \;\; 0\right]^{\top}, \qquad l_{uv11}(t) \in \mathbf{r}_1(u, v),$$

$$l_{uv21}(t) = [u(t) \;\; v(t)]^{\top} = \left[\frac{\pi t}{2} \;\; 0\right]^{\top}, \qquad l_{uv21}(t) \in \mathbf{r}_2(u, v),$$

$$0 \leqslant t \leqslant 1;$$

$$l_{uv12}(t) = [u(t) \;\; v(t)]^{\top} = \left[\frac{\pi}{2} \;\; \frac{\pi t}{2}\right]^{\top}, \qquad l_{uv12}(t) \in \mathbf{r}_1(u, v),$$

$$l_{uv22}(t) = [u(t) \;\; v(t)]^{\top} = \left[\frac{\pi}{2} \;\; \frac{\pi t}{2}\right]^{\top}, \qquad l_{uv22}(t) \in \mathbf{r}_2(u, v),$$

$$0 \leqslant t \leqslant 1;$$

$$l_{uv13}(t) = [u(t) \;\; v(t)]^{\top} = \left[\frac{\pi(1-t)}{2} \;\; \frac{\pi}{2}\right]^{\top}, \qquad l_{uv13}(t) \in \mathbf{r}_1(u, v),$$

$$l_{uv23}(t) = [u(t) \;\; v(t)]^{\top} = \left[\frac{\pi(1-t)}{2} \;\; \frac{\pi}{2}\right]^{\top}, \qquad l_{uv23}(t) \in \mathbf{r}_2(u, v),$$

$$0 \leqslant t \leqslant 1;$$

$$l_{uv14}(t) = [u(t) \;\; v(t)]^{\top} = \left[0 \;\; \frac{\pi(1-t)}{2}\right]^{\top}, \qquad l_{uv14}(t) \in \mathbf{r}_1(u, v),$$

$$l_{uv24}(t) = [u(t) \;\; v(t)]^{\top} = \left[0 \;\; \frac{\pi(1-t)}{2}\right]^{\top}, \qquad l_{uv24}(t) \in \mathbf{r}_2(u, v),$$

$$0 \leqslant t \leqslant 1.$$

Поверхности (2.12) и (2.13) для построения сферического тела приведены на рис. 2.4.

Для построения граней тела достаточно знать местную систему координат $\mathbf{p}$, $\mathbf{i}_x$, $\mathbf{i}_y$,$\mathbf{i}_z$ и радиус r сферы.

Тороидальное тело.

Начало местной системы координат **p** для тороидального тела поместим в его центр, а орт $\mathbf{i}_z$ направим по оси симметрии тела. Пусть больший радиус тора равен R, а малый радиус тора равен *r*. Тороидальное тело имеет одну грань, описываемую тороидальной поверхностью

$$\mathbf{r}(u,\, v) = \mathbf{p} + (R + r\cos v)\cos u\,\mathbf{i}_x + (R + r\cos v)\sin u\,\mathbf{i}_y + r\sin v\,\mathbf{i}_z$$
$$0 \leqslant u \leqslant 2\pi, \quad -\pi \leqslant v \leqslant \pi, \tag{2.14}$$

два ребра

$$\begin{gathered} \boldsymbol{l}_{uvR1}(t) = \left[u(t) \quad v(t)\right]^\top = [t \quad -\pi]^\top \\ \boldsymbol{l}_{uvR2}(t) = \left[u(t) \quad v(t)\right]^\top = [t \quad \pi]^\top, \\ 0 \leqslant t \leqslant 2\pi, \\ \boldsymbol{l}_{uvr1}(t) = \left[u(t) \quad v(t)\right]^\top = [0 \quad t]^\top, \\ \boldsymbol{l}_{uvr2}(t) = \left[u(t) \quad v(t)\right]^\top = [2\pi \quad t]^\top, \\ -\pi \leqslant t \leqslant \pi, \end{gathered} \tag{2.15-2.16}$$

и одну вершину в точке пересечения ребер $\mathbf{p} + (R\text{-}r)\,\mathbf{i}_x$. Нормаль поверхности и грани тороидального тела совпадают по направлению. Грань тела имеет один цикл. Цикл грани состоит из списка ребер с соответствующими флагами:

ребро на базе кривой (2.15) — флаг положительный,

ребро на базе кривой (2.16) — флаг положительный,

ребро на базе кривой (2.15) — флаг отрицательный,

ребро на базе кривой (2.16) — флаг отрицательный.

Тороидальная поверхность грани является замкнутой по обоим параметрам, поэтому оба ребра оболочки тела замкнуты и являются швами. Тороидальное тело, его ребра и цикл грани показаны на рис. 2.5.

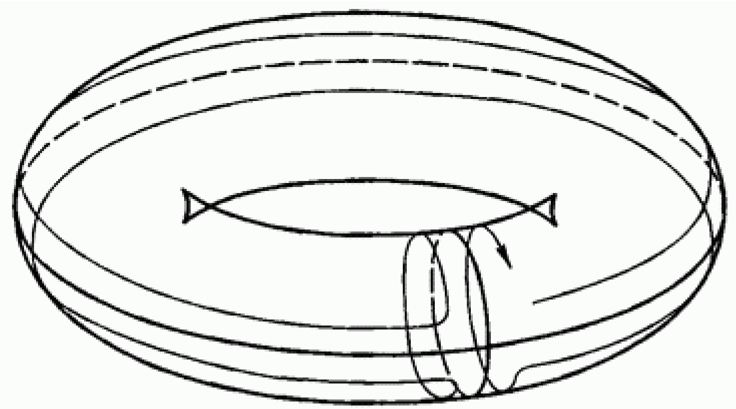


Рис. 2.5. Ориентация цикла грани тороидального тела

Для построения граней тела достаточно знать местную систему координат $\mathbf{p}$, $\mathbf{i}_x$, $\mathbf{i}_y$,$\mathbf{i}_z$ больший радиус R и меньший радиус r.

Принцип построения тел.

Рассмотренные тела иллюстрируют принцип построения математической модели тел. Эти тела называют твердотельными примитивами. К ним могут быть отнесены еще некоторые тела простой формы, например, треугольная призма (клин). Все тела построены по такому же принципу, что и твердотельные примитивы.

Может возникнуть вопрос: для чего требуется так усложнять модель тела, в частности, для чего требуется строить ребра на кривых пересечения поверхностей? Действительно, тела можно было бы описать набором поверхностей, не используя ни грани, ни ребра, ни вершины. У реальных деталей эти поверхности могут иметь очень сложную форму, как в смысле кривизны, так и в смысле границ, и эти поверхности каким-то образом нужно построить. Одним из удобных способов построения поверхностей, описывающих тело, является способ одновременного построения всех требуемых поверхностей с помощью операций над телами. Для этого берется одно из простых тел, и далее в определенных местах к нему добавляется объем или от него отнимается объем. Например, для того чтобы просверлить отверстие в некотором теле, выполняется булева операция вычитания из этого тела цилиндрического тела, играющего роль сверла. Аналогично выполняются пазы и вырезы. Для того чтобы сварить модели двух деталей, выполняется булева операция объединения тел. Использование топологических объектов необходимо для корректного выполнения этих операций. Пусть требуется отрезать от одного из описанных выше простейших тел некоторую часть и пусть резка производится плоскостью. Тогда мы вынуждены найти линии пересечения поверхностей тела с этой плоскостью и по этим линиям обрезать поверхности тела и саму плоскость. Кроме того, нужно найти место стыковки обрезанной плоскости с частью исходного тела. Для плоскости нужно определить, какая ее сторона будет смотреть наружу тела, а какая — внутрь. Все это приводит к тому, что нужно знать топологию исходного тела и строить его по общим правилам.

При проектировании приходится рассматривать несколько вариантов деталей и сборочных единиц. Различные варианты одной и той же детали можно получить путем изменения требуемых параметров ее исходного варианта. Для этого в математической модели детали необходимо иметь информацию о пути и способах ее построения. Таким образом, геометрическая модель детали или сборочной единицы должна быть дополнена еще некоторой информацией о последовательности ее построения.

12.3. Построение тела по плоским сечениям

Многие детали можно построить по их плоским сечениям в определенных местах. Пусть имеется несколько плоских контуров $c_i(t)$, i=0,1,2,…,n, одинаково ориентированных, расположенных на некотором расстоянии друг от друга и состоящих из одинакового числа кривых. Построим тело, сечения которого ограничены данными контурами. Тело имеет две торцевые грани, построенные на крайних контурах. Торцевые грани тела представляют собой плоскости, ограниченные одним из крайних контуров. Боковыми гранями оболочки тела могут служить сглаживающие поверхности или поверхности . Их количество равно количеству кривых в контуре. В каждой боковой грани участвует по одной кривой каждого контура.

Эти кривые будем называть соответствующими. Между кривыми в контурах должно быть установлено соответствие, например, по порядковому номеру кривой в контуре. Тело, построенное по плоским сечениям, приведено на рис. 3.1.

Рис. 3.1. Тело, построенное по плоским сечениям

В общем случае все контуры, за исключением торцевых контуров, могут быть не плоскими. Тело, построенное по плоским сечениям, может быть замкнутым и иметь топологию тора.

Тело, построенное по сечениям, описываемым незамкнутыми линиями, приведено на рис. 3.2.

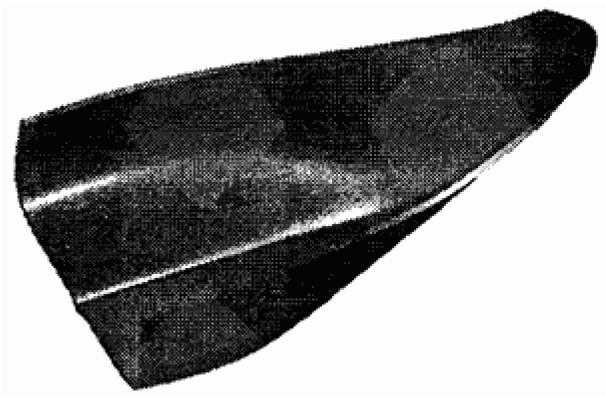


Рис.3.2. Тело, построенное по плоским сечениям

По незамкнутым линиям в любом случае строятся замкнутые контуры, в которых участвуют эквидистантные кривые.

12.4. Тело в форме листа

На основе поверхности произвольной формы можно построить тело в форме листа конечной толщины. Пусть дана поверхность **b**(u,v). Выбрав толщину листа h, построим эквидистантную к ней поверхность

$$\mathbf{r}(u, v) = \mathbf{b}(u, v) + h\mathbf{m}(u, v), \quad (4.1)$$

где **m**(u,v)— нормаль к поверхности **b**(u,v). На этих поверхностях построим две основные грани листового тела. Остальные (боковые) грани построим на линейчатых поверхностях, одной базовой линией которых является граничная линия на поверхности **b**(u,v), а второй — соответствующая ей линия на поверхности **r**(u,v)

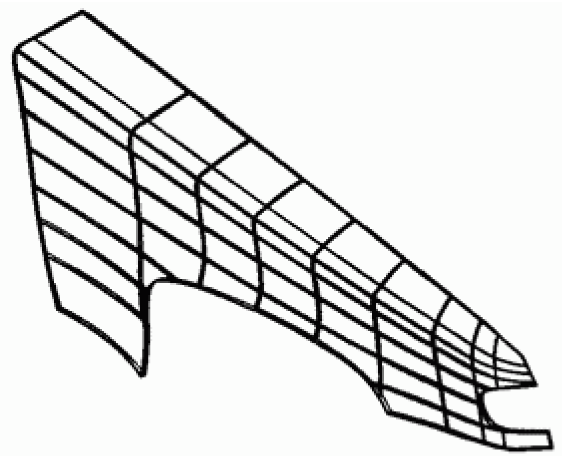


Рис. 4.1. Тело в форме листа

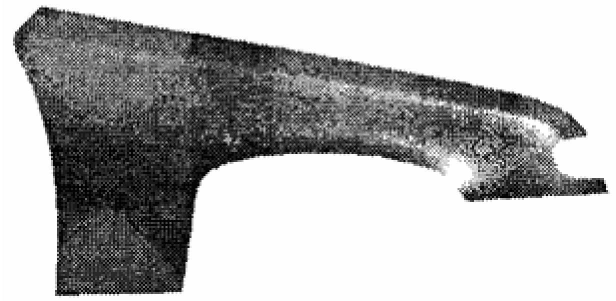


Рис. 4.2. Тело в форме листа

Если область определения параметров поверхности **b**(u, v) имеет прямоугольную форму, то листовое тело будет иметь четыре боковые грани. В общем случае листовое тело будет иметь столько боковых граней, сколько линий содержат двухмерные контуры, описывающие область определения параметров поверхности **b**(u, v).

Направление нормалей основных граней противоположно друг другу в соответствующих точках. Если поверхность **b**(u, v) является замкнутой по одному из параметров, то листовое тело будет иметь форму трубы и топологию тора. Если поверхность замкнута по обоим параметрическим направлениям, то мы получим тело с пустотой внутри. Такое тело будет иметь внешнюю оболочку на основе поверхности **r**(u, v) и внутреннюю оболочку на основе поверхности **b**(u, v).

Векторное изображение тела в форме листа, построенного по ограниченной контурами NURBS поверхности, приведено на рис. 4.1, а его тоновое изображение — на рис. 4.2.

С помощью тел в форме листа можно моделировать детали кузова автомобиля и планера самолета. В структуре данных тела достаточно иметь базовую поверхность **b**(u, v) и толщину тела h. Тело, построенное по сечениям, и тело в форме листа, так же как все рассмотренные выше тела, служат заготовками для моделей деталей. Дальнейшее моделирование деталей связано операциями над телами.

12.5. Булевы операции над телами

Над телами, как и над другими геометрическими объектами, можно выполнять операции — совокупность действий над одним или несколькими исходными телами, которая приводит к рождению нового тела. Одними из основных операций для двух тел являются булевы операции.

Булевыми операциями называют операции объединения, пересечения и вычитания тел, так как они выполняют одноименные операции над внутренними объемами тел (над множествами точек пространства, находящимися внутри тел). Булеву операцию объединения тел будем обозначать формулой $S=S_1 \bigcup S_2$, где S_1 и S_2— исходные тела, S — результирующее тело. Булеву операцию пересечения тел будем обозначать формулой $S=S_1 \bigcap S_2$. Булеву операцию вычитания тел будем обозначать формулой $S=S_1 - S_2$. В порядке следования тел-операндов будем называть их первым телом и вторым телом.

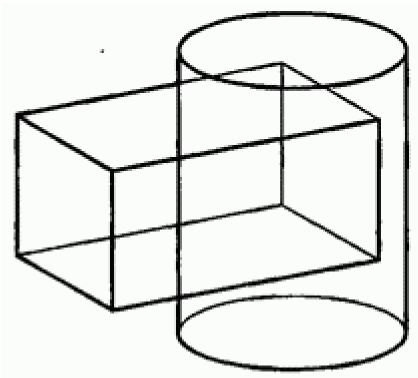


Рис. 5.1. Два исходных тела

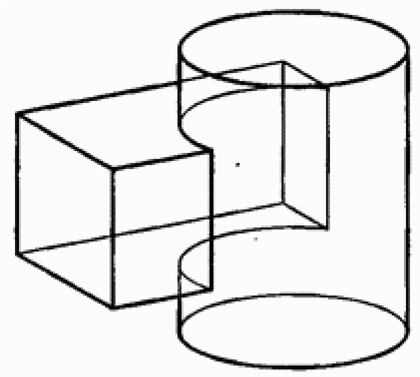


Рис. 5.2. Объединение тел

Результатом операции объединения двух тел является тело, которое содержит точки, принадлежащие внутреннему объему или первого, или второго тела. Результатом операции пересечения двух тел является тело, которое содержит точки, принадлежащие внутреннему объему как первого, так и второго тела.

Результатом операции вычитания двух тел является тело, которое содержит точки, принадлежащие внутреннему объему первого, но не принадлежащие внутреннему объему второго тела.

На рис. 5.1 приведены исходные для булевой операции тела. На рис. 5.2 приведен результат операции объединения тел, на рис. 5.3 приведен результат операции пересечения, на рис. 5.4 и 5.5 приведены результаты операции вычитания тел.

Операцию вычитания тел можно свести к операции пересечения тел; для этого нужно вывернуть второе тело наизнанку и найти точки его объема, одновременно принадлежащие и объему первого тела.

Вывернутое наизнанку тело S будем обозначать S^-.

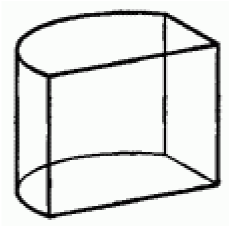


Рис.5.3. Пересечение тел

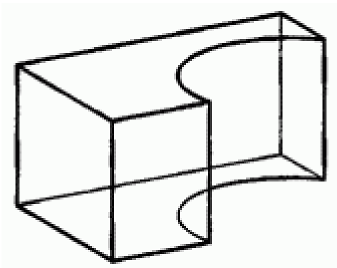


Рис. 5.4. Разность тел

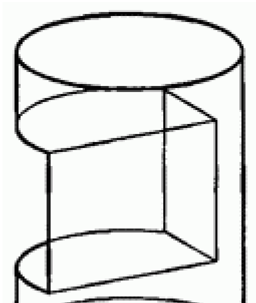


Рис. 5.5. Разность тел

При выворачивании тела наизнанку внутренние стороны граней становятся наружными сторонами, а наружные — внутренними и изменяются направления циклов на противоположные, в результате чего внутренним объемом тела становится та часть пространства, которая до этого находилась снаружи тела. Математически операция вычитания сводится к операции пересечения тел $S=S_1 - S_2 = S_1 \cap \bar{S_2}$. Конструктор при проектировании использует операции объединения и вычитания (они могут называться по-другому, например, операции сварки и сверления), а математический аппарат выполняет соответственно операции объединения и пересечения. Все булевы операции содержат много общего и выполняются по единому алгоритму.

Объединение тел.

Рассмотрим булеву операцию объединения тел. Кратко суть операции можно описать следующим образом: нужно найти линии пересечения граней тел, удалить ту часть первого тела, которая попала внутрь второго тела и ту часть второго тела, которая попала внутрь первого тела, а из всего остального построить новое тело. Операцию условно разобьем на три этапа. На первом этапе построим линии пересечения поверхностей граней и на их базе — новые ребра. Построенные новые ребра будем называть ребрами пересечения, а ребра тел будем называть старыми ребрами. На втором этапе определим точки пересечения новых ребер со старыми ребрами и в этих точках разрежем старые ребра на несколько ребер. На третьем этапе операции перестроим циклы пересекшихся граней. После этого добавим к пересекшимся граням тел грани, топологически связанные ними. Рассмотрим этапы построения тела более подробно.

Первый этап операции объединения тел начнем с того, что построим линии пересечения каждой грани первого тела с каждой гранью второго тела, если таковые имеются. Для этого используем алгоритм пересечения поверхностей граней.

Пусть грани первого тела описываются поверхностями

$$\mathbf{r}_i(u_i, v_i), \qquad i = 1, 2, \ldots, m, \tag{5.1}$$

а грани второго тела описываются поверхностями

$$\mathbf{s}_j(a_j,\, b_j), \qquad j = 1,\, 2,\, \ldots,\, n. \tag{5.2}$$

На базе линий пересечения граней первого и второго тел

$$\begin{gathered} l_{uv}(t) = \left[u_i(t) \quad v_i(t)\right]^{\top}, \qquad l_{uv}(t) \in \mathbf{r}_i(u_i,\, v_i), \\ l_{ab}(t) = \left[a_j(t) \quad b_j(t)\right]^{\top}, \qquad l_{ab}(t) \in \mathbf{s}_j(a_j,\, b_j), \\ t_{\min} \leqslant t \leqslant t_{\max}, \end{gathered} \tag{5.3}$$

построим ребра пересечения. При этом ребрам пересечения дадим направление векторного произведения нормали грани первого тела с нормалью грани второго тела: $t_{edge}=\mathbf{m}_r\times\mathbf{m}_s$. Направление ребра определяется признаком совпадения направления производной линии пересечения с требуемым направлением ребра. За положительное направление нормали грани примем направление наружу тела. Нормаль грани может совпадать с нормалью ее поверхности или иметь противоположное направление в зависимости от признака их совпадения. На рис. 5.6 показаны направления ребер пересечения грани первого тела с двумя гранями второго тела.

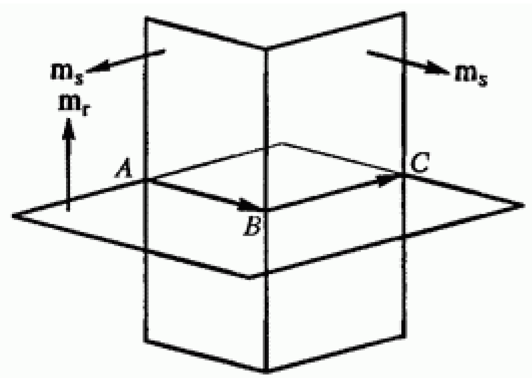


Рис. 5.6. Направление ребер пересечения граней тел

Ребра пересечения должны быть построены так, чтобы они полностью лежали внутри циклов граней исходных тел. Ребра пересечения могут подходить к границам грани только своими концами. В точках А, В и С (рис. 5.6) старые ребра граней должны быть разбиты каждое на два ребра, так как в результирующее тело войдет только часть исходной грани.

На втором этапе разрежем старые ребра тела, к которым подходят ребра пересечения. Резка старого ребра осуществляется путем рассечения кривой, на которой базируется ребро. Из одной кривой получим две кривые, в совокупности заменяющие исходную кривую ребра. Одна из этих кривых останется геометрическим носителем разрезаемого ребра, а на базе второй построим новое ребро, которое получит от исходного ребра всю необходимую информацию. Как было сказано, каждое ребро строится на базе кривой пересечения поверхностей. Кривую пересечения поверхностей составляют две поверхности и две соответствующие им двухмерных кривые. Рассечению подлежат именно двухмерные кривые на двух разных поверхностях. Как до, так и после рассечения эти кривые должны иметь одинаковые области определения параметров и соответствие точек при всех значениях параметра.

Точки пересечения нового ребра со старым ребром грани ищутся как точки пересечения двухмерных кривых, заданных на общей для них плоскости параметров. От каждого ребра в формуле точек пересечения линий участвует по одной двухмерной кривой, входящей в линию пересечения.

Параметры точек пересечения ребер и сами координаты кривых, являющиеся параметрами поверхности, должны быть определены с заданной точностью.

Если разрезаемое ребро базируется на кривой пересечения, заданной отдельными точками (двухмерные линии являются ломаными и точно совпадают в пространстве только в характеристических точках), то прежде чем разрезать такую кривую, нужно в обе линии вставить дополнительные точки, соответствующие точки пересечения трех поверхностей — двух поверхностей, лежащих по обе стороны разрезаемого ребра, и поверхности грани другого исходного тела. Например, если режется линия пересечения поверхностей $\mathbf{r_i}$ (u_i, v_i), $\mathbf{r_n}$ (u_n, v_n) поверхностью $\mathbf{s_j}$ (a_j, b_j), то в линии $\boldsymbol{l}_{uv}(q)=[u_i(q)\ v_i(q)]^T$ и

$\boldsymbol{l}_{uv}(q)=[u_n(q)\ v_n(q)]^T$ нужно вставить дополнительные двухмерные характеристические точки, соответствующие пересечению поверхностей $\mathbf{r_i}\ (u_i, v_i)$, $\mathbf{r}_n\ (u_n, v_n)$, $\mathbf{s}_j\ (a_j, b_j)$. Задача пересечения трех поверхностей сводится к решению системы шести скалярных уравнений относительно шести параметров $u_i, v_i, u_n, v_n, a_j, b_j$. Начальное приближение решения известно достаточно точно.

Так как каждое ребро исходных тел входит в циклы двух смежных граней, то после резки ребер исходных тел необходимо произвести корректировку этих циклов с учетом разрезанных ребер.

После первых двух этапов мы получили совокупность ребер пересечения, ориентированных определенным образом, стыкующихся друг с другом и с ребрами исходных тел только в вершинах. Далее необходимо перестроить циклы пересекшихся граней.

Третий этап завершает булеву операцию. Для того чтобы каждую из показанных на рис. 5.6 граней разрезать на части, нужно перестроить ее циклы и в соответствии с циклами изменить контуры, описывающие область определения параметров поверхности грани. На рис. 5.7 показаны две пересекшиеся грани (тонкими линиями со стрелками показано направление циклов граней исходных тел) и ребро пересечения. На рис. 5.8 показаны те части граней, которые войдут в объединение тел.

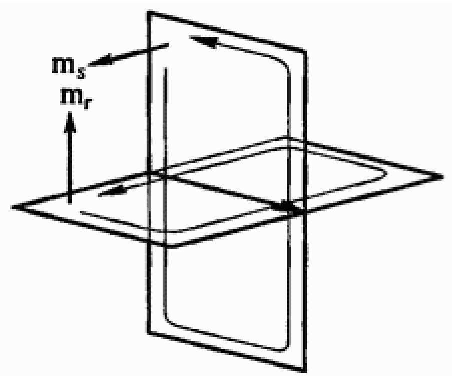


Рис. 5.7. Исходные грани тела

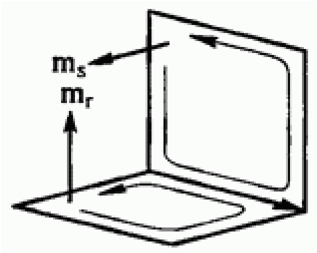


Рис. 5.8. Обрезанные грани

Стрелками показано направление перестроенных циклов граней тел. Каждый цикл представляет собой список ребер в порядке их следования и список флагов ориентации этих ребер в цикле.

Из рисунка видно, что при принятой ориентации ребер пересечения ($t_{edge}=\mathbf{m}_r\times\mathbf{m}_s$) в циклы граней первого тела они войдут с отрицательным флагом, а в циклы граней второго тела они войдут с положительным флагом.

Старые ребра исходных тел, которые сохранятся в результирующем теле, войдут в перестроенные циклы, сохранив свои флаги. При перестройке циклов пересекшихся граней будем использовать следующий алгоритм.

Рассмотрим одну из двух пересекшихся граней, принадлежащих первому телу. Берем любое ребро пересечения рассматриваемой грани и начинаем с него составлять список ребер цикла. Ребро пересечения должно войти в цикл грани первого тела с отрицательным флагом, следовательно, цикл будет иметь направление, противоположное первому ребру. Для продолжения цикла среди ребер пересечения и среди старых ребер грани ищем все ребра, стыкующиеся с данным ребром в его начальной вершине. Среди найденных ребер выберем то, которое лежит слева от остальных (заворачивает влево на больший угол по сравнению с другими найденными ребрами, если смотреть вдоль цикла с наружной стороны грани). Выбранное ребро ставим в список цикла. Если выбранное ребро является старым, то оно сохраняет свой флаг в цикле. Если выбранное ребро есть ребро пересечения, то оно получит отрицательный флаг. Ребра пересечения обладают преимущественным правом по отношению к старым ребрам быть выбранными. То есть, если левее других оказались два совпадающих ребра, одно из которых является старым, а другое — ребром пересечения, то для продолжения цикла должно быть выбрано ребро пересечения.

Процесс перестроения цикла грани будем продолжать до тех пор, пока цикл не замкнется. На этом построение очередного цикла заканчивается. Если у рассматриваемой грани при построении цикла использованы не все ребра пересечения, то с любого из оставшихся ребер пересечения начинаем строить еще один цикл грани. Циклы перестраиваем до тех пор, пока не используем все ребра пересечения. Таким способом мы построим в общем случае несколько новых циклов рассматриваемой грани первого тела.

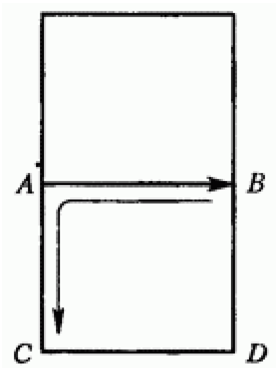


Рис. 5.9

Поиск стыкующихся ребер и определение угла поворота в точке стыка удобно выполнять по двухмерным кривым ребер грани. При этом можно работать даже с такими ребрами, которые в пространстве стянуты в точку (например, ребро в вершине конуса или полюсе сферы). На рис. 5.9 показано, что с новым ребром В А в точке А стыкуются несколько ребер. В данном случае для продолжения цикла, начатого с ребра пересечения В А, следует выбрать ребро АС.

Вновь построенные циклы рассортируем по группам, каждая из которых состоит из внешнего цикла и входящих в него внутренних циклов. У рассматриваемой грани могут остаться нетронутыми один или несколько старых циклов. Нетронутыми мы будем называть старые циклы грани, ни одно ребро которых не вошло в перестроенные циклы. Среди старых нетронутых циклов отберем те, которые необходимо включить в состав описания перестроенной грани. Ими являются старые внутренние циклы грани, лежащие внутри новых внешних циклов. Еще нужно определить, не потребуется ли включить в результат старый внешний цикл грани, если он остался нетронутым. Это необходимо сделать, если для некоторых новых внутренних циклов не найден новый внешний цикл и они лежат внутри старого внешнего цикла. Сортировку циклов удобно выполнять с помощью двухмерных контуров, соответствующих каждому циклу.

Если в результате сортировки внешних циклов получилось больше одного, то это означает, что из исходной грани в результате операции образовалось несколько граней. На рис.5.10 а - 5.13 а приведены варианты исходных граней первого тела.

На рис. 5.10 б-5.13 б. приведены грани с добавлением ребер пересечения (ребра пересечения выделены). На рис.5.10 в-5.13 в приведены грани, которые получились в результате операции.

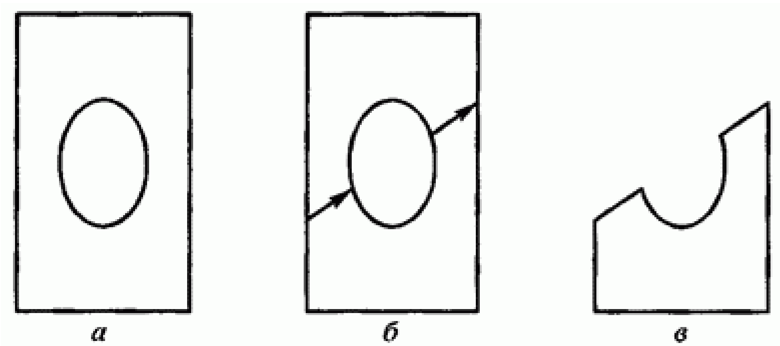


Рис. 5.10. Исходная грань (а), грань с. добавлением ребер пересечения (б), результат операции (б)

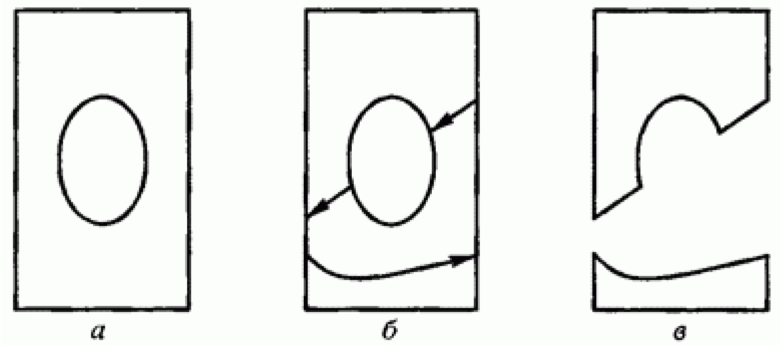


Рис. 5.11. Исходная грань (а), грань с добавлением ребер пересечения (б), результат операции (в)

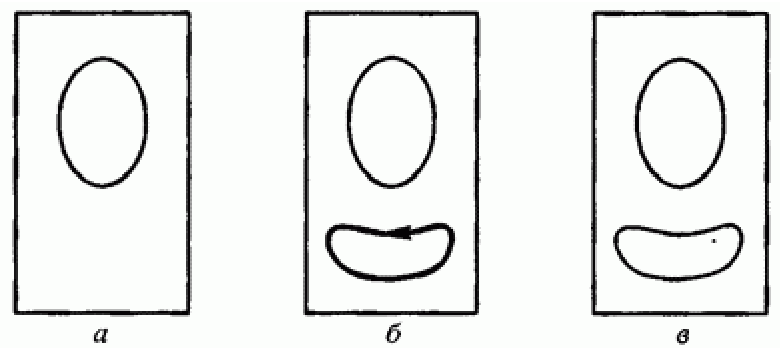


Рис. 5.12. Исходная грань (а), грань с добавлением ребер пересечения (б), результат операции (в)

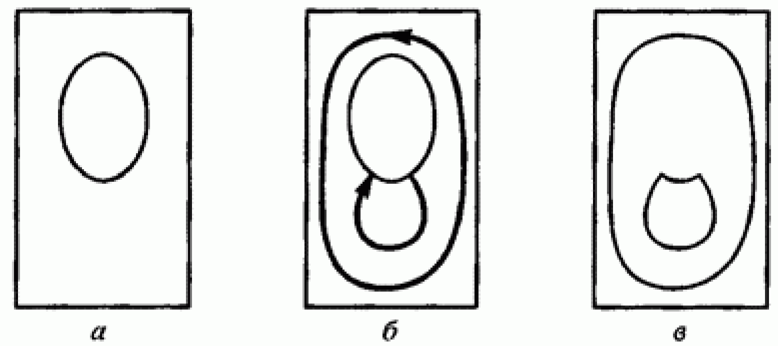


Рис. 5.13. Исходная грань (а), грань с добавлением ребер пересечения (б), результат операции (в)

На рис. 5.10 грань с двумя циклами была разрезана и получилась одна грань с одним циклом. На рис. 5.11 грань с двумя циклами дала две грани. В примере, приведенном на рис. 5.12, потребовалось использовать старый внешний и внутренний циклы. На рис. 5.13 из одной грани получено две грани, причем для одной из них потребовалось использовать старый внешний цикл.

Описанное перестроение циклов выполняется для каждой пересекшейся грани первого тела.

С пересекшимися гранями второго тела поступим аналогично тому, как мы поступили с гранями первого тела, но с одной небольшой разницей. Ребра пересечения должны войти в перестроенные циклы граней второго тела с положительным флагом (в перестроенные циклы граней первого тела они вошли с отрицательным флагом). В этом состоит отличие перестроения циклов граней второго тела. Все остальные действия над гранями первого и второго объединяемых тел одинаковы.

Мы перестроили пересекшиеся грани объединяемых тел. Все их включим в оболочку результирующего тела. Для получения результирующего тела остается к этим граням в оболочку тела добавить не пересекшиеся в операции грани, которые топологически связаны с пересекшимися гранями. Для этого будем брать последовательно ребра, входящие в оболочку нового тела, и включать в оболочку смежные грани ребер (если они там отсутствуют).

Продолжив эти действия для ребер всех добавленных граней, получим оболочку результирующего тела.

Пересечение тел.

Коротко суть булевой операции пересечения тел можно описать следующим образом: нужно найти линии пересечения тел, удалить ту часть первого тела, которая не попала внутрь второго, и ту часть второго тела, которая не попала внутрь первого, а из всего остального построить новое тело.

Эта операция имеет много общего с операцией объединения тел. Вернемся к рис. 5.6. На нем приведены пересекающиеся грани: одна грань первого тела и две грани второго тела. Грани режут друг друга, так что в результирующую оболочку войдут только части этих граней.

В пересечение тел войдет часть грани первого тела, лежащая внутри второго тела, и часть грани второго тела, лежащая внутри первого тела (рис. 5.14) (в объединение тел вошла часть грани первого тела, лежащая вне второго тела, и часть грани второго тела, лежащая вне первого тела).

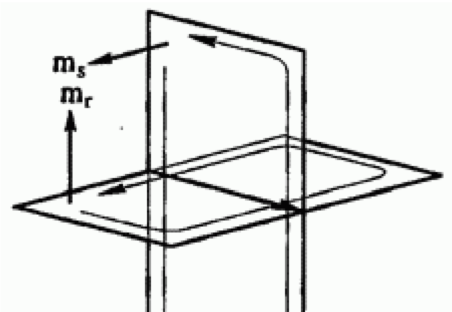


Рис. 5.14. Исходные грани

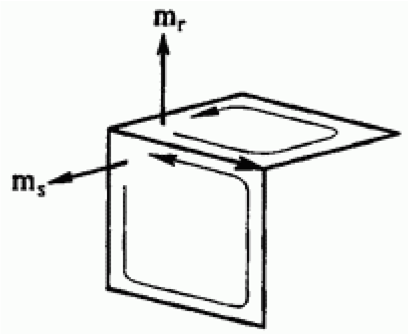


Рис. 5.15. Перестроенные грани пересечения тел

В этом и состоит основное отличие операций объединения и пересечения тел.

Операцию разобьем на три этапа. Первый и второй этапы операции пересечения тел полностью совпадают с соответствующими этапами операции объединения тел.

Третий этап операции пересечения тел выполняется аналогично третьему этапу операции объединения тел, но имеет одно отличие в том, с каким флагом входят в циклы ребра пересечения. Ребра пересечения входят в перестроенные циклы граней первого тела с положительным флагом, а в перестроенные циклы граней второго тела — с отрицательным флагом (в объединении тел флаги ребер пересечения в циклах имеют противоположное значение). Все остальные действия над гранями обоих тел в обеих операциях одинаковы.

Разность тел.

Коротко суть булевой операции вычитания тел можно описать следующим образом: нужно найти линии пересечения тел, удалить ту часть первого тела, которая попала внутрь второго, и ту часть второго тела, которая не попала внутрь первого, а из всего остального построить новое тело.

Булева операция вычитания тел сводится к булевой операции пересечения уменьшаемого тела и вывернутого наизнанку вычитаемого тела. Вывернутое наизнанку тело мы получим из исходного тела путем переориентации направлений нормалей граней и направлений циклов граней. Переориентация направления нормали грани производится изменением признака совпадения нормали поверхности и нормали ее грани. Переориентация направления цикла грани производится перестроением списка ребер (изменением на обратный порядок следования ребер в списке) и заменой на противоположные флагов ребер в списках. Для вывернутого тела внутренним объемом является часть пространства, находящаяся вне его оболочки. Поэтому при пересечении уменьшаемого тела и вывернутого наизнанку вычитаемого тела результирующая оболочка будет содержать ту часть объема уменьшаемого тела, которая лежит вне вычитаемого тела.

Пересекающиеся ребра.

Наиболее трудоемким и требующим определенной точности в процессе выполнения булевых операций является построение ребер пересечения. Ребра пересечения не должны иметь выступающих за пределы грани частей, они должны обязательно стыковаться или друг с другом или со старыми ребрами граней. При корректном выполнении операции пересечения поверхностей эти условия обеспечиваются. Ребра пересечения не должны пересекать друг друга вне крайних точек.

В большинстве случаев новые ребра не пересекают друг друга, но в некоторых частных случаях это возможно. На рис. 5.16 показаны два цилиндрических тела одинакового диаметра, оси которых пересекаются — крест из цилиндров. При булевом объединении этих тел возможна ситуация, когда будут построены всего два замкнутых ребра пересечения. Такие ребра имеют две точки пересечения: А и В, по крайней мере, одна из которых не будет совпадать с вершинами ребер, а будет лежать где-то на ребре. Точки пересечения ребер лежат в точках касания цилиндров. Эти ребра должны быть разрезаны в точках А и В и у их частей должна быть уточнена ориентация, так как при прохождении точки касания поверхностей в данном случае векторное произведение нормалей к ним, по которому ориентируются новые ребра, меняет свое направление на противоположное. Обнаружить пересечение новых ребер можно по пересечению кривых на поверхностях, из которых состоит линия пересечения.

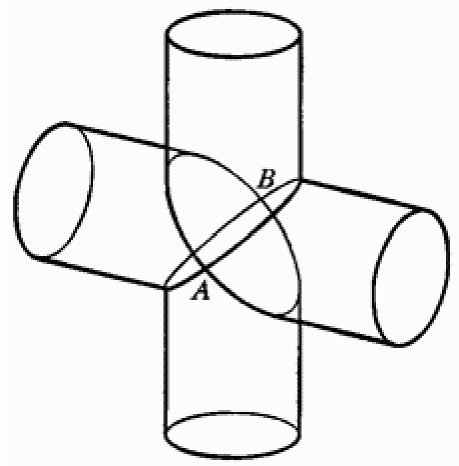


Рис. 5.16. В точках А и В нормали граней совпадают

Совпадающие ребра.

Описанный алгоритм выполнения булевых операций работает корректно, если ребра пересечения не совпадают с ребрами исходных тел. В противном случае он нуждается в уточнении. Рассмотрим примеры.

В булевых операциях возможен случай, когда какая-либо грань одного тела пересекает другое тело по его ребру. Возможен также случай, когда при построении ребер пересечения граней одного тела с гранями другого тела мы получим два новых ребра, совпадающих в пространстве друг с другом и с ребром одного из тел. Все ребра являются разными, так как в них стыкуются разные грани. На рис. 5.17 приведены два тела, при выполнении булевой операции над которыми, ребра пересечения совпадут с ребрами меньшего тела. Следуя общему алгоритму, мы в данном случае получим восемь ребер пересечения, половина из которых должна быть опущена (или не должна быть построена).

Правило для ребер пересечения.

При наличии ребер пересечения, совпадающих с ребрами граней исходных тел, и в некоторых других случаях будем выполнять следующую проверку. Построим в плоскости каждой из двух пересекаемых граней по два вектора, ортогональные ребру пересечения. На рис. 5.18 это векторы $\mathbf{a}_1$ и $\mathbf{b}_1$ для грани с нормалью $\mathbf{m}_1$ и векторы $\mathbf{a}_2$ и $\mathbf{b}_2$ для грани с нормалью $\mathbf{m}_2$. Векторы $\mathbf{a}_1$ и $\mathbf{a}_2$ поворачивают влево от ребра пересечения в соответствующих плоскостях, а векторы $\mathbf{b}_1$ и $\mathbf{b}_2$ поворачивают вправо от ребра пересечения в соответствующих плоскостях.

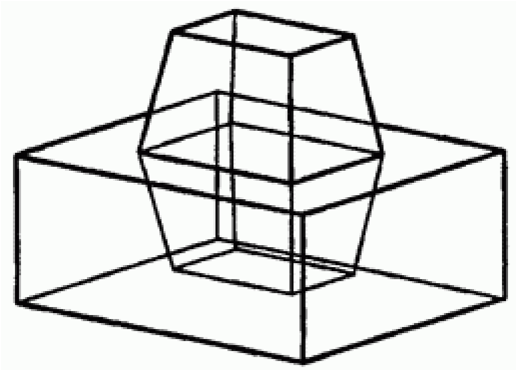


Рис. 5.17. Совпадение ребер пересечения с ребрами меньшего тела

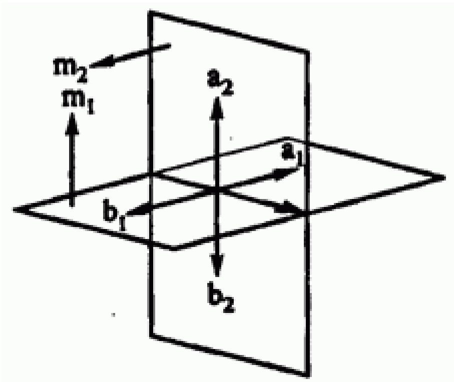


Рис. 5.18. Пересечение граней

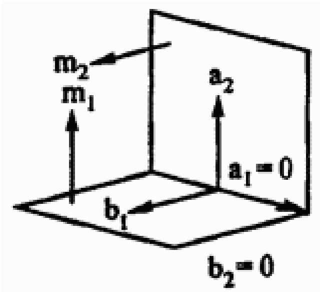


Рис. 5.19. Пересечение граней по ребру

Если грань не имеет продолже ния за ребро пересечения (ребро пересечения частично или полностью совпадает с ребром тела), то соответствующий вектор положим равным нулю (рис.5.19). Используя векторы $\mathbf{a}_1$**,** $\mathbf{a}_2$ **,** $\mathbf{b}_1$, $\mathbf{b}_2$ и нормали $\mathbf{m}_1$, $\mathbf{m}_2$ мы можем определить, будет ли данное ребро пересечения использовано в операции или оно должно быть опущено. Обратим внимание на следующее обстоятельство.

В операции объединения тел грань первого тела мы сможем перестроить, еслиона имеет продолжение справа от ребра пересечения вне второго тела, а грань второго тела мы сможем перестроить, если она имеет продолжение слева от ребра пересечения вне первого тела.

Таким образом, в булевой операции объединения тел для ребра пересечения должны быть выполнены условия

$$\mathbf{m}_1 \cdot \mathbf{a}_2 > 0 \quad \text{и} \quad \mathbf{m}_2 \cdot \mathbf{b}_1 > 0. \qquad (5.7)$$

В противном случае рассматриваемое ребро пересечения в булевой операции объединения тел строить не следует (если оно построено, то должно быть опущено).

В булевой операции пересечения тел грань первого тела мы сможем перестроить, если она имеет продолжение слева от ребра пересечения внутрь второго тела, а грань второго тела мы сможем перестроить, если она имеет продолжение справа от ребра пересечения внутрь первого тела. Таким образом, в булевой операции пересечения тел для ребра пересечения должны быть выполнены условия

$$\mathbf{m}_1 \cdot \mathbf{b}_2 < 0 \quad \text{и} \quad \mathbf{m}_2 \cdot \mathbf{a}_1 < 0. \tag{5.8}$$

В противном случае рассматриваемое ребро пересечения в булевой операции пересечения тел строить не следует (если оно построено, то должно быть опущено).

Как уже было сказано, ребра пересечения обладают преимущественным правом по отношению к старым ребрам тела быть включенными в перестраиваемый цикл.

Так как булева операция вычитания тел сводится к операции пересечения тел, то для нее должно выполнятся правило (5.8) для вывернутой наизнанку оболочки тела.

Принадлежность точки пространству внутри тела.

Для ответа на вопрос, внутрь или вне тела продолжается от ребра пересечения грань другого тела, нужно уметь определять, принадлежит ли некоторая точка **p** пространству внутри тела — классифицировать точку относительно тела. Нам известно, что нормаль каждой грани направлена вне объема тела. Для точки **p** найдем ближайшую точку $\mathbf{p}_0$ на ближайшей грани. Построим вектор **n** из найденной точки $\mathbf{p}_0$ в точку **p**. Вычислим нормаль **m** тела в точке $\mathbf{p}_0$. Если точка $\mathbf{p}_0$ лежит на границе граней (является точкой ребра или вершины), то в качестве нормали m тела возьмем среднюю нормаль граничащих граней. Если $\mathbf{m} \cdot \mathbf{n} \geq 0$, то точка **p** принадлежит внутреннему пространству тела. В противном случае — нет.

Перекрывающиеся грани.

В булевых операциях довольно часто можно встретить случаи частичного совпадения некоторых граней двух исходных тел. На рис. 5.20 приведены два тела, некоторые грани которых частично перекрывают друг друга.

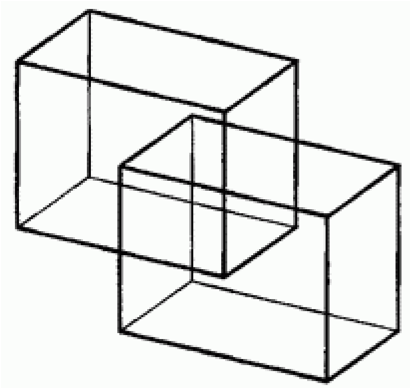


Рис. 5.20. Совпадающие грани тел

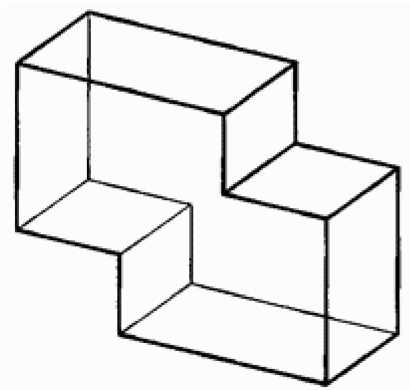


Рис. 5.21. Объединение тел

В данном случае не все ребра пересечения граней войдут в булев результат; некоторые ребра должны быть опущены (или не должны строиться). Результат объединения тел приведен на рис. 5.21.

Аналогичная ситуация возникает при выполнении булевой операции над двумя соосными цилиндрами одинакового радиуса и еще во многих случаях. Грани могут перекрываться полностью, частично или всего лишь по одной линии. Нормали $\mathbf{m}_1$ и $\mathbf{m}_2$ частично перекрывающихся граней должны быть одинаково направлены в общих точках. Перекрывающимися являются грани, которые можно перенести на одну общую поверхность. Таких граней может быть больше двух. При наличии у тел-операндов перекрывающихся граней все ребра пересечения этих граней должны быть проверены на предмет присутствия их в булевом результате. В операции объединения тел нам потребуются только те ребра пересечения перекрывающихся граней, смежные грани которых имеют продолжение вне одного из тел. В операции пересечения тел нам потребуются только те ребра пересечения перекрывающихся граней, смежные грани которых имеют продолжение внутрь одного из тел.

Циклы перекрывающихся граней тел должны быть перестроены по несколько иному алгоритму, чем для остальных граней. Прежде всего, все ребра этих граней должны быть приведены к одному общему носителю (к поверхности одной из перекрывающихся граней). В операции объединения тел старые ребра перекрывающихся граней первого тела, лежащие внутри второго тела, и старые ребра перекрывающихся граней второго тела, лежащие внутри первого тела, не войдут в результирующее тело. В операции пресечения тел старые ребра перекрывающихся граней первого тела, лежащие вне второго тела, и старые ребра перекрывающихся граней второго тела, лежащие вне первого тела, также не войдут в результирующее тело.

Построение некоторого цикла перекрывающихся граней начнем с ребра, которое точно должно войти в результат. Для продолжения цикла среди ребер пересечения и среди старых оставшихся ребер грани найдем все ребра, стыкующиеся с данным ребром в его соответствующей вершине. Среди найденных ребер выберем то, которое лежит справа от остальных (заворачивает вправо на больший угол по сравнению с другими найденными ребрами, если смотреть вдоль цикла с наружной стороны грани). Выбранное ребро ставим в список цикла. Отличие алгоритма состоит в том, что при продолжении цикла мы выбираем самое правое (а не самое левое) ребро.

Тела с несколькими оболочками.

Если какое-либо из тел-операндов имеет пустоты и, соответственно, описывается несколькими оболочками, то нетронутые операцией внутренние (внешние) оболочки должны быть проверены на вхождение в оболочку булева результата. В общем случае результирующее тело также может иметь несколько оболочек.

Дерево построения тел.

Булевы операции над телами показывают необходимость привлечения топологических понятий для построения тел.

Рассмотренные в предыдущих параграфах тела будем называть простыми, в отличие от тел, полученных в результате операций, которые будем называть сложными. Структура данных простых тел содержит минимум информации, по которой могут быть построены все грани тела. В структуре данных тела, полученного в результате булевой операции, положим структуры данных исходных тел и тип булевой операции. Оболочки тела всегда могут быть построены по этой информации. Таким образом, структура данных тела, полученного в результате многократного выполнения булевых операций, будет содержать **дерево построения**. Результирующее тело находится в корне дерева, а его ветви начинаются в простых телах. Пример дерева приведен на рис. 5.22. В узлах дерева находятся тела. Дерево имеет несколько ярусов. Операции между телами обозначены соответствующими знаками. Операции выполняются между телами одного яруса.

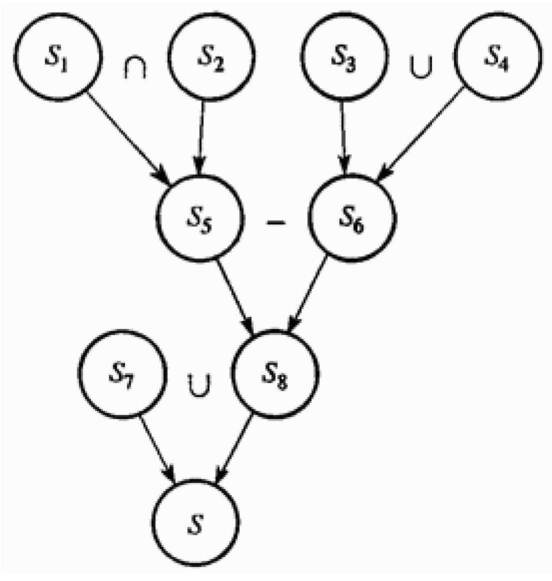


Рис. 5.22. Дерево построения тела

Структура данных тела в виде дерева построения может быть использована для всех тел — как для тел, полученных в результате некоторой операции, так и для простых тел.

12.6. Резка тела поверхностью

Если поверхность полностью пересекает тело, то тело можно разрезать этой поверхность на две части. Результатом такой операции является новое тело, представляющее собой часть исходного тела, лежащую по одну или другую сторону режущей поверхности. Положительной стороной будем называть сторону поверхности, в которую направлены нормаль к ней (если мы смотрим на положительную сторону поверхности, то нормаль направлена на нас). Другую сторону поверхности будем называть отрицательной.

Операцию резки тела поверхностью мы сведем к одной из булевых операций. На базе поверхности построим оболочку. Эта оболочка будет состоять из одной грани. Нормаль этой грани пусть совпадает с нормалью поверхности. Построенную незамкнутую оболочку мы будем рассматривать как некоторое незаконченное тело.

Пусть требуется построить ту часть тела, которая лежит с положительной стороны режущей поверхности. В этом случае выполним булеву операцию вычитания из этого тела незаконченного тела, построенного на основе режущей поверхности.

Пусть требуется построить ту часть тела, которая лежит с отрицательной стороны режущей поверхности. В этом случае выполним булеву операцию пересечения этого тела с незаконченным телом, построенным на основе режущей поверхности.

На рис. 6.1 показаны тело, режущая поверхность, ее нормаль и ребра пересечения граней тела с поверхностью. При резке тела поверхностью мы выполняем операцию пересечения тела с частью пространства, лежащей по ту или иную сторону поверхности.

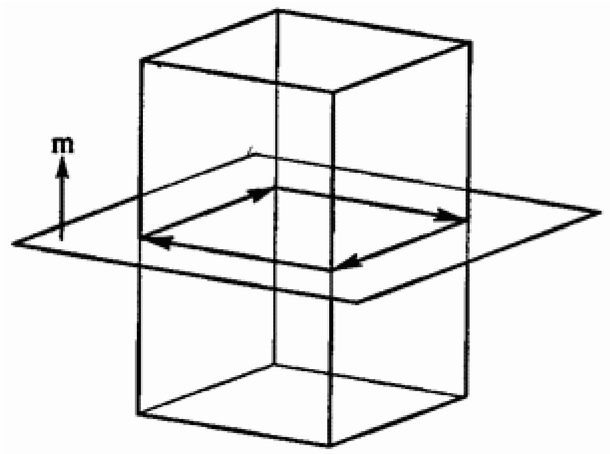


Рис. 6.1. Направление ребер пересечения

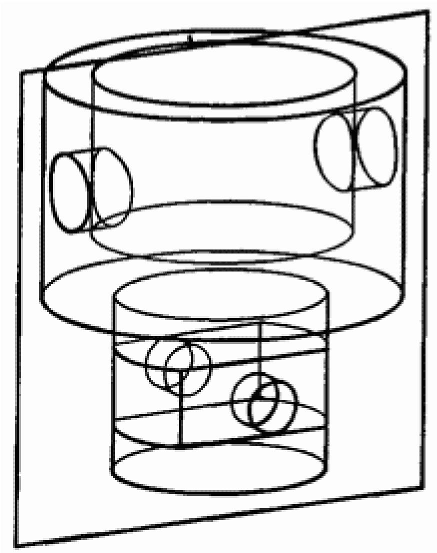


Рис. 6.2. Исходное тело и режущая поверхность

В частном случае тело можно разрезать плоскостью. На рис. 6.2 приведены исходное тело и режущая плоскость. На рис. 6.3 приведено разрезанное плоскостью тело.

С помощью рассматриваемой операции можно получить сложный разрез тела. Для этого оболочка незаконченного тела должна состоять из нескольких стыкующихся между собой граней.

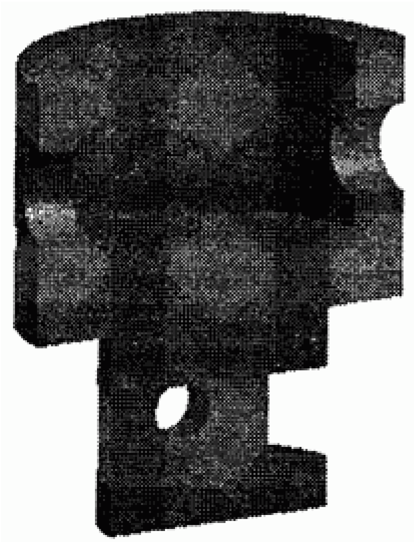


Рис. 6.3. Разрезанное тело

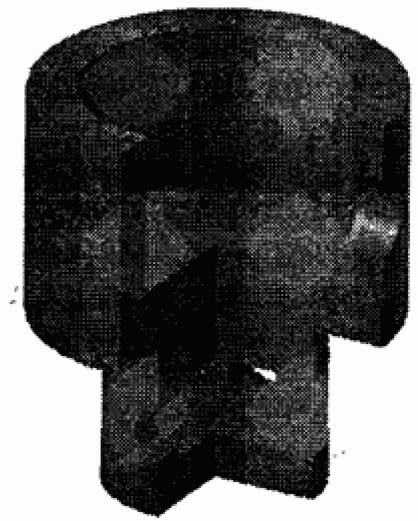


Рис. 6.4. Сложный разрез тела

Для получения сложного разреза возьмем составную кривую на плоскости и построим по ней незаконченное тело выдавливания так, чтобы оно пересекало заданное тело требуемым образом. Далее выполним булеву операцию над заданным телом и незаконченным телом. Результат сложного разреза тела показан на рис. 7.4.

12.7. Построение симметричного тела

Пусть имеется тело и плоскость. Построим тело, симметричное данному телу относительно данной плоскости. Исходное тело будем называть базовым. Симметричное тело будет зеркальной копией данного тела. Геометрия тела описывается точками, кривыми и поверхностями, которые в конечном итоге описываются точками векторами и скалярами. Поэтому построение симметричного тела в конечном итоге сводится к преобразованию симметрии радиус-векторов точек, свободных векторов и скаляров. Скалярные величины при преобразовании симметрии не изменяются, свободные векторы меняют свое направление, а точки — свое положение.

Пусть плоскость симметрии описывается формулой

$$\mathbf{p}(x, y) = \mathbf{p}_0 + x\mathbf{i}_1 + y\mathbf{i}_2,$$

где $\mathbf{i}_1$ и $\mathbf{i}_2$— ортогональные векторы единичной длины. Тогда матрица преобразования симметрии тела относительно этой плоскости определяется формулой

$$\mathbf{A} = 2\mathbf{i}_1\mathbf{i}_1 + 2\mathbf{i}_2\mathbf{i}_2 - \mathbf{E}, \quad (7.1)$$

где $\mathbf{i}_1\ \mathbf{i}_1$ и $\mathbf{i}_2\ \mathbf{i}_2$ — диадные произведения векторов. Свободный вектор $\mathbf{v}_0$ после преобразования симметрии относительно этой плоскости будет описываться радиус-вектором

$$\mathbf{v} = \mathbf{A} \cdot \mathbf{v}_0. \quad (7.2)$$

Произвольная точка $\mathbf{r}_0$ после преобразования симметрии относительно этой плоскости будет описываться радиус-вектором

$$\mathbf{r} = \mathbf{p}_0 + \mathbf{A} \cdot (\mathbf{r}_0 - \mathbf{p}_0). \quad (7.3)$$

После преобразования симметрии все точки и линии на поверхностях останутся неизменными. А вот нормали поверхностей изменят свое направление на противоположное. Там, где нормаль грани и нормаль ее поверхности в исходном теле совпадали по направлению, в зеркальной копии будут иметь противоположное направление, и наоборот: там, где нормаль грани и нормаль ее поверхности в исходном теле не совпадали по направлению, в зеркальной копии будут иметь одинаковое направление. Кроме того, циклы граней зеркальной копии будут иметь направление, противоположное своим оригиналам. Поэтому в гранях зеркальной копии нужно произвести изменение признаков совпадения нормали поверхности и нормали ее грани на противоположные значения и переориентировать циклы. Переориентация направления цикла грани производится перестроением списка ребер (порядок следования ребер в списке обратный) и заменой флагов ребер в списках на противоположные флаги. Такая же переориентация производилась и при выворачивании тела наизнанку. Если выполнить только преобразование геометрических данных тела по матрице (7.1), то получим зеркальное отражение тела, вывернутое наизнанку. Таким образом, построение симметричного тела сводится к преобразованию его копии по матрице (7.1) и выворачивании ее наизнанку (переориентации его граней).

Симметрия части тела.

Пусть плоскость симметрии пересекает тело. Представим, что оболочка тела разрезана плоскостью, одна из отрезанных частей удалена, по другой части выполнена зеркальная копия и склеена с ней. Мы получим тело, симметричное относительно плоскости, состоящее из двух половинок, одна из которых совпадает с частью исходного тела.

Плоскостью симметрии может служить плоская грань исходного тела. Построение такого симметричного тела имеет много общих моментов с резкой тела на части и операцией объединения тел.

Не теряя общности, будем строить симметричное тело по его части, находящейся с положительной стороны плоскости — с той стороны, в которую направлена нормаль плоскости. Операцию условно разобьем на три этапа.

На первом этапе построим линии пересечения поверхностей граней тела с плоскостью. Пусть пересеченные плоскостью грани тела описываются поверхностями $\mathbf{r}_i(u_i, v_i)$, $i=1,2,..,n$. Тогда каждая линия пересечения будет состоять из пары двухмерных кривых

$$l_{uv}(t) = \begin{bmatrix} u_i(t) & v_i(t) \end{bmatrix}^\top, \qquad l_{uv}(t) \in \mathbf{r}_i(u_i, v_i),$$
$$l_{xy}(t) = \begin{bmatrix} x(t) & y(t) \end{bmatrix}^\top, \qquad l_{xy}(t) \in \mathbf{p}(x, y), \tag{7.4}$$

имеющих общий параметр t. Первая кривая лежит на поверхности $\mathbf{r}_i(u_i, v_i)$, а вторая — на плоскости симметрии. На базе линий пересечения построим новые ребра (ребра пересечения). С помощью признака совпадения направления ребра и его кривой пересечения ребрам пересечения дадим направление векторного произведения нормали грани тела с нормалью к плоскости: $t_{edge} = \mathbf{m}_r \times \mathbf{m}_p$.

На втором этапе определим точки пересечения новых и старых ребер тела, в этих точках построим вершины и этими вершинами разрежем старые ребра на несколько ребер. Резка ребер описана в булевых операциях над телами.

Далее построим симметричную относительно плоскости копию части оболочки тела, лежащей над плоскостью. Для этого построим копию части оболочки, преобразуем ее по матрице (7.1) и вывернем ее наизнанку. При выворачивании оболочки нужно произвести изменение признаков совпадения нормали поверхности и нормали ее грани на противоположные значения и переориентировать циклы.

Остается сшить симметричные половинки тела по ребрам пересечения тела с плоскостью симметрии. Но прежде выполним замену линий на плоскости, входящих в кривые пересечения построенных ребер, на линии на симметричной оболочке. Для этого вместо кривых (7.4) в ребра положим кривые

$$l_{uv}(t) = \left[u_i(t) \quad v_i(t)\right]^{\mathsf{T}}, \qquad l_{uv}(t) \in \mathbf{r}_i(u_i, v_i),$$
$$l_{uv}'(t) = \left[u_i(t) \quad v_i(t)\right]^{\mathsf{T}}, \qquad l_{uv}'(t) \in \mathbf{r}_i'(u_i, v_i). \tag{7.5}$$

Первую двухмерную кривую $l_{uv}(t)$ кривой пересечения сохраним нетронутой, а вместо второй возьмем двухмерную кривую $l_{uv}'(t)$ являющуюся точной копией $l_{uv}(t)$ и лежащую на симметричной копии поверхности $\mathbf{r}_i(u_i, v_i)$. Теперь каждое ребро пересечения будет базироваться на двух симметричных половинках искомого тела.

На третьем этапе перестроим циклы пересеченных плоскостью граней тела. Перестроение циклов грани подробно описано в булевой операции объединения тел. Каждое ребро пересечения должно войти в цикл исходной грани с отрицательным флагом, а цикл ее симметричной части — с положительным флагом.

К пересеченным граням тела добавим непересеченные грани, которые топологически связаны с первыми, и соответствующие симметричные копии. Так мы получим оболочку симметричного тела.

При построении симметричной оболочки следует учитывать все случаи совпадения ребер пересечения и старых ребер тала и частичное перекрытие граней, описанные в булевых операциях над телами. Кроме того, следует учесть наличие нескольких оболочек у исходного тела.

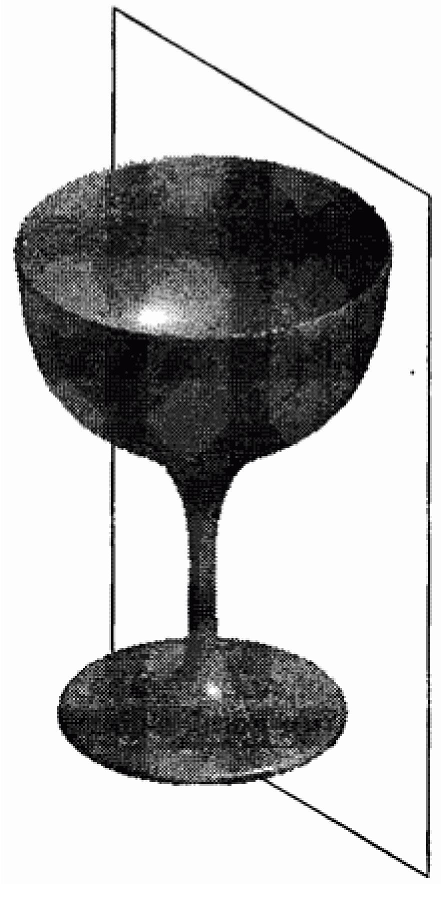


Рис. 7.1. Тело и плоскость симметрии

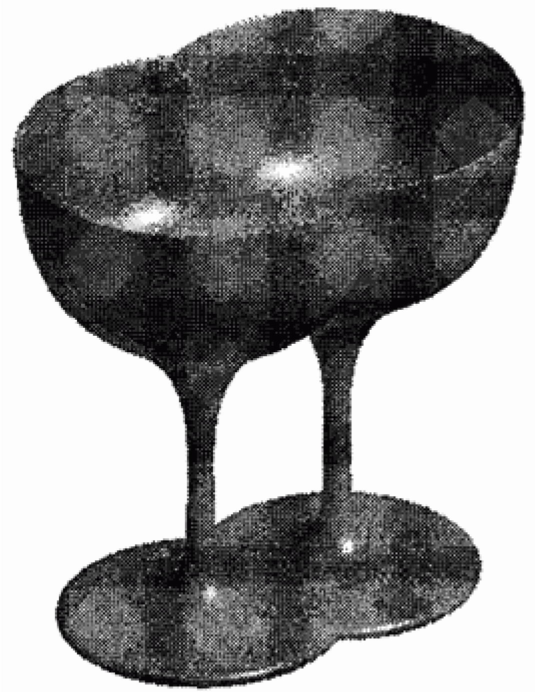


Рис. 7.2. Симметричное тело

На рис. 7.1 приведено исходное тело и плоскость симметрии. На рис. 7.2 приведен результат построения симметричного тела.

12.8. Построение эквидистантной оболочки тела

По данному телу можно построить тело с эквидистантной оболочкой. Оболочка нового тела расположена на заданном расстоянии по нормали от оболочки исходного тела. Это расстояние будем называть параметром эквидистанты и обозначим через h. Тело, по которому строится тело с эквидистантной оболочкой, будем называть базовым, а новое тело будем называть эквидистантным. Параметр эквидистанты этой операции может принимать как положительные, так и отрицательные значения. Если $h > 0$, то базовое тело располагается внутри эквидистантного тела. Если $h < 0$, то эквидистантное тело располагается внутри базового тела. Недопустимыми значениями параметра являются такие, при которых оболочка нового тела получается самопересекающейся или вырожденной. Процесс построения оболочки эквидистантного тела условно разобьем на четыре этапа.

На первом этапе для каждой грани базового тела построим эквидистантную грань. Эквидистантная грань базируется на поверхности, эквидистантной к соответствующей поверхности базового тела. На рис. 8.1 приведены три грани базового тела, имеющие общую вершину А, и эквидистантные им поверхности.

Каждая эквидистантная поверхность должна быть продолжена до пересечения с соседними эквидистантными поверхностями. Радиус-вектор эквидистантной поверхности определяется известной формулой. На продолжении эквидистантной поверхности за пределы области определения параметров ее радиус-вектор будем вычислять по одной из известных формул в зависимости от замкнутости базовой поверхности.

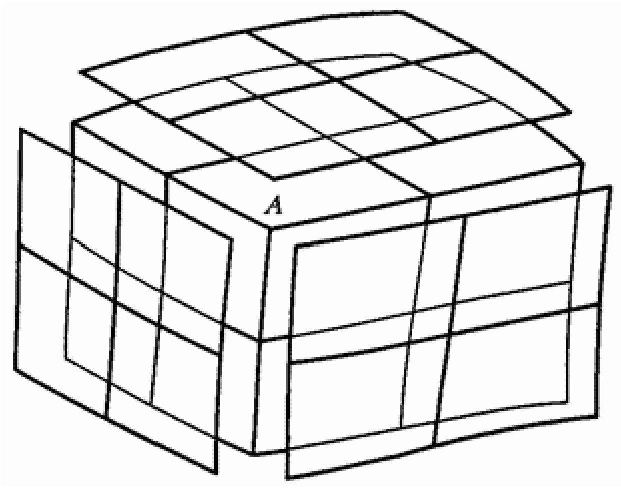


Рис. 8.1. Эквидистантные поверхности

Для построения эквидистантного тела нам остается построить его вершины и ребра. Для этого необходимо найти линии пересечения эквидистантных поверхностей. Каждая линия пересечения должна начинаться и оканчиваться в вершине тела.

На втором этапе построим вершины эквидистантной оболочки. Рассмотрим последовательно вершины базового тела. В каждой вершине стыкуется несколько ребер. Нам нужно знать, какие ребра и какие грани базового тела стыкуются в данной вершине.

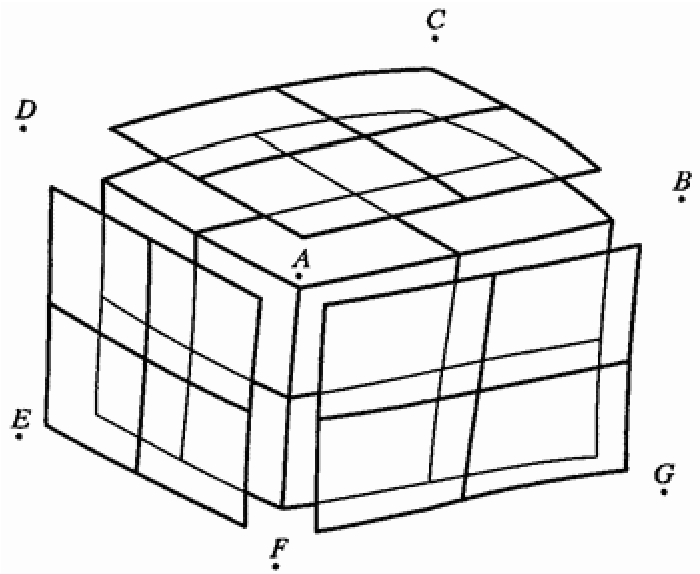


Рис. 8.2. Вершины эквидистантного тела

Вершине базового тела будут соответствовать одна или несколько вершин эквидистантного тела. В каждой вершине эквидистантного тела будут также стыковаться несколько ребер. Вершина эквидистантного тела базируется на точке пересечения эквидистантных поверхностей.

Вычислив эту точку, мы найдем параметры эквидистантных поверхностей, которые будут служить нам в качестве начальных и конечных точек линий пересечения поверхностей.

На рис. 8.2 приведены точки пересечения продолженных эквидистантных поверхностей, на которых будут базироваться вершины.

На третьем этапе построим ребра эквидистантной оболочки. Рассмотрим последовательно ребра базового тела. Для каждого ребра построим соответствующее ребро эквидистантной оболочки тела. Ребро будет базироваться на линии пересечения эквидистантных поверхностей.

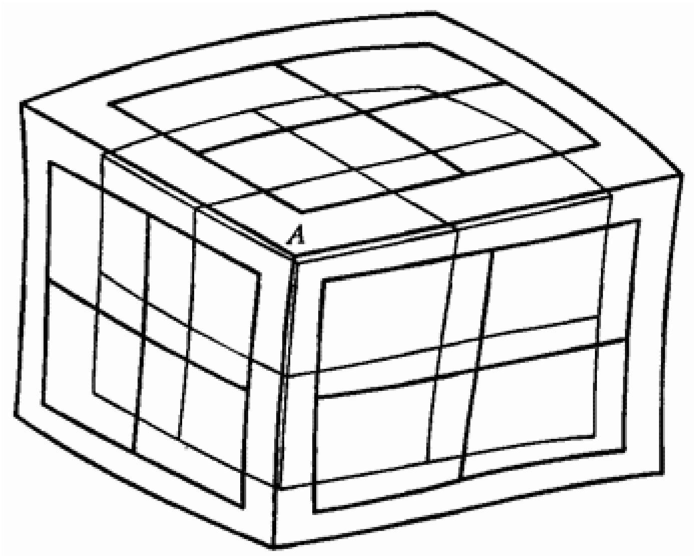


Рис. 8.3. Линии пересечения эквидистантных поверхностей

Начальные и конечные точки ребер нам известны из второго этапа. На рис. 8.3 приведены линии пересечения продолженных эквидистантных поверхностей, на которых будут базироваться ребра.

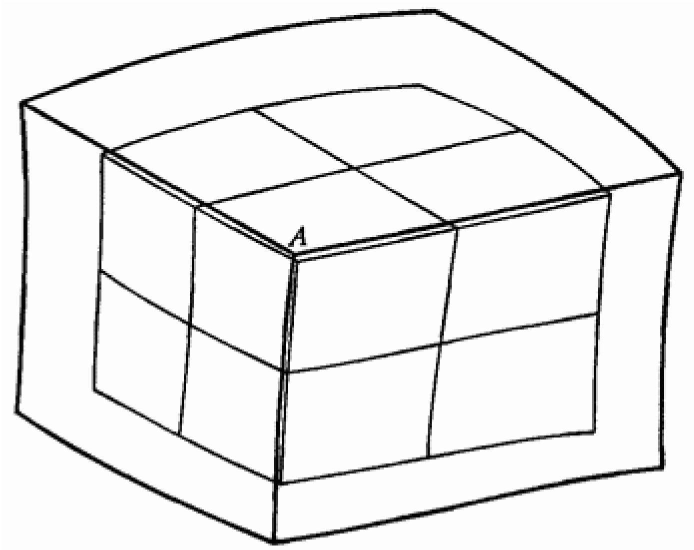


Рис. 8.4. Эквидистантные грани

На последнем, четвертом этапе построим циклы эквидистантных граней (рис. 8.4). Этот процесс аналогичен процессу перестроения циклов граней в булевых операциях.

Таким образом, мы получим оболочку эквидистантного тела.

Пример построения эквидистантного тела приведен на рис.8.5 (исходное тело показано внутри тонкими линиями).

Заметим, что топология эквидистантного тела (количество вершин, ребер, граней и их взаимосвязь) не всегда совпадает с топологией базового тела.

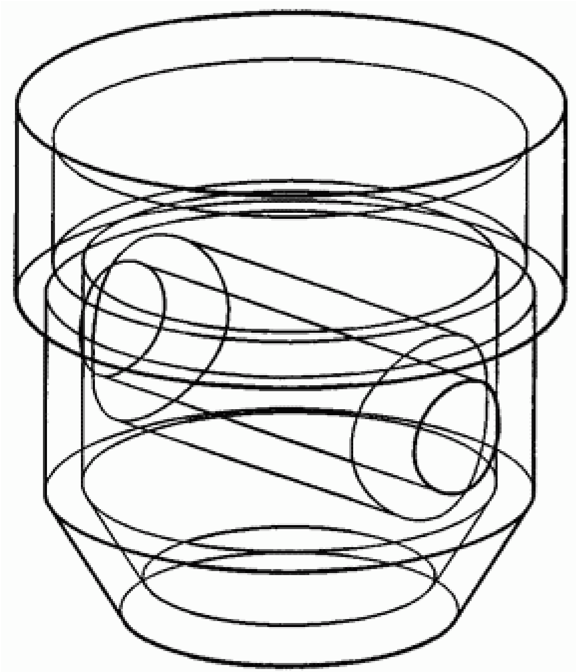


Рис. 8.5. Эквидистантное тело

Если в вершине базового тела стыкуется более трех ребер (не являющихся швами), то b эквидистантном теле этой вершине будет соответствовать несколько вершин и новых ребер. На рис. 8.6 приведена пирамида и эквидистантное пирамиде тело с отрицательным параметром эквидистанты, а на рис. 8.7 приведена аналогичная пирамида и эквидистантное пирамиде тело с положительным параметром эквидистанты.

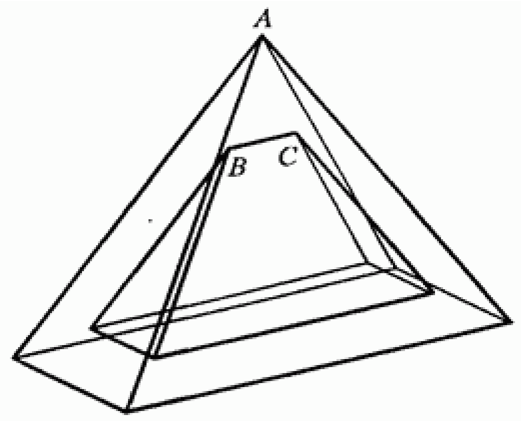


Рис. 8.6. Эквидистантная пирамида (h < 0)

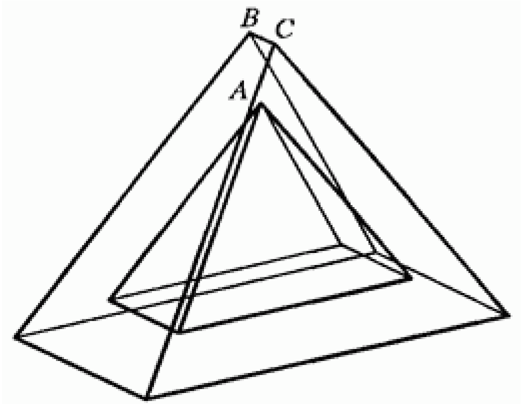


Рис. 8.7. Эквидистантная пирамида (h > 0)

В вершине А пирамиды стыкуется четыре ребра. В эквидистантном теле данной вершине соответствует две вершины В и С и одно дополнительное ребро ВС, что мы и наблюдаем на рис. 8.6 и 8.7.

Возможна и другая ситуация, когда некоторой вершине или некоторому ребру базового тела в эквидистантной оболочке не будет аналога. В таких случаях топология эквидистантного тела будет отличаться от топологии базового тела. Для обработки подобных ситуаций следует проанализировать расположение соседних вершин и ребер.

12.9. Построение тонкостенного тела

Рассмотрим построение тонкостенных тел двух типов. Тела обоих типов будут строиться по некоторому базовому телу. Тело первого типа будет иметь замкнутую полость внутри. Оно имеет две оболочки, делящие пространство на три части: одна из них лежит вне тела, вторую занимает тело, а третья является внутренней полостью тела. Тело второго типа будет представлять собой открытое тонкостенное тело, имеющее одну оболочку. Тело первого типа будем называть закрытым тонкостенным телом, а тело второго типа — открытым тонкостенным телом. Закрытое тонкостенное тело можно считать частным случаем открытого тонкостенного тела.

Закрытое тело.

Закрытое тонкостенное тело с толщиной стенки h получим следующим образом. Построим эквидистантную оболочке базового тела оболочку. Процесс построения эквидистантной оболочки был описан при построении эквидистантного тела. Далее вывернем наизнанку одну из этих оболочек. Если $h > 0$, то вывернем наизнанку оболочку базового тела, если $h < 0$, то вывернем наизнанку эквидистантную оболочку. Эти две оболочки и создадут тонкостенное тело. Закрытое тонкостенное тело можно представить по рис.8.5.

Данное тонкостенное тело является в отличие от других рассмотренных тел телом с пустотами. В общем случае тело с пустотами имеет несколько оболочек. Одна из них является внешней, а остальные — внутренними и лежат внутри внешней оболочки. Все оболочки не должны пересекать друг друга. Вектор нормали к внешней оболочке направлен вне объема оболочки, а векторы нормалей к внутренним оболочкам направлены внутрь объема, ограниченного ими.

Открытое тело.

Открытое тонкостенное тело строится на базе некоторого тела путем удаления одной или нескольких граней последнего и «придания оставшимся граням конечной толщины». Конечно, грань не может иметь толщину, поэтому к оставшейся после удаления некоторых граней открытой оболочке строится эквидистантная открытая оболочка, а затем эти оболочки замыкаются частями удаляемых граней. В результате получается одна замкнутая оболочка.

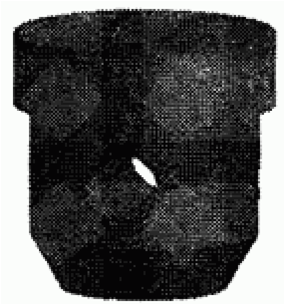


Рис.9.1. Исходное тело

Рис.9.2. Тонкостенное тело

На рис.9.2 приведен пример открытого тонкостенного тела ($h > 0$), построенного путем вскрытия одной грани тела, показанного на рис.9.1.

Рассмотрим процесс построения оболочки открытого тонкостенного тела. Прежде всего, рассортируем грани базового тела на две группы: к первой группе отнесем вскрываемые грани базового тела, а ко второй группе отнесем остальные грани базового тела, которые будем называть сохраняемыми. Сгруппируем ребра базового тела: к первой группе отнесем ребра, по которым пересекаются между собой вскрываемые грани базового тела, а ко второй группе отнесем ребра сохраняемых граней базового тела.

Процесс построения оболочки открытого тонкостенного тела имеет много общего с процессом построения эквидистантного тела.

Для каждой сохраняемой грани базового тела построим эквидистантную грань. Поверхность каждой эквидистантной грани и каждой вскрываемой грани должна иметь возможность быть продолженной до пересечения с поверхностями соседних граней. Радиус-вектор эквидистантной поверхности определяется известной формулой. На продолжении эквидистантной поверхности за пределы области определения параметров ее радиус-вектор будем вычислять по одной из известных формул в зависимости от замкнутости базовой поверхности.

Далее рассмотрим вершины сохраняемых граней базового тела. Каждой рассматриваемой вершине будут соответствовать одна или несколько вершин тонкостенного тела. В каждой вершине тонкостенного тела будут стыковаться несколько ребер. Вершина тонкостенного тела базируется на точке пересечения эквидистантных поверхностей или на точке пересечения эквидистантных поверхностей с поверхностями вскрываемых граней. Вычислив эти точки, мы найдем параметры пересекающихся поверхностей, которые будут служить нам в качестве начальных и конечных точек линий пересечения поверхностей.

Рассмотрим последовательно ребра сохраняемых граней базового тела (ребра второй группы). Для каждого ребра построим соответствующее ребро тонкостенного тела. Для этого найдем линии пересечения эквидистантных поверхностей между собой и с поверхностями вскрываемых граней. Начальные и конечные точки ребер нам известны.

По построенным вершинам перестроим ребра пересечения вскрываемых граней (ребра первой группы).

Далее вывернем наизнанку часть граней.

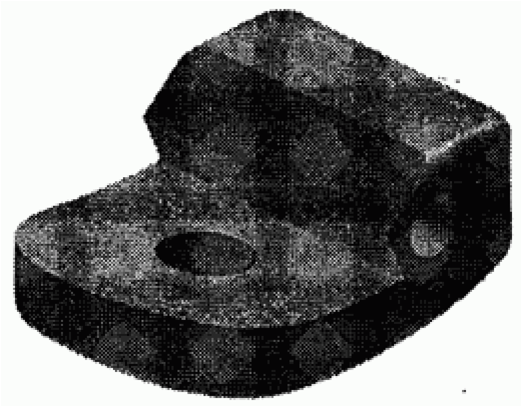


Рис.9.3. Базовое тело

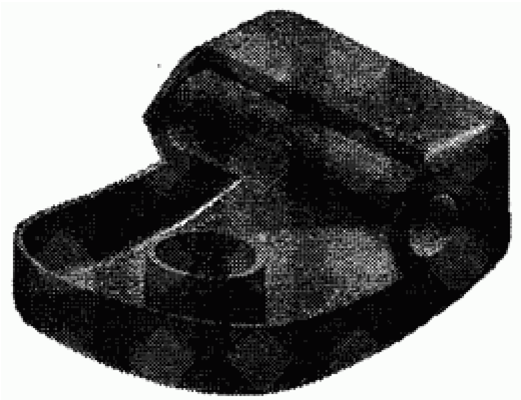


Рис. 9.4. Тонкостенное тело, построенное внутрь от базового тела

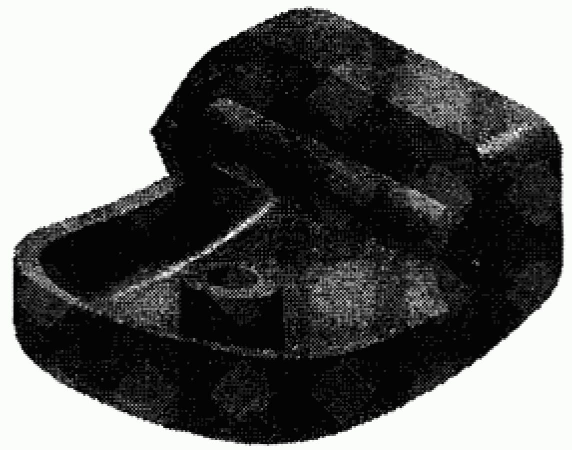


Рис.9.5. Тонкостенное тело, построенное наружу от базового тела

Если h<0, то изменим на противоположные направления циклов и нормалей сохраняемых граней базового тела и направления циклов вскрываемых граней базового тела. Если $h < 0$, то изменим на противоположные направления нормалей эквидистантных граней (циклов они еще не имеют).

На последнем этапе построим циклы эквидистантных граней и перестроим циклы вскрываемых граней. Таким образом, оболочка открытого тонкостенного тела будет составлена из сохраняемых граней, эквидистантных к ним граней и частей вскрываемых граней.

Открытое тонкостенное тело может быть построено как наружу ($h > 0$), так и внутрь ($h < 0$) от базового тела. На рис.9.4 приведено открытое тонкостенное тело, построенное внутрь от базового тела путем вскрытия трех граней. Базовое тело приведено на рис.9.3. На рис.9.5 приведено открытое тонкостенное тело, построенное на основе этого же базового тела наружу от него тела путем вскрытия тех же трех граней (с большей толщиной стенки).

Операция построения тонкостенного тела наряду с булевыми операциями является мощным средством для построения тел сложной формы. На рис.9.6 приведено открытое тонкостенное тело, построенное на базе тела, показанного на рис.9.5, путем вскрытия двух граней.

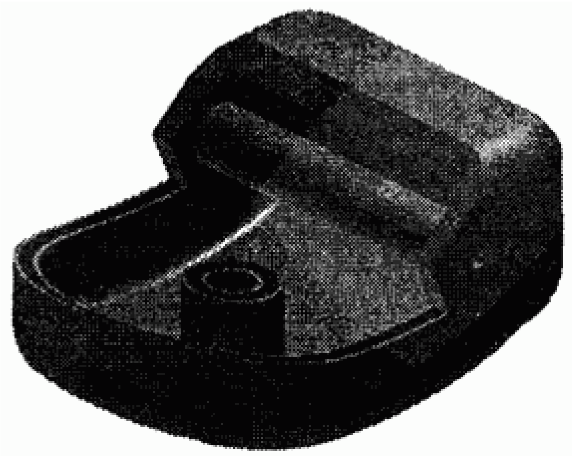


Рис.9.6. Тонкостенное тело на базе тонкостенного тела

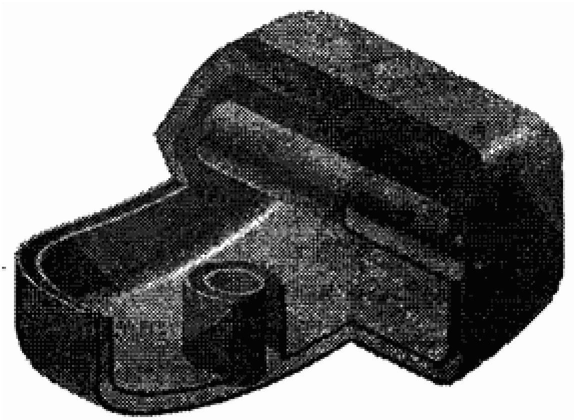


Рис.9.7. Разрез тонкостенного тела

На рис.9.7 приведен разрез двумя плоскостями этого тонкостенного тела, на котором видны внутренние полости тела.

Тонкостенное тело принадлежит к сложным телам. В его дерево построения положим структуру данных исходного тела, толщину стенки и список удаляемых граней.

12.10. Скругление ребер тела

Операция скругления ребер тела позволяет построить плавный переход от одной грани к другой. На рис. 10.1-11.4 приведены примеры скругления ребер тел. На рис. 10.1, 10.2 скруглены выпуклые ребра.

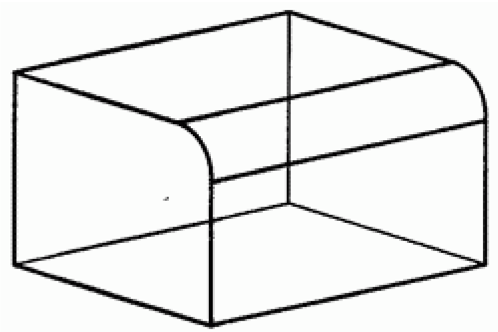


Рис. 10.1. Призма со скругленным ребром

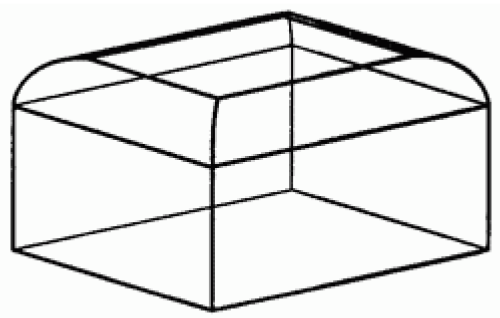


Рис. 10.2. Призма со скругленными ребрами

На рис. 10.3, 10.4 скруглены вогнутые ребра.

Рассмотрим общий случай скругления произвольного криволинейного ребра. Пусть грань, лежащая справа от скругляемого ребра (если смотреть снаружи тела вдоль ребра), базируется на поверхности $\mathbf{r}(u,v)$, а грань, лежащая слева от скругляемого ребра, базируется на поверхности $\mathbf{s}(a,b)$.

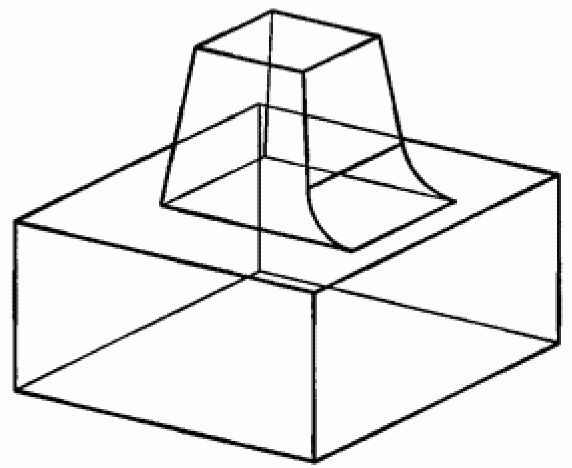


Рис. 10.3. Скругление ребра

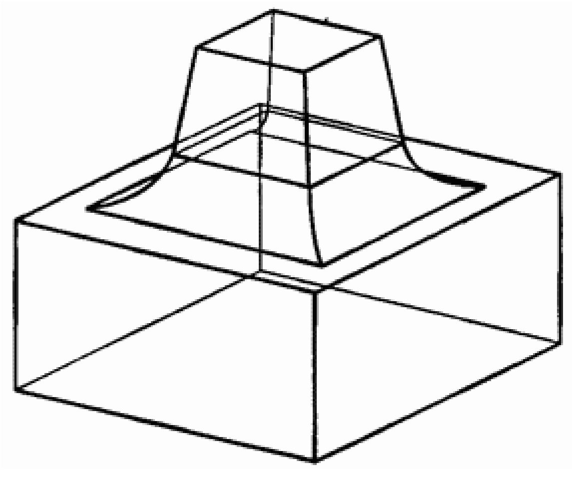


Рис. 10.4. Вариант скругления ребер

Эти грани будем называть сопрягаемыми. Пусть скругляемое ребро базируется на линии пересечения **c**(s) поверхностей **r**(*u*,*v*) и **s**(*a*,*b*). По кривой **c**(s) построим поверхность скругления **q**(*t*,*z*), которая определяется известной формулой. Краями поверхности скругления являются линии, совпадающие с двухмерными линиями ()

$$l_{uv}(t) = \begin{bmatrix} u(t) & v(t) \end{bmatrix}^{\top}, \qquad l_{uv}(t) \in \mathbf{r}(u,\, v),$$
$$l_{ab}(t) = \begin{bmatrix} a(t) & b(t) \end{bmatrix}^{\top}, \qquad l_{ab}(t) \in \mathbf{s}(a,\, b) \qquad (10.1)$$

на сопрягаемых поверхностях. Построим два ребра вдоль краев поверхности скругления на базе линий пересечения

$$l_{uv}(t) = \begin{bmatrix} u(t) & v(t) \end{bmatrix}^{\top}, \qquad l_{uv}(t) \in \mathbf{r}(u,\, v),$$
$$l_{tz}(t) = [t \quad 0]^{\top}, \qquad l_{tz}(t) \in \mathbf{q}(t,\, z),$$
$$l_{ab}(t) = \begin{bmatrix} a(t) & b(t) \end{bmatrix}^{\top}, \qquad l_{ab}(t) \in \mathbf{s}(a,\, b),$$
$$l_{tz}(t) = [t \quad 1]^{\top}, \qquad l_{tz}(t) \in \mathbf{q}(t,\, z).$$

(10.2-10.3)

Эти ребра будем называть продольными, так как они направлены вдоль скругляемого ребра и имеют ту же ориентацию.

Если скругляемое ребро замкнуто, то поверхность скругления также будет замкнутой по параметру t. Тогда на базе линии пересечения

$$l_{tz1}(w) = [t_{\min} \quad w]^{\top}, \qquad l_{tz1}(w) \in \mathbf{q}(t,\, z),$$
$$l_{tz2}(w) = [t_{\max} \quad w]^{\top}, \qquad l_{tz2}(w) \in \mathbf{q}(t,\, z),$$
$$0 \leqslant w \leqslant 1, \qquad (10.4)$$

построим ребро, которое будет являться швом.

Если скругляемое ребро не замкнуто, то найдем все грани (за исключением сопрягаемых), пересекающиеся с поверхностью скругления (большая их часть стыкуется в начальной и конечной вершинах скругляемого ребра), и построим линии их пересечения с поверхностью скругления. На базе этих линий пересечения создадим ребра, которые будем называть поперечными.

На базе поверхности $\mathbf{q}(t,z)$ построим грань скругления. Цикл этой грани будет состоять из продольных и поперечных ребер (или шва). Для грани скругления определим признак совпадения ее нормали с направлением нормали поверхности. Нормаль грани скругления должна быть направлена наружу тела (нормаль поверхности скругления может совпадать с ней или быть ей противоположной) .

После этого перестроим сопрягаемые поверхности $\mathbf{r}(u,v)$ и $\mathbf{s}(a,b)$ и циклы граней на них. Для этого найдем пересечение линий (10.2) и (10.3) с ребрами сопрягаемых граней или их продолжениями и изменим эти ребра. В одних случаях упомянутые ребра нужно обрезать (рис. 10.1), в других случаях их нужно продлить (рис. 10.3). Вместо скругляемого ребра в цикл грани $\mathbf{r}(u,v)$ поставим ребро на базе линии пересечения (10.2), а в цикл грани $\mathbf{s}(a,b)$ поставим ребро на базе линии пересечения (10.3) (рис. 10.5).

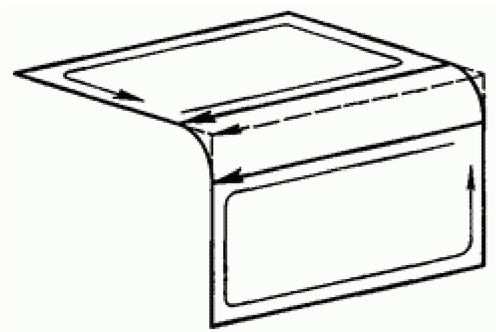


Рис. 10.5. Перестроение циклов граней при скруглении ребра

Далее перестроим циклы остальных граней, пересекшихся с поверхностью скругления $\mathbf{q}(t,z)$. В циклы этих граней войдут поперечные ребра. Перестроение циклов производится проверкой последовательности стыковки ребер между собой и составлением списка ребер цикла в порядке их следования. Поперечные ребра определят область изменения параметра t поверхности скругления $\mathbf{q}(t,z)$.

Скругление сопряженных ребер.

Будем называть стыкующиеся ребра сопряженными., если в точках стыковки они имеют общую касательную. Если в точках стыковки ребра претерпевают излом, то такие ребра будем называть несопряженными.

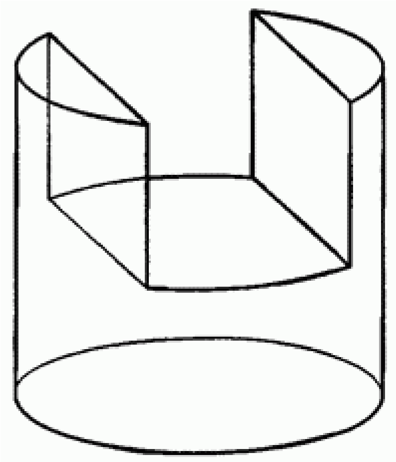


Рис. 10.6. Исходное тело

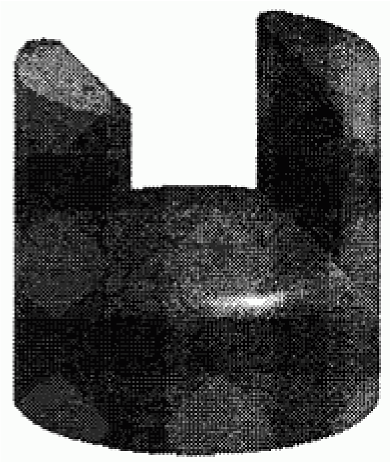


Рис. 10.7. Тело со скруглением нескольких ребер

Если скругляются сразу несколько несопряженных ребер тела, то выполним их скругление последовательно одно за другим. Скругление ребер тел, приведенных на рис. 10.2, 10.4, 10.7, выполнено последовательно.

В случаях скругления нескольких сопряженных ребер различные стадии операции скругления каждого ребра следует выполнять параллельно — одновременно для нескольких ребер. Перед началом операции скругления следует составить группы гладко стыкующихся ребер и далее работать с этими группами как с отдельным ребром описанным выше образом.

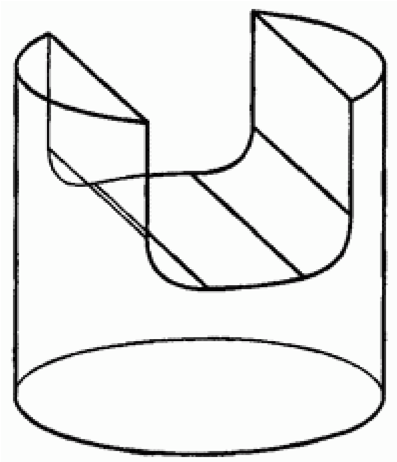


Рис. 10.8. Исходное тело

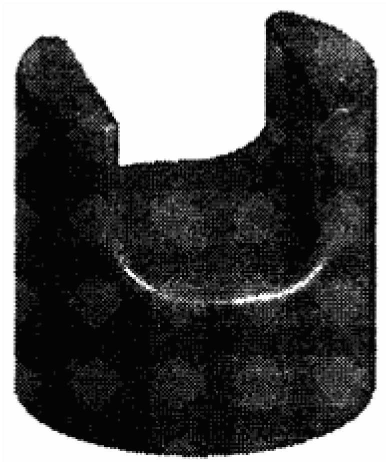


Рис. 10.9. Тело со скруглением цепочек ребер

Случай скругления нескольких сопряженных ребер приведен на рис. 10.9. Исходное тело, показанное на рис. 10.8, получено скруглением двух ребер тела, приведенного на рис. 10.6.

Для группы сопряженных ребер необходимо сначала построить все поверхности скругления, усечь этими поверхностями ребра тела, построить все продольные и поперечные ребра пересечения поверхностей скругления и граней тела, и только после этого построить грани скругления и произвести перестроение циклов граней тела.

На рис. 10.11 приведено тело со скруглением сопряженных ребер, у которых на концах продольные ребра грани скругления сходятся в одну точку (рис. 10.10).

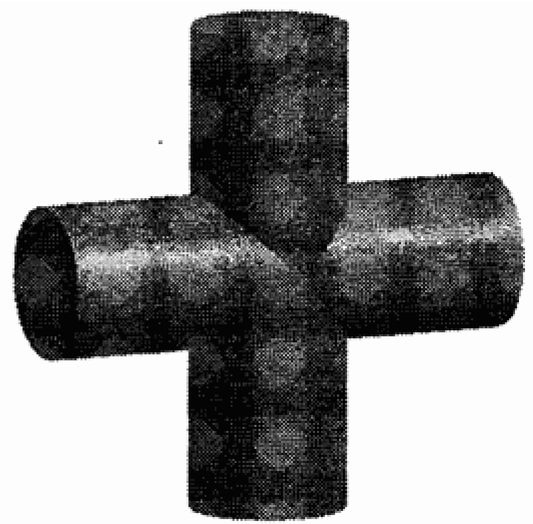


Рис. 10.10. Исходное тело

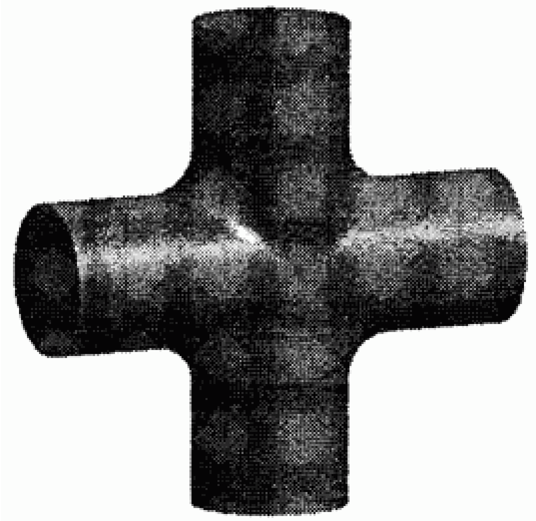


Рис. 10.11. Тело со скруглением цепочек ребер

В данном примере между сопряженными гранями поперечные ребра стянуты в точку, а поверхности скругления в них вырождаются.

Продольные ребра могут не полностью лежать в области сопрягаемых граней. В этом случае участки продольных ребер, выходящие за область определения сопрягаемых граней должны быть заменены на ребра пересечения грани скругления и соседних с сопрягаемой гранью граней.

Скругление вершин.

Если скруглить три ребра, стыкующиеся в одной вершине, то в вершине полупим картину, приведенную на рис. 10.12. Как правило, такую вершину скругляют. Скруглить вершину можно описанным выше способом: ребро пересечения двух граней скругления сопряжено с третьим ребром и может быть скруглено вместе с ним.

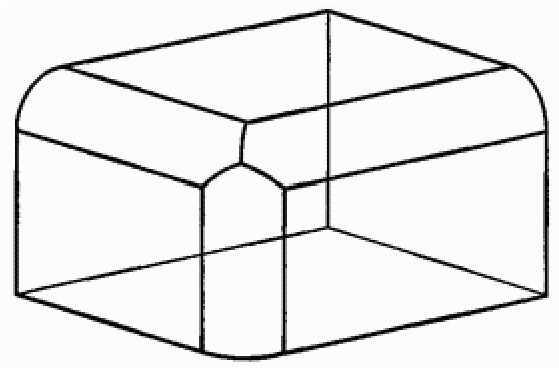


Рис. 10.12. Подлежащая скруглению вершина

Существует еще один способ скругления вершины. Для этого построим по одной линии на каждой поверхности скругления, используя то, что каждая из трех поверхностей скругления пересекается с двумя другими.

Рассмотрим построение упомянутой линии на одной из поверхностей скругления типа $\mathbf{q}(t,z)$ Пусть продольная линия $l_r(t)$ этой поверхности скругления пересекается с продольной линией другой поверхности скругления в точке с параметром t_r, а продольная линия $l_s(t)$ рассматриваемой поверхности скругления пересекается с продольной линией третьей поверхности скругления в точке с параметром t_s. Построим на рассматриваемой поверхности скругления двухмерный отрезок прямой из точки $[t_t\, 0]^T$в точку $[t_s\, 0]^T$:

$$l_{tz}(z) = \left[(1-z)t_r \quad zt_s\right]^T, \quad l_{tz}(z) \in \mathbf{q}(t,\, z), \quad 0 \leqslant z \leqslant 1. \tag{10.5}$$

По отрезку и поверхности построим пространственную линию $l_q(z)$. Линии (10.5) построим на каждой из трех поверхностей скругления, имеющих общую точку пересечения. Отрежем и опустим ту часть каждой грани скругления, которая лежит за построенной линией (10.5) (со стороны общей вершины). По трем линиям (10.5) построим поверхность, а на ее базе построим грань скругления вершины.

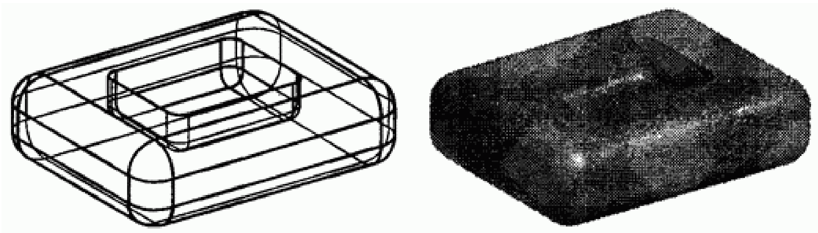


Рис. 10.13. Скругление ребер и вершин тела

На рис. 10.13 приведен пример скругления ребер и вершин призмы описанным способом.

Аналогично можно скруглить вершину, в которой стыкуются четыре скругляемых ребра. Для этого следует использовать поверхность, построенную по четырем кривым.

12.11. Построение фасок ребер тела

Фаски ребер тела строятся аналогично скруглению ребер с той лишь разницей, что поверхности скругления заменяются поверхностями фасок. Фаска описывается линейчатой поверхностью, построенной по двум линиям на пересекающихся по ребру поверхностях.

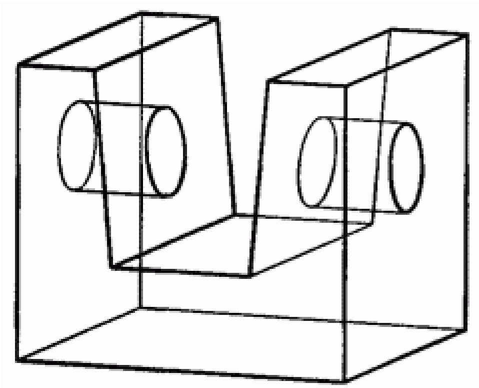


Рис. 11.1. Исходное тело

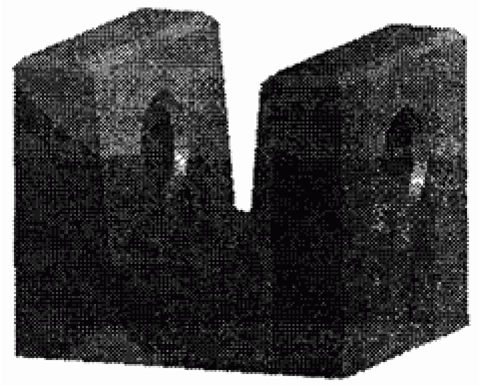


Рис. 11.2. Тело с фасками ребер

Результат построения фасок ребер тела, показанного на рис. 11.1, приведен на рис. 11.2.

Фаска вершин.

Если поверхности фасок строятся для трех ребер, стыкующихся в одной вершине, то общую вершину, как правило, срезают. Срез вершины выполним аналогично скруглению вершины.

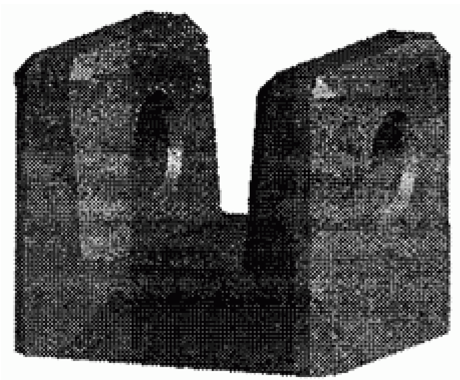


Рис. 11.3. Тело с фасками ребер и вершин

Для этого найдем три линии пересечения поверхностей фасок и построим по ним треугольную поверхность. На базе трех линий пересечения построим три ребра, а на базе треугольной поверхности построим грань. Результат среза вершины, в которой стыкуются три ребра, приведен на рис. 11.3.

12.12. Некоторые способы построения тел

Описанный выше метод выполнения булевых операций применим в случае полной определенности оболочек операндов. На практике часто требуется выполнить булеву операцию, когда оболочка одного из операндов определена не полностью и должна быть достроена в процессе операции. Например, к заданному телу нужно добавить часть тела, полученного выдавливанием заданного плоского контура, лежащую со стороны контура (рис. 12.1).

Результат такой операции приведен на рис. 12.2.

В данном случае требовалось «выдавить» заданный контур до ближайших к нему граней заданного тела. В разных местах контура ближайшими могут оказаться разные грани, поэтому точки контура должны быть «выдавлены» на различные расстояния.

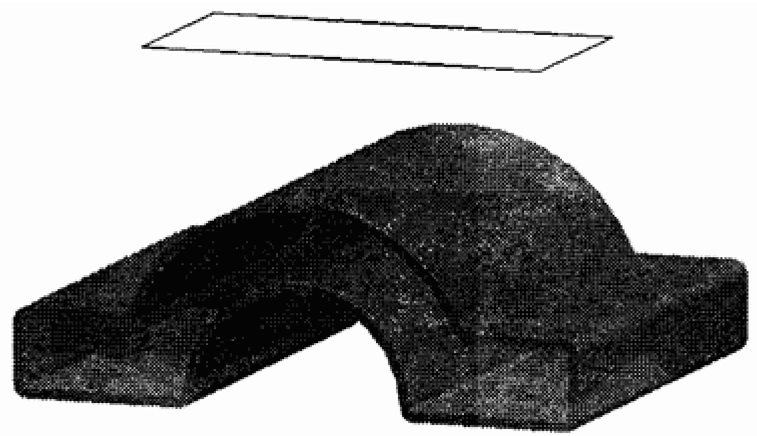


Рис. 12.1. Исходное тело и плоский контур

Выполнить данную операцию можно следующим образом. По заданному контуру построим тело выдавливания достаточной глубины для пересечения с заданным телом. Далее из тела выдавливания вычтем заданное тело. В результате операции мы в общем случае получим несколько оболочек. Выберем из них ближайшую к заданному контуру.

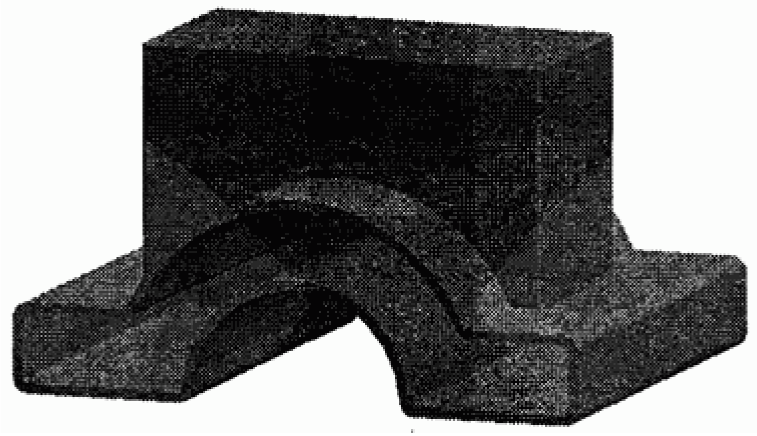


Рис. 12.2. Выдавливание контура до ближайшей поверхности тела

Тело с выбранной оболочкой объединим с заданным телом. При выполнении данных действий достаточно строить ребра пересечения только один раз. Во второй части операции (объединение тела с выбранной оболочкой) следует использовать только те ребра, которые принадлежат телу с выбранной оболочкой.

Другим примером может служить операция «вырезки» ближайших граней заданного тела заданным контуром (рис. 12.3).

В данном примере контуром вырезаются ближайшие стенки тела и не трогаются грани, лежащие за ними. В разных местах контура его точки «выдавливаются» на различные расстояния.

Вырезать ближайшие грани заданного тела заданным контуром можно следующим образом. По заданному контуру построим тело выдавливания достаточной глубины для пересечения с заданным телом.

Далее найдем пересечение тела выдавливания с заданным телом. В результате операции мы в общем случае получим несколько оболочек. Выберем из них ближайшую к заданному контуру. Тело с выбранной оболочкой вычтем из заданного тела. При выполнении данных действий достаточно строить ребра пересечения только один раз.

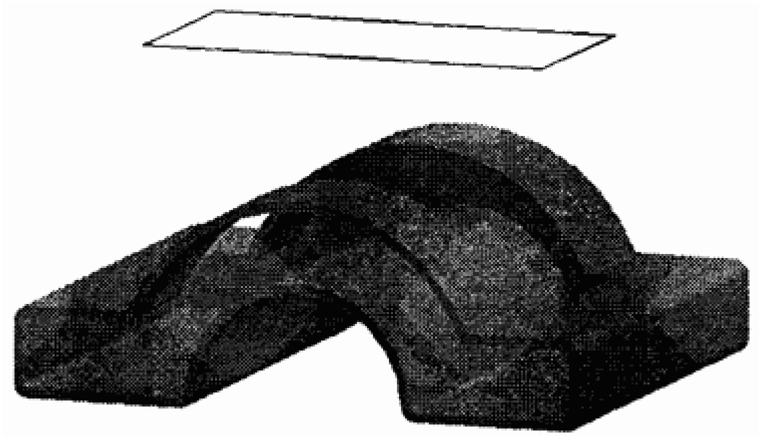


Рис. 12.3. Вырез контуром ближайших поверхностей тела

Во многих конструкциях используются ребра жесткости. Операция построения ребер жесткости выполняется по той же схеме, что и операция «выдавливания» контура до ближайших к нему граней тела. Тело с ребрами жесткости приведено на рис. 12.4.

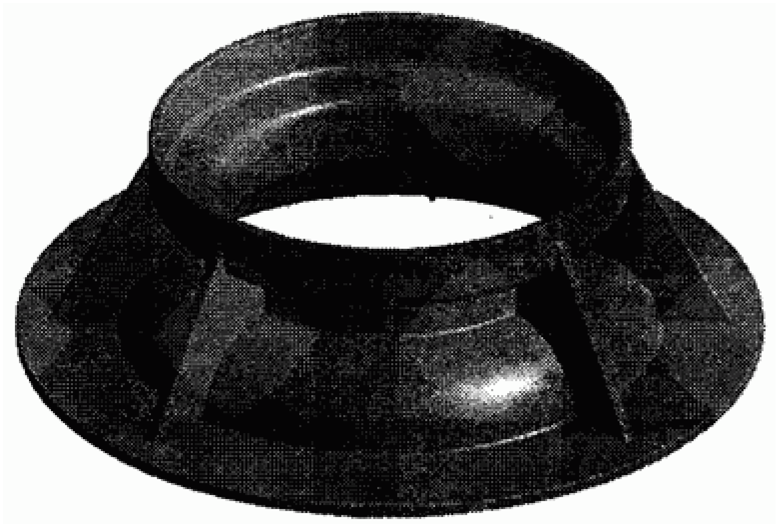


Рис. 12.4. Тело с ребрами жесткости

Когда оболочка одного из тел-операндов должна быть построена в процессе выполнения операции, после выполнения операций следует проверить оболочку результирующего тела на замкнутость.

12.13. Последовательность моделирования тел

Моделирование некоторого объекта может включать построение одного тела или построение нескольких тел — сборки тел.

Создание одиночного тела начинается с построения или одного из простых тел, или тела на базе линий, или тела на базе поверхности. Эти способы построения тел приведены в левой части рис. 13.1. Если исходное тело создается на базе плоских линий, то для их построения используются конструктивные плоскости.

Перед построением тела на базе поверхности нужно сначала создать исходную поверхность. Во многих случаях при построении поверхности также используются конструктивные плоскости. Далее путем выполнения операций, приведенных в правой части рис. 13.1, из исходного тела можно получить тело с более сложной оболочкой.

Операции над телом могут выполняться многократно и в произвольной последовательности, что отражено на рис. 13.1.

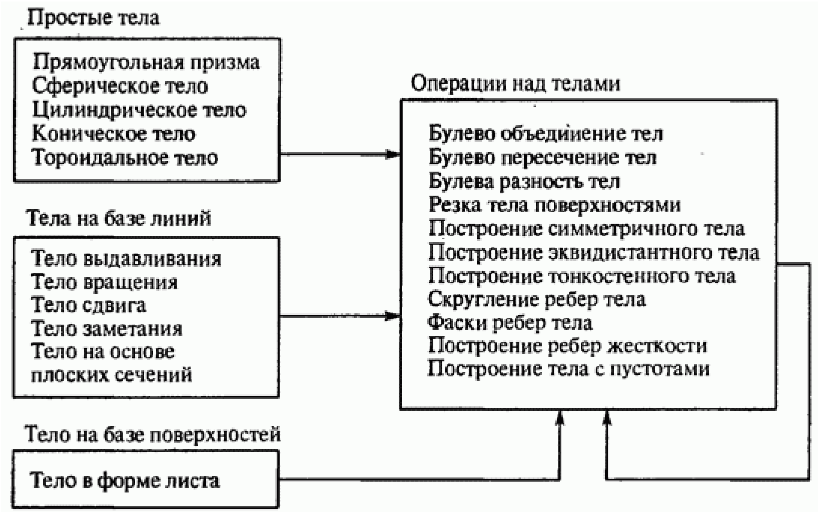


Рис. 13.1. Способы построения тел

Процесс построения оболочки сложного тела близок к процессу изготовления моделируемого объекта. С помощью булевой операции объединения к телу можно добавить требуемый объем. Для этого нужно построить еще одно тело и объединить исходное тело с ним. Аналогично с помощью булевых операций пересечения или вычитания из тела можно убрать требуемый объем. От тела может быть отрезана лишняя часть объема. Ребра тела могут быть скруглены или с них могут быть сняты фаски. Из тела можно получить тонкостенное тело путем «вскрытия» одной или нескольких граней и «придания оставшимся граням конечной толщины». Для симметричных тел можно построить только одну половину тела, а затем получить требуемое тело с помощью операции создания симметричного тела. К телу могут быть добавлены ребра жесткости.

Из нескольких тел можно получить сборку тел. В сборке все тела равноправны. Для взаимного расположения и ориентации тел сборки можно использовать преобразования сдвига, поворота, масштабирования или симметрии.

Построение отдельных тел и сборок должно сопровождаться протоколом построения, который называют деревом построения. Дерево построения позволяет выполнять редактирование тел и их сборок, создавать наборы однотипных моделей и управлять ими.

Наряду с деревом построения для редактирования кривых линий, поверхностей, тел, сборок тел и управления ими может использоваться механизм вариационных связей. Этому механизму посвящена следующая глава.

13. ВАРИАЦИОННЫЕ СВЯЗИ ГЕОМЕТРИЧЕСКИХ ОБЪЕКТОВ

13.1. Наложение вариационных связей

В данной главе мы рассмотрим один из эффективных способов управления геометрическими объектами. Каждый геометрический объект имеет свою структуру данных. **Структура данных вместе с набором необходимых объекту функций представляет собой численную модель геометрического объекта**. **Скалярные величины, компоненты векторов, координаты точек,** лежащие в структуре данных геометрического объекта, которые подлежат редактированию, будем называть **параметрами этого объекта**. Именно через эти параметры мы и будем осуществлять управление геометрическими объектами.

До сих пор геометрические объекты строились и существовали независимо друг от друга, т. е. редактирование одного из объектов не сказывалось на остальных объектах. Редактирование объекта сводится к изменению численных значений его параметров. Независимость геометрических объектов отражает тот факт, что значения параметров одного объекта не зависят от значений параметров других объектов. С практической точки зрения наложение зависимостей на параметры геометрических объектов является очень полезным.

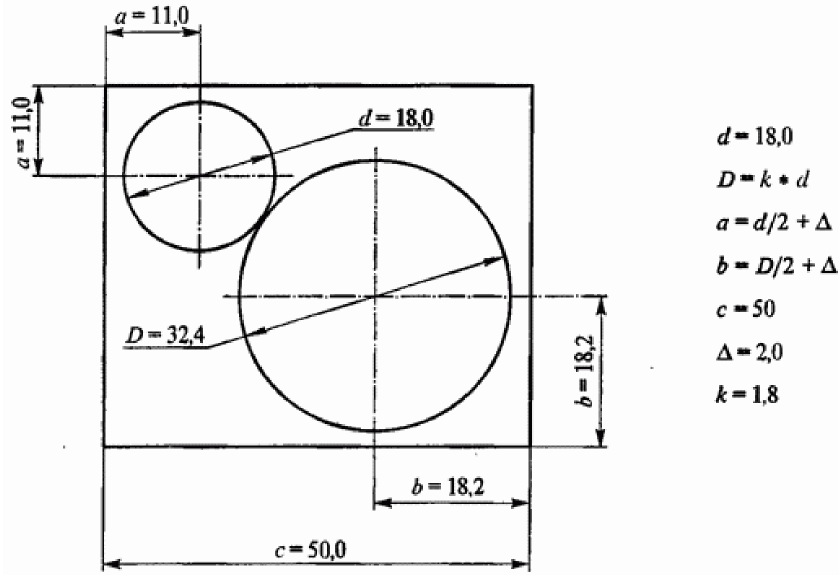


Рис. 1.1. Вариационные связи геометрических объектов

На рис. 1.1 приведен пример наложения зависимостей на две окружности и замкнутую ломаную линию на плоскости.

Зависимость геометрических объектов заключается в том, что окружности должны касаться друг друга, ломаная должна представлять собой прямоугольник, диаметры d и D окружностей связаны заданным коэффициентом к, центры окружностей должны отстоять на заданных расстояниях a и b от сторон прямоугольника, размер горизонтальной стороны прямоугольника должен быть равен c. Перечисленные связи представлены размерами и алгебраическими уравнениями, приведенными на рис. 1.1. **Эскиз автоматически перестраивается при изменении одной или нескольких зависимостей**. На рис. 1.2 приведены те же геометрические объекты при другом отношении диаметров окружностей.

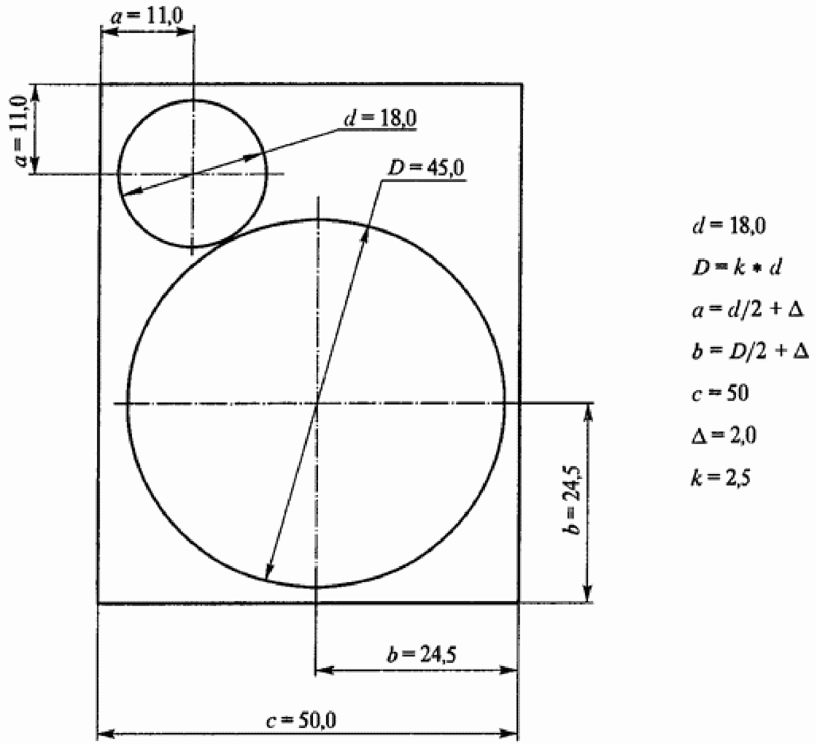


Рис. 1.2. Результат изменения одной из связей геометрических объектов

Изменение коэффициента отношения диаметров к привело *k* изменению размера вертикальной стороны прямоугольника, изменению диаметра и положения центра большей окружности. Все заданные уравнения при новом коэффициенте *k* также выполняются. Аналогично могут быть изменены и другие размеры и уравнения, что приведет к автоматическому перестроению геометрических объектов.

Наложение связей облегчает труд при проектировании нескольких однотипных деталей и при сборке различных деталей. Достигается это путем установления определенных зависимостей между параметрами геометрических объектов. Эти зависимости представляют собой некоторые уравнения

$$\begin{array}{l} f_1(x_1, x_2, \ldots, x_n) = 0, \\ f_2(x_1, x_2, \ldots, x_n) = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ f_k(x_1, x_2, \ldots, x_n) = 0, \end{array}$$

относительно параметров $x_1, x_2, \ldots, x_n$ объектов.

Будем называть связывающие параметры уравнения **вариационными связями.** Кроме координат точек, компонент векторов, скалярных величин из структуры данных геометрических объектов параметрами могут служить дополнительные скалярные величины и координаты точек. Они могут существовать отдельно или могут быть введены в структуру данных геометрических объектов для удобства управления объектами посредством вариационных связей. Для наложения вариационных связей нужно, чтобы геометрические объекты предоставляли свои параметры «во внешнее пользование» и перестраивались по измененным параметрам. Вариационными связями в общем случае могут служить любые алгебраические уравнения. Естественно, что каждой геометрической задаче соответствует определенное уравнение (или несколько уравнений) для определенных параметров.

Управление геометрическими объектами с помощью вариационных связей осуществляется следующим образом. Пусть задан набор вариационных связей. Каждая связь накладывает одно или несколько уравнений на определенные параметры геометрических объектов. Связями формируется система уравнений относительно участвующих в связях параметров. Пусть в исходном состоянии параметры удовлетворяют этой системе уравнений. Предположим, что мы изменили некоторую константу в уравнении связи (например, изменили значение размера) или изменили некоторые параметры геометрического объекта (например, переместили объект) и зафиксировали их. Чтобы после этого система уравнений связей удовлетворялась, необходимо изменить остальные параметры. Для этого мы должны решить систему уравнений связей и определить новые значения параметров. О результатах решения следует сообщить геометрическим объектам. Они должны перестроиться в соответствии с этим решением (в соответствии с новыми значениями своих данных). Так как уравнения в общем случае являются нелинейными, то их решение осуществляется итерационным методом. Объекты должны перестраиваться на каждой итерации решения, так как одни параметры объекта могут влиять на другие параметры этого же геометрического объекта. Уравнения будут изменять значения параметров, а объекты — перестраиваться в соответствии с новыми значениями параметров. Когда все уравнения системы будут удовлетворены с требуемой точностью, объекты будут перестроены соответствующим образом. Если систему уравнений удовлетворить нельзя, то всем параметрам присваиваются первоначальные значения. На практике часто приходится иметь дело с системой уравнений, содержащей большее число параметров, чем число уравнений. В последнем случае мы воспользуемся некоторым критерием, определяющим поведение всей системы параметров, и с помощью этого критерия сформируем систему уравнений для определения всех параметров.

Дополним вариационные связи информацией связываемых ими геометрических объектах, а также функциями общения с объектами и системой уравнений. В результате вариационные связи можно будет называть вариационными объектами. Они несут и обрабатывают геометрическую информацию. Как и геометрические объекты, вариационные объекты имеют свою структуру данных и свой набор функций. Состав структуры данных и функций определяется выполняемыми связями задачами. Каждый вариационный объект отвечает за то, чтобы заданные параметры удовлетворяли заданным уравнениям. В структуре данных вариационного объекта должны находиться: информация о связываемых им варьируемых параметрах, исходные значения параметров, информация об уравнениях связи (одном или нескольких).

В набор функций вариационных объектов должны войти функции, предоставляющие и изменяющие необходимую информацию о варьируемых параметрах, функции, предоставляющие информацию об уравнениях связей, функции, изменяющие параметры в процессе решения и после удовлетворения уравнений всех связей, функции восстановления исходного состояния параметров в случае неудачи в процессе решения.

Сначала мы рассмотрим отдельные вариационные связи и их уравнения. Далее введем критерий поведения геометрических объектов, позволяющий сформировать систему уравнений, в которой варьируемые параметры обладают равноправием при любом их числе.

Этот критерий в общем случае требует привлечение методов **вариационного исчисления**. На примере использования критерия поведения геометрических объектов мы рассмотрим вариационные связи двухмерных объектов.

13.2. Фиксирующие связи

Простейшей связью является фиксирующая связь. Она описывается одним уравнением, содержащим один параметр. Если параметр обозначить через q, то уравнение имеет вид

$$q = q^{(0)} = \mathrm{const}, \quad (2.1)$$

где $q(0)$— заданное значение параметра, которое должно быть сохранено в дальнейшем. Фиксирующая связь используется тогда, когда требуется, чтобы некоторый параметр не изменялся в процессе решения системы уравнений связей. Если некоторый параметр неизменен, то он может быть удален из списка переменных системы уравнений связей. Иногда удобнее сохранить неизменный параметр, но при этом следует ввести для него фиксирующее уравнение (2.1).

Примером фиксирующей связи может служить вариационная зависимость, называемая **закреплением точки**. Она содержит три уравнения типа (2.1) и фиксирует три координаты точки $\mathbf{p}=[x\ y\ z]^T$ Закрепление точки описывается одним векторным уравнением

$$\mathbf{p} = \mathbf{p}^{(0)}, \quad (2.2)$$

состоящим из трех скалярных уравнений

$$x = x^{(0)}, \quad y = y^{(0)}, \quad z = z^{(0)}, \quad (2.3)$$

где $\mathbf{p}^{(0)}=[x^{(0)}\ y^{(0)}\ z^{(0)}]^T$— заданное положение точки. Координаты закрепленных точек остаются неизменными, поэтому частные производные всех других уравнений системы по закрепленным параметрам равны нулю.

Уравнение (2.1) может использоваться для любого параметра геометрического объекта — скалярной величины, компоненты вектора или координаты точки.

Фиксирующие связи занимают особое место среди вариационных связей. Количество параметров в фиксирующих связях всегда равно количеству уравнений, сами параметры остаются неизменными, поэтому фиксирующие связи могут быть выделены из общей системы связей в самостоятельную группу. Система уравнений фиксирующих связей и система уравнений остальных связей не зависят друг от друга и могут быть решены отдельно.

13.3. Вариационные связи точек в пространстве

Наиболее простыми являются вариационные связи, накладываемые на координаты радиус-векторов точек. Эти связи могут быть наложены на две или несколько точек. При этом число уравнений связей может быть меньше числа участвующих в связях параметров. В этом случае можно составить дополнительные уравнения, необходимые для решения задачи. Мы рассмотрим пример, построения вариационной связи, фиксирующей расстояние между двумя точками в пространстве. В этом примере мы составим дополнительные уравнения исходя из симметрии поведения геометрических объектов. Дополнительные уравнения нам потребуются, чтобы проиллюстрировать процесс решения системы уравнений связи.

Линейный размер.

Пусть имеются две точки $\mathbf{p}_1=[x_1\ y_1\ \ z_1]^T$ и $\mathbf{p}_2=[x_2\ y_2\ \ z_2]^T$. Наложим на точки вариационную связь, называемую линейным размером. Она описывается уравнением

$$|\mathbf{p}_2 - \mathbf{p}_1| - d = 0, \quad (3.1)$$

где d — требуемый размер. В координатном представлении уравнение связи имеет вид

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} - d = 0.$$

Линейный размер приведен на рис. 3.1. Точка **p**$_1$ лежит в структуре данных тора и определяет положение центра его местной системы координат. Точка **p**$_2$ лежит в структуре данных сферы и определяет положение ее центра.

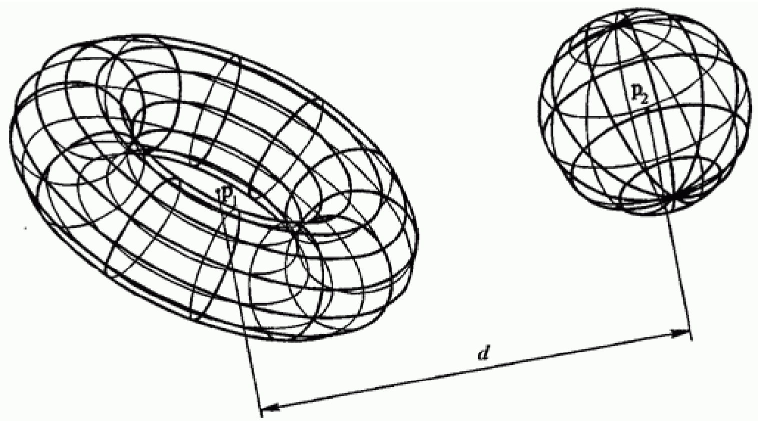


Рис. 3.1. Линейный размер между точками центров тора и сферы

В общем случае это могут быть любые точки из структуры данных геометрических объектов. Мы имеем одно уравнение (3.1), связывающее в общем случае шесть параметров: $x_1, y_1, z_1, x_2, y_2, z_2$. В уравнении (3.1) величину d будем считать константой. В общем случае величина d также может быть варьируемым параметром.

Рассмотрим принцип решения поставленной задачи. В данном случае для того, чтобы найти все параметры, необходимы шесть уравнений. Существует несколько способов получить систему уравнений с числом варьируемых параметров (неизвестных) равным числу уравнений. В данном случае мы составим некоторые дополнительные уравнения, связывающие те же параметры. Набор дополнительных уравнений может быть различным, но от этих дополнительных уравнений будет зависеть поведение точек. Потребуем, например, чтобы при изменении размера точки перемещались симметрично, т. е. точки перемещались бы вдоль соединяющей их линии, а центр тяжести точек оставался бы неподвижным. Эти условия описываются уравнениями

$$(x_2 - x_1)(y_2{}^{(0)} - y_1{}^{(0)}) = (y_2 - y_1)(x_2{}^{(0)} - x_1{}^{(0)}),$$

$$(y_2 - y_1)(z_2{}^{(0)} - z_1{}^{(0)}) = (z_2 - z_1)(y_2{}^{(0)} - y_1{}^{(0)}),$$

$$x_2 + x_1 = x_2{}^{(0)} + x_1{}^{(0)},$$

$$y_2 + y_1 = y_2{}^{(0)} + y_1{}^{(0)},$$

$$z_2 + z_1 = z_2{}^{(0)} + z_1{}^{(0)},$$

(3.2-3.6)

где $\mathbf{p_1}^{(0)}=[x_1{}^{(0)}\ y_1{}^{(0)}\ z_1{}^{(0)}]^{\mathrm{T}}$ и $\mathbf{p_2}^{(0)}=[x_2{}^{(0)}\ y_2{}^{(0)}\ z_2{}^{(0)}]^{\mathrm{T}}$ — исходное положение точек, которое в общем случае не удовлетворяет уравнению (3.1). Добавив (3.2)-(3.6) к (3.1), получим системы шести уравнений относительно шести параметров. Перепишем все уравнения, придав им вид $f_i(x_1, y_1, z_1, x_2, y_2, z_2)=0$, $i=1,2,\ldots,6$. Первое уравнение является нелинейным.

Решим систему уравнений методом Ньютона. Организуем итерационный процесс

$$x_i{}^{(r+1)} = x_i{}^{(r)} + \Delta x_i,$$

$$y_i{}^{(r+1)} = y_i{}^{(r)} + \Delta y_i,$$

$$z_i{}^{(r+1)} = z_i{}^{(r)} + \Delta z_i,$$

$$i = 1,\ 2, \qquad (3.7)$$

где r— номер итерации. Приращения координат Δx_i Δy_i Δz_i, $i=1,2$ на каждой итерации определим из системы линейных алгебраических уравнений

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & 0 & a_{24} & a_{25} & 0 \\ 0 & a_{32} & a_{33} & 0 & a_{35} & a_{36} \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \Delta x_1 \\ \Delta y_1 \\ \Delta z_1 \\ \Delta x_2 \\ \Delta y_2 \\ \Delta z_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix} \qquad (3.8)$$

где

$$a_{11} = -a_{14} = \frac{x_1 - x_2}{d^{(r)}}, \quad a_{12} = -a_{15} = \frac{y_1 - y_2}{d^{(r)}}, \quad a_{13} = -a_{16} = \frac{z_1 - z_2}{d^{(r)}},$$
$$a_{21} = -a_{24} = {y_1}^{(0)} - {y_2}^{(0)}, \quad a_{22} = -a_{25} = {x_2}^{(0)} - {x_1}^{(0)},$$
$$a_{32} = -a_{35} = {z_1}^{(0)} - {z_2}^{(0)}, \quad a_{33} = -a_{36} = {y_2}^{(0)} - {y_1}^{(0)},$$
$$b_1 = d - d^{(r)},$$
$$b_2 = ({y_2}^{(r)} - {y_1}^{(r)})({x_2}^{(0)} - {x_1}^{(0)}) - ({x_2}^{(r)} - {x_1}^{(r)})({y_2}^{(0)} - {y_1}^{(0)}),$$
$$b_3 = ({z_2}^{(r)} - {z_1}^{(r)})({y_2}^{(0)} - {y_1}^{(0)}) - ({y_2}^{(r)} - {y_1}^{(r)})({z_2}^{(0)} - {z_1}^{(0)}),$$
$$b_4 = {x_2}^{(0)} + {x_1}^{(0)} - {x_2}^{(r)} - {x_1}^{(r)},$$
$$b_5 = {y_2}^{(0)} + {y_1}^{(0)} - {y_2}^{(r)} - {y_1}^{(r)},$$
$$b_6 = {z_2}^{(0)} + {z_1}^{(0)} - {z_2}^{(r)} - {z_1}^{(r)},$$
$$d^{(r)} = \sqrt{({x_2}^{(r)} - {x_1}^{(r)})^2 + ({y_2}^{(r)} - {y_1}^{(r)})^2 + ({z_2}^{(r)} - {z_1}^{(r)})^2}.$$

Система уравнений (3.8) в матричной записи имеет вид

$$\mathbf{A} \cdot \Delta\mathbf{x} = \mathbf{b}. \qquad (3.9)$$

Эта система имеет единственное решение, если определитель $|\mathbf{A}| \neq 0$.На каждой итерации матрица А вычисляется заново, так как точки перемещаются в пространстве, и может случиться так, что определитель матрицы **A** на очередной итерации окажется равным нулю. Тогда точки следует вернуть в исходное положение, а размер положить равным расстоянию между ними. Каждая строка матрицы **A** соответствует определенному уравнению. Для *i*-го уравнения $f_i(x_1, y_1, z_1, x_2, y_2, z_2)=0$ в системе (3.8) коэффициенты *i*-й строки матрицы **A** определяются по формулам

$$a_{i1} = \frac{\partial f_i}{\partial x_1}, \quad a_{i2} = \frac{\partial f_i}{\partial y_1}, \quad a_{i3} = \frac{\partial f_i}{\partial z_1}, \quad a_{i4} = \frac{\partial f_i}{\partial x_2}, \quad a_{i5} = \frac{\partial f_i}{\partial y_2}, \quad a_{i6} = \frac{\partial f_i}{\partial z_2},$$

Итерационный процесс (3.7) закончим, когда приращения координат на очередной итерации станут меньше заданной величины.

Возможны и другие варианты дополнительных уравнений. Например, если требуется сохранить точку $\mathbf{p}_1$ в заданном положении, то в качестве дополнительных уравнений можно использовать уравнения (3.2), (3.3) и

$$\mathbf{p}_1 = \mathbf{p}^{(0)},$$

где $\mathbf{p}^{(0)}$— заданное положение точки. В дальнейшем для решения системы уравнений связей мы не будем использовать дополнительные уравнения, так как они нарушают равноправие варьируемых параметров.

Размер вдоль координаты.

Более простой, чем линейный размер, вариационной связью двух точек является размер вдоль одной координаты.

Она связывает две соответствующие координаты двух точек и описывается одним из уравнений

$$x_2 - x_1 = d_x,$$
$$y_2 - y_1 = d_y,$$
$$z_2 - z_1 = d_z, \tag{3.10}$$

где d_x, d_y, d_z — требуемые размеры. Уравнения (3.10), примененные к координатам точек отрезка $\mathbf{r}(t)=(1-t)\mathbf{p}_1+t\mathbf{p}_2$ могут сделать отрезок параллельным соответствующей координатной плоскости. Два уравнения (3.10) могут сделать отрезок параллельным одной из координатных осей.

Размер вдоль направления.

Еще одной вариационной связью двух точек является размер вдоль заданного направления. Она описывается уравнением

$$|(\mathbf{p}_2 - \mathbf{p}_1) \cdot \mathbf{m}| = d, \quad (3.11)$$

где $\mathbf{m}=[m_x\ m_y\ m_z\ ]^T$— задающий направление вектор единичной длины, d — требуемый размер. Уравнение (7.3.11) устанавливает, что проекция вектора $\mathbf{p}_2 - \mathbf{p}_1$ построенного между заданными точками, на направление вектора m по абсолютной величине равна d. В координатной записи уравнение (3.11) имеет вид

$$|(x_2 - x_1)m_x + (y_2 - y_1)m_y + (z_2 - z_1)m_z| = d. \quad (3.12)$$

Уравнение (3.12) связывает шесть параметров. Если $\mathbf{m}=[m_x\ m_y\ m_z\ ]^T$, $\mathbf{1}=[l_x\ l_y\ l_z\ ]$, $\mathbf{n}=[n_x\ n_y\ n_z\ ]^T$ — три линейно независимых вектора единичной длины, то можно задать относительное положение двух точек $\mathbf{p}_2$ и $\mathbf{p}_1$ сразу тремя размерами вдоль трех некомпланарных направлений **m,1,n**:

$$\begin{aligned} |(\mathbf{p}_2 - \mathbf{p}_1) \cdot \mathbf{m}| &= d_m, \\ |(\mathbf{p}_2 - \mathbf{p}_1) \cdot \mathbf{l}| &= d_l, \\ |(\mathbf{p}_2 - \mathbf{p}_1) \cdot \mathbf{n}| &= d_n. \end{aligned} \quad (3.13)$$

В координатной записи уравнения (3.13) имеют вид

$$|(x_2 - x_1)m_x + (y_2 - y_1)m_y + (z_2 - z_1)m_z| = d_m,$$
$$|(x_2 - x_1)l_x + (y_2 - y_1)l_y + (z_2 - z_1)l_z| = d_l,$$
$$|(x_2 - x_1)n_x + (y_2 - y_1)n_y + (z_2 - z_1)n_z| = d_n.$$

Совмещение точек.

Можно установить вариационную связь, которая сливает две точки в одну. Эта связь совмещения точек. Она приравнивает координаты двух заданных точек и описывается векторным уравнением

$$\mathbf{p}_2 = \mathbf{p}_1. \qquad (3.14)$$

Данная вариационная связь содержит три скалярных уравнения

$$x_2 - x_1 = 0,$$
$$y_2 - y_1 = 0,$$
$$z_2 - z_1 = 0. \qquad (3.15)$$

Симметрия точек относительно плоскости.

Рассмотрим вариационную связь симметрии точек $\mathbf{p}_2$ и $\mathbf{p}_1$ относительно плоскости

$$\mathbf{r}(x, y) = \mathbf{p} + x\mathbf{i}_1 + y\mathbf{i}_2. \qquad (3.18)$$

Плоскость будем считать неподвижной, поэтому ее параметры варьироваться не будут.

Симметрия точек относительно плоскости описывается тремя уравнениями

$$(\mathbf{p}_1 - \mathbf{p}) \cdot \mathbf{i}_1 = (\mathbf{p}_2 - \mathbf{p}) \cdot \mathbf{i}_1,$$

$$(\mathbf{p}_1 - \mathbf{p}) \cdot \mathbf{i}_2 = (\mathbf{p}_2 - \mathbf{p}) \cdot \mathbf{i}_2,$$

$$(\mathbf{p}_1 - \mathbf{p}) \cdot \mathbf{i}_3 + (\mathbf{p}_2 - \mathbf{p}) \cdot \mathbf{i}_3 = 0, \quad (3.19\text{-}3.21)$$

где $\mathbf{i}_3=\mathbf{i}_1 \times \mathbf{i}_2$. Три уравнения (3.19)-(3.21) связывают шесть параметров. Симметрия точек показана на рис. 3.2.

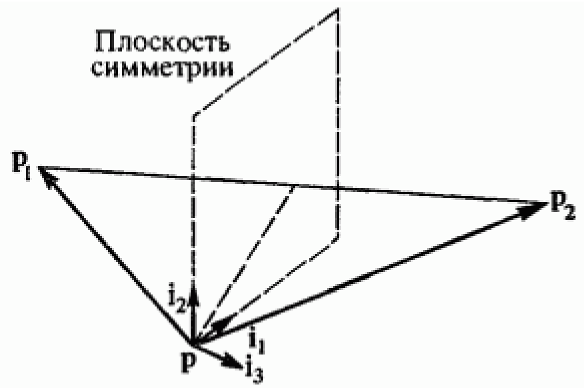


Рис. 3.2. Симметрия точек относительно плоскости

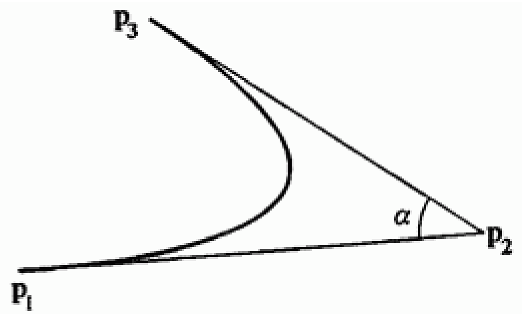


Рис. 3.3. Угловой размер между тремя точками NURBS кривой

Угловой размер.

Для трех точек можно установить угловую зависимость. Пусть имеется три точки $\mathbf{p_1}=[x_1\ y_1\ z_1\ ]^T$, $\mathbf{p_2}=[x_2\ y_2\ z_2\ ]^T$, $\mathbf{p_3}=[x_3\ y_3\ z_3\ ]^T$. На три точки можно наложить вариационную связь, определяющую угловой размер между отрезками $\mathbf{p_1}-\mathbf{p_2}$, $\mathbf{p_3}-\mathbf{p_2}$. На рис. 3.3 точками, на которые наложена вариационная связь, являются характеристические точки рациональной кривой. В общем случае это могут быть любые точки из структуры данных геометрических объектов.

Уравнение, описывающее угловой размер между тремя точками, имеет вид

$$\operatorname{arctg}\frac{s}{c}=\alpha, \qquad (3.22)$$

где

$$s=|(\mathbf{p}_1-\mathbf{p}_2)\times(\mathbf{p}_3-\mathbf{p}_2)|,$$
$$c=(\mathbf{p}_1-\mathbf{p}_2)\cdot(\mathbf{p}_3-\mathbf{p}_2)=$$
$$=(x_1-x_2)(x_3-x_2)+(y_1-y_2)(y_3-y_2)+(z_1-z_2)(z_3-z_2), \qquad (3.23\text{-}3.24)$$

α — заданный угол.

Величина s равна произведению длин векторов $\mathbf{p_1}-\mathbf{p_2}$ и $\mathbf{p_3}-\mathbf{p_2}$ на $\sin\alpha$. Величина c равна произведению длин векторов $\mathbf{p_1}-\mathbf{p_2}$ и $\mathbf{p_3}-\mathbf{p_2}$ на $\cos\alpha$. Квадрат s определяется формулой

$$s^2=\left((y_1-y_2)(z_3-z_2)-(z_1-z_2)(y_3-y_2)\right)^2+$$
$$+\left((z_1-z_2)(x_3-x_2)-(x_1-x_2)(z_3-z_2)\right)^2+$$
$$+\left((x_1-x_2)(y_3-y_2)-(y_1-y_2)(x_3-x_2)\right)^2. \qquad (3.25)$$

Уравнение (3.22) связывает девять параметров.

Другой вариационной связью трех точек $\mathbf{p}_1$, $\mathbf{p}_2$ и $\mathbf{p}_3$ является угловой размер в плоскости. Эта связь устанавливает угловой размер между проекциями векторов $\mathbf{p}_1 - \mathbf{p}_2$ и $\mathbf{p}_3 - \mathbf{p}_2$ на плоскость, ортогональную единичному вектору **m**, и описывается уравнением вида (3.22), где

$$s = \pm|\mathbf{q}_1 \times \mathbf{q}_3|,$$

$$c = \mathbf{q}_1 \cdot \mathbf{q}_3,$$

$$\mathbf{q}_1 = \mathbf{p}_1 - \mathbf{p}_2 - \mathbf{m}\big((\mathbf{p}_1 - \mathbf{p}_2)\cdot\mathbf{m}\big),$$

$$\mathbf{q}_2 = \mathbf{p}_3 - \mathbf{p}_2 - \mathbf{m}\big((\mathbf{p}_3 - \mathbf{p}_2)\cdot\mathbf{m}\big). \quad (3.26\text{-}3.27)$$

Векторы $\mathbf{q}_1$ и $\mathbf{q}_3$ являются составляющими векторов $\mathbf{p}_1 - \mathbf{p}_2$ и $\mathbf{p}_3 - \mathbf{p}_2$ ортогональными единичному вектору **m**. Знак плюс в (3.26) выбирается в случае совпадения направления векторов $\mathbf{q}_1 \times \mathbf{q}_3$ и **m**, знак минус — в противном случае.

Угол между векторами.

Пусть имеется четыре точки $\mathbf{p}_1=[x_1\ y_1\ z_1]^T$, $\mathbf{p}_2=[x_2\ y_2\ z_2]^T$, $\mathbf{p}_3=[x_3\ y_3\ z_3]^T$, $\mathbf{p}_4=[x_4\ y_4\ z_4]^T$. По четырем точкам можно построить два вектора. Построим векторы на точках $\mathbf{p}_1$, $\mathbf{p}_2$ и $\mathbf{p}_3$, $\mathbf{p}_4$. Будем считать, что эти векторы не коллинеарны. Два вектора определяют семейство плоскостей, параллельных одновременно им обоим. Каждая плоскость семейства будет ортогональна вектору $\mathbf{n}=(\mathbf{p}_1-\mathbf{p}_2)\times(\mathbf{p}_3-\mathbf{p}_4)$.

На четыре точки можно наложить вариационную связь, определяющую угол между векторами $\mathbf{p}_1 - \mathbf{p}_2$, $\mathbf{p}_3 - \mathbf{p}_4$. Эта связь аналогична угловому размеру (3.22). Уравнение, описывающее угол между четырьмя точками, имеет тот же вид, что и уравнение (3.22):

$$\operatorname{arctg}\frac{s}{c} = \alpha, \quad (3.28)$$

где

$$s = |(\mathbf{p}_1 - \mathbf{p}_2) \times (\mathbf{p}_3 - \mathbf{p}_4)|,$$

$$c = (\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{p}_3 - \mathbf{p}_4) = \\ = (x_1 - x_2)(x_3 - x_4) + (y_1 - y_2)(y_3 - y_4) + (z_1 - z_2)(z_3 - z_4), \qquad (3.29\text{-}3.30)$$

α— заданный угол. Величина s равна произведению длин векторов $\mathbf{p}_1 - \mathbf{p}_2$ и $\mathbf{p}_3 - \mathbf{p}_2$ на sinα. Величина c равна произведению длин векторов $\mathbf{p}_1 - \mathbf{p}_2$ и $\mathbf{p}_3 - \mathbf{p}_2$ на cosα.

Квадрат s определяется формулой

$$s^2 = \left((y_1 - y_2)(z_3 - z_4) - (z_1 - z_2)(y_3 - y_4)\right)^2 + \\ + \left((z_1 - z_2)(x_3 - x_4) - (x_1 - x_2)(z_3 - z_4)\right)^2 + \\ + \left((x_1 - x_2)(y_3 - y_4) - (y_1 - y_2)(x_3 - x_4)\right)^2.$$

Уравнение (3.28) связывает двенадцать параметров.

Потребуем, чтобы при изменении угла между векторами точки перемещались ортогонально вектору **n**. Для нахождения точек скрещения построим две линии

$$\mathbf{r}_1(t_1) = (1 - t_1)\mathbf{p}_1 + t_1\mathbf{p}_2, \\ \mathbf{r}_3(t_3) = (1 - t_3)\mathbf{p}_3 + t_3\mathbf{p}_4. \qquad (3.31\text{-}3.32)$$

Точками скрещения пространственных линий являются точки $\mathbf{r}_1$ и $\mathbf{r}_3$ на линиях $\mathbf{r}_1(t_1)$ и $\mathbf{r}_3(t_3)$ для которых выполняются равенства

$$(\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{r}_1 - \mathbf{r}_3) = 0, \\ (\mathbf{p}_3 - \mathbf{p}_4) \cdot (\mathbf{r}_1 - \mathbf{r}_3) = 0. \qquad (3.33)$$

Вектор, построенный по точкам скрещения, ортогонален обеим прямым. Параметры t_1 и t_3 прямых, соответствующие точкам скрещения, определим из системы уравнений

$$(\mathbf{p}_1 - \mathbf{p}_2) \cdot \big((1-t_1)\mathbf{p}_1 + t_1\mathbf{p}_2 - (1-t_3)\mathbf{p}_3 - t_3\mathbf{p}_4\big) = 0,$$
$$(\mathbf{p}_3 - \mathbf{p}_4) \cdot \big((1-t_1)\mathbf{p}_1 + t_1\mathbf{p}_2 - (1-t_3)\mathbf{p}_3 - t_3\mathbf{p}_4\big) = 0 \qquad (3.35)$$

или

$$(\mathbf{p}_1 - \mathbf{p}_2)\cdot(\mathbf{p}_2 - \mathbf{p}_1)t_1 + (\mathbf{p}_1 - \mathbf{p}_2)\cdot(\mathbf{p}_3 - \mathbf{p}_4)t_3 = (\mathbf{p}_1 - \mathbf{p}_2)\cdot(\mathbf{p}_3 - \mathbf{p}_1),$$
$$(\mathbf{p}_3 - \mathbf{p}_4)\cdot(\mathbf{p}_2 - \mathbf{p}_1)t_1 + (\mathbf{p}_3 - \mathbf{p}_4)\cdot(\mathbf{p}_3 - \mathbf{p}_4)t_3 = (\mathbf{p}_3 - \mathbf{p}_4)\cdot(\mathbf{p}_3 - \mathbf{p}_1).$$

По найденным параметрам найдем точки скрещения.

Ортогональность векторов.

Частным случаем вариационной связи (3.28) является ортогональность векторов, построенных по точкам $\mathbf{p}_1$, $\mathbf{p}_2$ и $\mathbf{p}_3$, $\mathbf{p}_4$. Данная вариационная связь определяется уравнением

$$(\mathbf{p}_1 - \mathbf{p}_2)\cdot(\mathbf{p}_3 - \mathbf{p}_4) = 0. \qquad (3.37)$$

Если на точках построены отрезки прямых (3.31) и (3.32), то уравнение (3.37) делает эти отрезки ортогональными.

Параллельность векторов.

Вариационная связь, делающая векторы $\mathbf{p}_1 - \mathbf{p}_2$ и $\mathbf{p}_3 - \mathbf{p}_4$ параллельными, определяется векторным уравнением

$$(\mathbf{p}_1 - \mathbf{p}_2)\times(\mathbf{p}_3 - \mathbf{p}_4) = 0. \qquad (3.38)$$

Связь (3.38) будем называть параллельностью векторов. Уравнение (3.38) содержит три скалярных уравнения

$$(y_1 - y_2)(z_3 - z_4) = 0, \qquad (z_1 - z_2)(x_3 - x_4) = 0, \qquad (x_1 - x_2)(y_3 - y_4) = 0.$$

13.4. Вариационные связи точек на кривых и поверхностях

Выше мы рассмотрели связи, накладываемые на существующие в структурах данных геометрических объектов параметры. Если требуется наложить вариационные связи на некоторые значения, которые отсутствуют в структуре данных геометрических объектов, то можно ввести в структуру данных дополнительные параметры. Например, для описания дуги окружности в структуре данных не нужны крайние ее точки, но их можно туда ввести, если необходимо накладывать на них вариационные связи. При этом остальные параметры должны пересчитываться в зависимости от положения крайних точек.

Вариационные связи можно накладывать на точки кривых и поверхностей, которые отсутствуют в их структурах данных. Например, мы хотим задать линейный размер между определенной точкой кривой $\mathbf{a}(t)$ и определенной точкой кривой $\mathbf{b}(w)$. Пусть для простоты кривыми являются отрезки

$$\mathbf{a}(t) = (1-t)\mathbf{a}_1 + t\mathbf{a}_2,$$
$$\mathbf{b}(w) = (1-w)\mathbf{b}_1 + w\mathbf{b}_2,$$

а в качестве точек на них мы возьмем середины отрезков. Середины отрезков определяются параметрами $t=t_0=\frac{1}{2}$, $w=w_0=\frac{1}{2}$ Будем считать, что t_0 и w_0 есть постоянные величины, если не будет объявлено, что они также являются варьируемыми параметрами уравнения. Линейный размер описывается уравнением (3.1), в котором теперь $\mathbf{p}_1=\mathbf{a}(t_0)$, $\mathbf{p}_2=\mathbf{b}(w_0)$. Координаты связываемых точек $\mathbf{p}_1=[x_1\ y_1\ z_1]^T$, $\mathbf{p}_2=[x_2\ y_2\ z_2]^T$ являются функциями координат точек кривых $\mathbf{a}_1=[x_{a1}\ y_{a1}\ z_{a1}]^T$, $\mathbf{a}_2=[x_{a2}\ y_{a2}\ z_{a2}]^T$, $\mathbf{b}_1=[x_{b1}\ y_{b1}\ z_{b1}]^T$, $\mathbf{b}_2=[x_{b2}\ y_{b2}\ z_{b2}]^T$. Число варьируемых параметров в данном случае будет равно 12. Ими являются координаты $x_{a1}\ y_{a1}\ z_{a1}$, $x_{a2}\ y_{a2}\ z_{a2}$, $x_{b1}\ y_{b1}\ z_{b1}$, $x_{b2}\ y_{b2}\ z_{b2}$. В общем случае варьируемыми параметрами будут являться параметры из структур данных связываемых кривых, которыми описываются точки $\mathbf{p}_1=\mathbf{a}(t_0)$ и $\mathbf{p}_2=\mathbf{b}(w_0)$.

Аналогично можно установить размер по одной из координат (3.10) или размер вдоль заданного направления (3.12) между определенной точкой кривой **a**(t) и определенной точкой кривой **b**(w). Можно связать определенные точки кривых вариационной зависимостью (3.14). Можно установить угловой размер (3.28) между заданными точками трех геометрических объектов.

Все сказанное остается в силе, если вместо точек кривых связать определенные точки поверхностей. Например, описанным образом можно установить линейный размер между определенной точкой поверхности **a**(u,v) и определенной точкой поверхности **b**(t,w) или установить угловой размер между определенными точками поверхностей **a**(u,v), **b**(t,w), **s**(x,y).

Между точкой **p** и произвольным геометрическим объектом **S** можно установить вариационную связь, «усаживающую» точку на геометрический объект. Эта связь описывается уравнением

$$\mathbf{p} = \mathbf{s}_p, \tag{4.1}$$

где s_p— проекция точки на геометрический объект. Векторное уравнение (4.1) содержит три скалярных уравнения относительно трех параметров — координат точки **p**. Геометрический объект в процессе решения уравнений связей можно считать или неподвижным, или подвижным, но единым целым.

Если требуется, чтобы геометрический объект после установления вариационных связей не изменял свою форму, то его параметры нужно считать неизменными или к уравнениям связи следует добавить уравнения, сохраняющие неизменными параметры, определяющие его геометрию.

13.5. Алгебраические связи

В уравнения связей (3.1), (3.12), (3.22), (3.28) входят величины d и α, которые до сих пор читались константами. На рис. 1.1 размеры связаны друг с другом алгебраическими уравнениями. Для того, чтобы величины d и α, могли изменяться, они должны также являться варьируемыми параметрами. Эти параметры не принадлежат ни какому-либо геометрическому объекту, ни какой-либо вариационной

связи. Будем называть их свободными параметрами.
Алгебраические уравнения, связывающие свободные параметры, будем называть алгебраическими связями. В частном случае алгебраические связи используются, чтобы сделать равными несколько параметров различных геометрических объектов. Для этого вводится свободный параметр и совокупность уравнений, приравнивающих требуемые параметры геометрических объектов введенному свободному параметру.

В качестве примера использования алгебраических связей рассмотрим два линейных размера, один из которых связывает точки $\mathbf{p}_1=[x_1\ y_1\ z_1\ ]^T$, $\mathbf{p}_2=[x_2\ y_2\ z_2\ ]^T$, а другой связывает точки $\mathbf{p}_3=[x_3\ y_3\ z_3\ ]^T$, $\mathbf{p}_4=[x_4\ y_4\ z_4\ ]^T$. Пусть линейные размеры описываются уравнениями

$$|\mathbf{p}_2-\mathbf{p}_1|=a,$$
$$|\mathbf{p}_3-\mathbf{p}_4|=b. \quad (5.1\text{-}5.2)$$

Пусть значения размеров a и b связаны некоторыми алгебраическими уравнениями

$$g(a,\ b)=0,$$
$$h(a,\ b)=0. \quad (5.3\text{-}5.4)$$

Уравнения (5.1) и (5.2) отличаются от уравнения (3.1) тем, что в последнем размер d не варьируется, а в уравнениях (5.1) и (5.2) параметры a и b являются варьируемыми. На (r+1)-й итерации решения системы уравнений изменяются как координаты точек, так и значения размеров:

$$x_i^{(r+1)}=x_i^{(r)}+\Delta x_i,\quad y_i^{(r+1)}=y_i^{(r)}+\Delta y_i,\quad z_i^{(r+1)}=z_i^{(r)}+\Delta z_i,$$
$$a^{(r+1)}=a^{(r)}+\Delta a,\quad b^{(r+1)}=b^{(r)}+\Delta b,$$
$$i=1,\ 2,\ 3,\ 4.$$

Свободными могут быть любые параметры уравнений связей, не являющиеся параметрами геометрических объектов. В общем случае алгебраические уравнения нельзя выделить из общей системы уравнений связей и решить отдельно, так как они могут быть связаны друг с другом через уравнения для параметров геометрических объектов.

Наряду с алгебраическими уравнениями вариационные связи могут строиться на неравенствах. Например, вместо уравнения (5.3) алгебраическая связь может определяться неравенством

$$a > b. \quad (5.5)$$

На каждой итерации решения системы уравнений неравенство заменяется соответствующим уравнением. Неравенство (5.5) на каждой итерации должно быть заменено уравнением

$$a - b = \varepsilon, \quad (5.6)$$

где

$$\begin{aligned} &\varepsilon = a^{(r)} - b^{(r)}, && \text{если} \quad a^{(r)} > b^{(r)}, \\ &\varepsilon = 0, && \text{если} \quad a^{(r)} \leqslant b^{(r)}. \end{aligned}$$

Если на очередной итерации неравенство выполняется, то оно заменяется тождественным равенством, если неравенство не выполняется, то оно заменяется строгим равенством. Аналогичным образом в процессе численного решения системы уравнений и другие неравенства заменяются соответствующими равенствами. Связи, построенные на неравенствах, будем называть нестрогими алгебраическими связями.

13.6. Минимизация изменения параметров

Идеальным является случай, когда число уравнений связей равно числу варьируемых параметров. Этот идеальный случай реализуется, когда на заданную группу параметров наложено максимально возможное число связей. Если делается попытка наложить связей больше, чем это возможно, то система уравнений будет переопределена и отвергнет лишние связи. Если число уравнений связей меньше числа параметров, то необходимо найти способ дополнить систему связей уравнениями. Один из способов — составление дополнительных уравнений.

На примере линейного размера между двумя точками показано использование дополнительных уравнений. Возможны ситуации, когда одна и та же группа параметров участвует в нескольких связях (например, на одну и ту же группу точек наложено несколько размерных зависимостей). Набор дополнительных уравнений может быть различным и существенным образом влияет на поведение геометрических объектов. Использование дополнительных уравнений нарушает равноправие варьируемых параметров и может привести к нежелательным последствиям в поведении геометрических объектов.

Мы откажемся от использования дополнительных уравнений. Необходимую систему уравнений для удовлетворения всех вариационных связей мы получим, наложив некоторое общее для всей совокупности геометрических объектов условие поведения. Поведение системы параметров мы будем определять некоторым критерием, описываемым функцией или функционалом. Уравнения для определения параметров мы получим из требования минимума или максимума функции или функционала данного критерия.

При наложении вариационных связей будем использовать метод минимизации суммы квадратов изменений параметров. Рассмотрим его применение на примере линейного размера между двумя точками, описываемого уравнением (3.1). Если ввести дополнительные уравнения (3.2)-(3.6), то поведение связанных точек будет симметричным, но сама система уравнений имеет несимметричный вид. Кроме того, если рассматриваемые точки участвуют еще в каких-нибудь связях, то возникает вопрос, какие из дополнительных уравнений следует оставить, а какие опустить. В большинстве случаев поведение такой системы связей теряет симметрию.

Обратим внимание на то, что в варианте дополнительных уравнений (3.2)-(3.6) суммарное перемещение связанных линейным размером точек является минимальным из возможных. Используем это свойство для составления системы уравнений. Квадрат суммарного перемещения связанных точек описывается функцией

$$\psi = \frac{1}{2}\left(|\mathbf{p}_1 - \mathbf{p}_1{}^{(0)}|^2 + |\mathbf{p}_2 - \mathbf{p}_2{}^{(0)}|^2\right).$$

В координатном представлении эта функция имеет вид

$$\psi = \frac{1}{2}\Big((x_1 - x_1{}^{(0)})^2 + (y_1 - y_1{}^{(0)})^2 + (z_1 - z_1{}^{(0)})^2 + \\ +(x_2 - x_2{}^{(0)})^2 + (y_2 - y_2{}^{(0)})^2 + (z_2 - z_2{}^{(0)})^2\Big). \tag{6.1}$$

Эта функция пропорциональна сумме квадратов изменений всех связанных параметров. Аргументами этой функции являются те же шесть параметров: $x_1, y_1, z_1, x_2, y_2, z_2$. Потребуем, чтобы сумма квадратов изменений параметров была минимальной. При этом необходимо, чтобы координаты точек удовлетворяли уравнению связи (3.1). Используем метод неопределенных множителей Лагранжа для отыскания минимума функции (6.1) при условии (3.1). Необходимым условием минимума функции (6.1) при условии (3.1) является равенство нулю частных производных по параметрам функции

$$F = \psi + \lambda f = \frac{1}{2}\left(|\mathbf{p}_1 - \mathbf{p}_1{}^{(0)}|^2 + |\mathbf{p}_2 - \mathbf{p}_2{}^{(0)}|^2\right) + \lambda\big(|\mathbf{p}_2 - \mathbf{p}_1| - d\big) = \\ = \frac{1}{2}\Big((x_1 - x_1{}^{(0)})^2 + (y_1 - y_1{}^{(0)})^2 + (z_1 - z_1{}^{(0)})^2 + (x_2 - x_2{}^{(0)})^2 + (y_2 - y_2{}^{(0)})^2 + \\ +(z_2 - z_2{}^{(0)})^2\Big) + \lambda\left(\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} - d\right), \quad (7.6.2 \tag{6.2}$$

где $f=|\mathbf{p}_2 - \mathbf{p}_1|-d$, λ— подлежащий определению множитель. Искомые координаты точек и множитель λ найдем из системы уравнений

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} - d = 0,$$

$$x_1 - x_1^{(0)} + \lambda \frac{x_1 - x_2}{|\mathbf{p}_1 - \mathbf{p}_2|} = 0,$$

$$y_1 - y_1^{(0)} + \lambda \frac{y_1 - y_2}{|\mathbf{p}_1 - \mathbf{p}_2|} = 0,$$

$$z_1 - z_1^{(0)} + \lambda \frac{z_1 - z_2}{|\mathbf{p}_1 - \mathbf{p}_2|} = 0,$$

$$x_2 - x_2^{(0)} + \lambda \frac{x_2 - x_1}{|\mathbf{p}_1 - \mathbf{p}_2|} = 0,$$

$$y_2 - y_2^{(0)} + \lambda \frac{y_2 - y_1}{|\mathbf{p}_1 - \mathbf{p}_2|} = 0,$$

$$z_2 - z_2^{(0)} + \lambda \frac{z_2 - z_1}{|\mathbf{p}_1 - \mathbf{p}_2|} = 0.$$

(6.3)

Данная система уравнений имеет решение

$$x_1 = x_1{}^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(x_1{}^{(0)} - x_2{}^{(0)}),$$
$$y_1 = y_1{}^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(y_1{}^{(0)} - y_2{}^{(0)}),$$
$$z_1 = z_1{}^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(z_1{}^{(0)} - z_2{}^{(0)}),$$
$$x_2 = x_2{}^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(x_2{}^{(0)} - x_1{}^{(0)}),$$
$$y_2 = y_2{}^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(y_2{}^{(0)} - y_1{}^{(0)}),$$
$$z_2 = z_2{}^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(z_2{}^{(0)} - z_1{}^{(0)}),$$
$$\lambda = -\frac{d - d^{(0)}}{2}, \tag{6.4}$$

где

$$d^{(0)} = \sqrt{(x_2^{(0)} - x_1{}^{(0)})^2 + (y_2^{(0)} - y_1{}^{(0)})^2 + (z_2^{(0)} - z_1{}^{(0)})^2}.$$

Можно доказать, что найденное решение является точкой минимума функции (6.2). Точки переместятся вдоль прямой, проходящей через точки $\mathbf{p}_1{}^{(0)}$ и $\mathbf{p}_2{}^{(0)}$ на одинаковое расстояние. В данной постановке задачи точки являются равноправными, а уравнения и решение системы уравнений являются симметричными.

Мы считали, что обе точки могут перемещаться. Если одна из точек закреплена, то на нее наложена связь (2.2) и ее координаты не варьируются. Пусть точка $\mathbf{p}_1$ закреплена. Тогда функция (6.2) будет иметь вид

$$F = \frac{1}{2}|\mathbf{p}_2 - \mathbf{p}_2{}^{(0)}|^2 + \lambda(|\mathbf{p}_2 - \mathbf{p}_1| - d) =$$
$$= \frac{1}{2}\left((x_2 - x_2{}^{(0)})^2 + (y_2 - y_2{}^{(0)})^2 + (z_2 - z_2{}^{(0)})^2\right) +$$
$$+\lambda\left(\sqrt{(x_2 - x_1{}^{(0)})^2 + (y_2 - y_1{}^{(0)})^2 + (z_2 - z_1{}^{(0)})^2} - d\right). \tag{6.5}$$

Координаты закрепленной точки в функцию (6.5) не входят. Функция (6.5) достигает минимума при

$$x_2 = x_2{}^{(0)} + \frac{d - d^{(0)}}{d^{(0)}}(x_2{}^{(0)} - x_1{}^{(0)}),$$
$$y_2 = y_2{}^{(0)} + \frac{d - d^{(0)}}{d^{(0)}}(y_2{}^{(0)} - y_1{}^{(0)}),$$
$$z_2 = z_2{}^{(0)} + \frac{d - d^{(0)}}{d^{(0)}}(z_2{}^{(0)} - z_1{}^{(0)}),$$
$$\lambda = -(d - d^{(0)}). \tag{6.6}$$

И в решении (6.4), и в решении (6.6) свободные точки перемещаются вдоль прямой, проходящей через точки $\mathbf{p}_1$ и $\mathbf{p}_2$ Требование минимума суммы квадратов варьируемых параметров автоматически создает симметрию в поведении системы. Симметрия присутствует и в системе уравнений. Это свойство особенно ценно для сложных связей, в которых много параметров и все они имеют различный геометрический смысл. На данном примере проиллюстрирован критерий, который мы будем использовать для удовлетворения произвольных вариационных связей.

Сумма квадратов изменений варьируемых параметров отражает реакцию геометрических объектов с наложенными вариационными связями, на изменение констант в уравнениях связей или на изменения самих объектов. Критерий изменения параметров сформулируем следующим образом. Реакция геометрических объектов с наложенными на них вариационными связями на любое возмущение должна быть такой, чтобы сумма квадратов изменений варьируемых параметров была бы минимальной. Данный критерий ставит все варьируемые параметры в равные условия. Критерий можно сформулировать в виде задачи минимизации некоторой функции или (в общем случае) в виде задачи минимизации функционала. Ниже мы рассмотрим обе формулировки критерия поведения геометрических объектов с наложенными вариационными связями.

13.7. Условный экстремум функции изменения параметров

Пусть имеется некоторая совокупность вариационных связей, наложенных на n параметров $x_1, x_2, \ldots, x_n$ геометрических объектов. Пусть эта совокупность связей накладывает m уравнений

$$\begin{array}{l} f_1(x_1, x_2, \ldots, x_n) = 0, \\ f_2(x_1, x_2, \ldots, x_n) = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ f_m(x_1, x_2, \ldots, x_n) = 0, \end{array} \tag{7.1}$$

на параметры объектов. Связи фиксации здесь учитывать не будем. Фиксированные параметры в процессе решения не изменяются, поэтому в других уравнениях связей они не варьируются и считаются константами. Если $m=n$, то все параметры могут быть найдены путем решения системы уравнений вариационных связей. Если $m<n$, то для определения параметров нужно составить систему уравнений, в которой число определяемых параметров было бы равно числу уравнений, и среди уравнений содержались бы уравнения (7.1). Систему уравнений получим из требования экстремума некоторой функции параметров, несущей определенный геометрический смысл. Это требование является критерием поведения параметров связей. Для вариационных связей, управляющих координатами точек и размерными параметрами, критерием поведения нам будет служить требование минимума суммы квадратов изменений варьируемых параметров, т. е. минимума функции

$$\psi = \frac{1}{2}\sum_{i=1}^{n}(x_i - x_i^{(0)})^2,$$

где $x_i^{(0)}$— исходные значения варьируемых параметров.

На примере линейного размера выше было показано применение этого критерия совместно с методом неопределенных множителей Лагранжа.

Рассмотрим теоретические основы использованного выше метода неопределенных множителей Лагранжа для нахождения условного минимума или максимума действительной функции нескольких переменных параметров.

Экстремум функции.

Пусть некоторое требование, предъявленное к варьируемым параметрам x_i, $i=1,2,\ldots,n$, определяется экстремумом функции

$$\psi = \psi(x_1, x_2, \ldots, x_n). \quad (7.2)$$

По определению действительная функция (7.2) в точке $x_i= a_i$, $i=1,2,\ldots,n$ (конкретную совокупность параметров будем называть точкой) имеет минимум (максимум), если существует такое положительное число δ, что при любых $\Delta x_1, \Delta x_2, \ldots, \Delta x_n$ удовлетворяющих неравенствам

$$0 < \sqrt{\Delta x_1^{\,2} + \Delta x_2^{\,2} + \ldots + \Delta x_n^{\,2}} < \delta,$$

приращение функции

$$\Delta\psi = \psi + (a_1 + \Delta x_1, a_2 + \Delta x_2, \ldots, a_n + \Delta x_n) - \psi(a_1, a_2, \ldots, a_n) \quad (7.3)$$

соответственно больше (меньше) нуля. Будем считать, что функция ψ в окрестности экстремума имеет непрерывные частные производные до второго порядка включительно. Разложим в ряд Тейлора в окрестности точки $x_i = a_i,\ i=1,2,\dots,n$ первое слагаемое правой части (7.3) и получим

$$\Delta\psi = \sum_{i=1}^{n} \frac{\partial\psi}{\partial x_i}\Delta x_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2\psi}{\partial x_i\,\partial x_j}\Delta x_i \Delta x_j + o(\Delta x^3), \tag{7.4}$$

Где $o(\Delta x^3)$— слагаемые третьего порядка относительно Δx_i. Слагаемые первого и второго порядка относительноΔx_i в (7.4) составляют первый $d\psi$ и второй $d^2\psi$ дифференциалы функции ψ.

Дифференцируемая функция $\psi(x_1, x_2, \dots, x_n)$ в точке $x_i = a_i,\ i=1,2,\dots,n$ имеет минимум или максимум лишь в том случае, когда ее первый дифференциал обращается в этой точке в нуль

$$d\psi = \frac{\partial\psi}{\partial x_1}\,dx_1 + \frac{\partial\psi}{\partial x_2}\,dx_2 + \ldots + \frac{\partial\psi}{\partial x_n}\,dx_n = 0. \tag{7.5}$$

Для этого необходимо выполнение равенств

$$\frac{\partial\psi}{\partial x_1} = 0, \quad \frac{\partial\psi}{\partial x_2} = 0, \quad \dots, \quad \frac{\partial\psi}{\partial x_n} = 0, \tag{7.6}$$

при $x_1 = a_1, x_2 = a_2, \dots, x_n = a_n$. Выполнение равенств (7.6) является необходимым условием экстремума. Достаточным условием экстремума функции $\psi(x_1, x_2, \dots, x_n)$ является знакоопределенность второго дифференциала

$$d^2\psi = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2\psi}{\partial x_i\,\partial x_j}\,dx_i\,dx_j, \tag{7.7}$$

при $x_1 = a_1, x_2 = a_2, \dots, x_n = a_n$. Если второй дифференциал (7.7) представляет собой положительно определенную квадратичную форму, то функция в рассматриваемой точке имеет локальный минимум. Если второй дифференциал (7.7) есть отрицательно определенная квадратичная форма, то функция в рассматриваемой точке имеет локальный максимум.

Если второй дифференциал (7.7) не является знакоопределенной квадратичной формой, то функция в рассматриваемой точке не достигает экстремума.

Условный экстремум функции.

Все сказанное об экстремуме функции (7.2) справедливо при отсутствии условий (7.1). Нас интересует минимум функции (7.2), аргументы которой удовлетворяют уравнениям связей (7.1). Будем говорить, что действительная функция (7.2) в точке $x_i = b_i,\ i=1,2,\ldots,n$ имеет условный минимум (условный максимум), если существует такая окрестность этой точки, в пределах которой значение функции в точке $x_i = b_i,\ i=1,2,\ldots,n$ является наименьшим (наибольшим) среди ее значений во всех точках окрестности, удовлетворяющих уравнениям связей (7.1).

Если уравнения (7.1) позволяют выразить любые m параметров через остальные n-m в виде функций, например,

$$\begin{aligned} x_1 &= g_1(x_{m+1}, x_{m+2}, \ldots, x_n), \\ x_2 &= g_2(x_{m+1}, x_{m+2}, \ldots, x_n), \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_m &= g_m(x_{m+1}, x_{m+2}, \ldots, x_n), \end{aligned} \tag{7.8}$$

то задача об условном экстремуме сведется к задача об обычном экстремуме функции. Для этого подставим равенства (7.8) в (7.2) и найдем безусловный экстремум функции

$$\psi = \psi(g_1(x_{m+1}, x_{m+2}, \ldots, x_n), g_2(x_{m+1}, x_{m+2}, \ldots, x_n), \ldots \\ \ldots, g_m(x_{m+1}, x_{m+2}, \ldots, x_n), x_{m+1}, x_{m+2}, \ldots, x_n), \tag{7.9}$$

зависящей от параметров $x_{n-m+1}, x_{n-m+2}, \ldots, x_n$.

Установим необходимые условия существования условного экстремума функции ψ когда невозможно получить выражения (7.8). В точке условного экстремума должен быть равен нулю первый дифференциал функции (7.5). Но при наличии условий

(7.1) не все дифференциалы параметров являются независимыми. Дифференциалы параметров в точке условного экстремума связаны уравнениями

$$\begin{aligned}
&\frac{\partial f_1}{\partial x_1}dx_1+\frac{\partial f_1}{\partial x_2}dx_2+\ldots+\frac{\partial f_1}{\partial x_n}dx_n=0,\\
&\frac{\partial f_2}{\partial x_1}dx_1+\frac{\partial f_2}{\partial x_2}dx_2+\ldots+\frac{\partial f_2}{\partial x_n}dx_n=0,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\frac{\partial f_m}{\partial x_1}dx_1+\frac{\partial f_m}{\partial x_2}dx_2+\ldots+\frac{\partial f_m}{\partial x_n}dx_n=0.
\end{aligned} \quad (7.10)$$

Действительно, если точки $x_i = b_i$ и $x_i = b_i + \Delta x_i$, $i=1,2,\ldots,n$ удовлетворяют уравнениям связей, то в пределе $\Delta x_i \to 0$, $i=1,2,\ldots,n$ разность уравнений (7.1) для этих точек даст уравнения (7.10). Из системы m уравнений (7.10) найдем любые m дифференциалов, пусть это будут $dx_1, dx_2, \ldots, dx_n$. Для этого необходимо, чтобы якобиан системы уравнений связей был отличен от нуля. Найденные дифференциалы подставим в первый дифференциал функции (7.5) и получим

$$d\psi = H_{m+1}\,dx_{m+1} + H_{m+2}\,dx_{m+2} + \ldots + H_n\,dx_n = 0, \quad (7.11)$$

где через H_{m+1}, H_{m+2}, H_n обозначены рациональные функции частных производных ψ, $f_1, f_2, \ldots, f_m$ в точке условного минимума. Дифференциалы $dx_{m+1}, dx_{m+2}, \ldots, dx_n$ являются независимыми переменными выражения (7.11), поэтому равенство нулю первого дифференциала (7.11) возможно, когда выполняются равенства

$$H_{m+1}=0, \quad H_{m+2}=0, \quad \ldots, \quad H_n=0. \quad (7.12)$$

Уравнения (7.12) совместно с уравнениями связей (7.1) представляют собой необходимое условие условного экстремума функции (7.2).

Метод множителей Лагранжа.

В описанном выше методе отыскания условного экстремума функции параметры $x_1, x_2, \ldots, x_n$ не являлись равноправными. Часть из них мы должны были выразить через остальные параметры. В результате система уравнений для определения условного экстремума являлась несимметричной относительно параметров. Лагранжем предложен метод, который делает параметры равноправными, а уравнения симметричными относительно параметров. Умножим равенства (7.10) на неопределенные пока постоянные множители $\lambda_1, \lambda_2, \ldots, \lambda_m$ и добавим к равенству (7.5). В результате получим равенство

$$\frac{\partial F}{\partial x_1} dx_1 + \frac{\partial F}{\partial x_2} dx_2 + \ldots + \frac{\partial F}{\partial x_n} dx_n = 0, \tag{7.13}$$

где

$$F = F(x_1, x_2, \ldots, x_n) = \psi + \lambda_1 f_1 + \lambda_2 f_2 + \ldots + \lambda_m f_m. \tag{7.14}$$

Функция F называется функцией Лагранжа. Так как параметры $x_1, x_2, \ldots, x_n$ связаны уравнениями связей, то их дифференциалы в (7.13) не являются независимыми. Будем считать, что первые параметров зависят через уравнения связей от остальных n-m параметров. Параметры $x_{m+1}, x_{m+2}, \ldots, x_n$ будем считать независимыми. Множители $\lambda_1, \lambda_2, \ldots, \lambda_m$ выберем так, чтобы выполнялись равенства

$$\frac{\partial F}{\partial x_1} = 0, \quad \frac{\partial F}{\partial x_2} = 0, \quad \ldots, \quad \frac{\partial F}{\partial x_m} = 0. \tag{7.15}$$

Это можно сделать, решив систему уравнений

$$\lambda_1\frac{\partial f_1}{\partial x_1}+\lambda_2\frac{\partial f_2}{\partial x_1}+\ldots+\lambda_m\frac{\partial f_m}{\partial x_1}=-\frac{\partial\psi}{\partial x_1},$$
$$\lambda_1\frac{\partial f_1}{\partial x_2}+\lambda_2\frac{\partial f_2}{\partial x_2}+\ldots+\lambda_m\frac{\partial f_m}{\partial x_2}=-\frac{\partial\psi}{\partial x_2},$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\lambda_1\frac{\partial f_1}{\partial x_m}+\lambda_2\frac{\partial f_2}{\partial x_m}+\ldots+\lambda_m\frac{\partial f_m}{\partial x_m}=-\frac{\partial\psi}{\partial x_m} \quad (7.16)$$

относительно множителей $\lambda_1, \lambda_2, \ldots, \lambda_m$. Как уже было сказано, определитель матрицы этой системы линейных алгебраических уравнений должен быть отличен от нуля. При выполнении равенств (7.15) для выполнения равенства (7.13) требуется, чтобы

$$\frac{\partial F}{\partial x_{m+1}}=0, \quad \frac{\partial F}{\partial x_{m+2}}=0, \quad \ldots, \quad \frac{\partial F}{\partial x_n}=0. \quad (7.17)$$

Объединив равенства (7.15), (7.17) и уравнения связи (7.1), придем к тому, что в точке условного экстремума функции (7.2) должны выполняться $n+m$ равенств

$$\frac{\partial\psi}{\partial x_1}+\lambda_1\frac{\partial f_1}{\partial x_1}+\lambda_2\frac{\partial f_2}{\partial x_1}+\ldots+\lambda_m\frac{\partial f_m}{\partial x_1}=0,$$
$$\frac{\partial\psi}{\partial x_2}+\lambda_1\frac{\partial f_1}{\partial x_2}+\lambda_2\frac{\partial f_2}{\partial x_2}+\ldots+\lambda_m\frac{\partial f_m}{\partial x_2}=0,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\frac{\partial\psi}{\partial x_n}+\lambda_1\frac{\partial f_1}{\partial x_n}+\lambda_2\frac{\partial f_2}{\partial x_n}+\ldots+\lambda_m\frac{\partial f_m}{\partial x_n}=0,$$
$$f_1(x_1, x_2, \ldots, x_n)=0,$$
$$f_2(x_1, x_2, \ldots, x_n)=0,$$
$$\cdots\cdots\cdots\cdots$$
$$f_m(x_1, x_2, \ldots, x_n)=0. \quad (7.18)$$

Эта система уравнений позволяет определить параметры условного экстремума функции ψ и множители Лагранжа. Мы видим, что поиск условного экстремума функции ψ привел к уравнениям безусловного экстремума функции Лагранжа F. **Практически метод Лагранжа реализуется следующим образом.** Составляется функция Лагранжа (7.14) и для нее находят возможные точки безусловного экстремума. Для исключения множителей $\lambda_1, \lambda_2, \ldots, \lambda_m$ привлекаются уравнения связей.

Найденное из системы уравнений (7.18) решение может не являться точкой экстремума. В точке условного минимума (условного максимума) функции ψдолжны выполняться достаточные условия. При наличии связей экстремумы функций ψ и F совпадают. Тогда из вышеизложенного следует, что для получения достаточного условия условного экстремума функции ψ к уравнениям (7.18) нужно добавить требование знакоопределенности d^2F

$$d^2F = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 F}{\partial x_i\,\partial x_j}\,dx_i\,dx_j.$$

В функции Лагранжа все параметры $x_1, x_2, \ldots, x_n$ можно рассматривать как независимые, если считать, что множители $\lambda_1, \lambda_2, \ldots, \lambda_m$ определены из уравнений (7.16).

В методе Лагранжа все параметры равноправны, а система уравнений (7.18) является симметричной относительно параметров. Кроме того, все необходимые уравнения получены из единого требования к поведению параметров. Платой за эту симметрию является увеличение числа неизвестных из-за введения неопределенных множителей.

Функция $\psi(x_1, x_2, \ldots, x_n)$ формулирует критерий поведения варьируемых параметров $x_1, x_2, \ldots, x_n$ при их переходе из начального состояния в конечное состояние, если изменены константы уравнений связей или сами уравнения. Принятый нами критерий в форме минимума функции

$$\psi(x_1, x_2, \ldots, x_n) = \frac{1}{2}\sum_{i=1}^{n}(x_i - x_i^{(0)})^2, \tag{7.20}$$

где $x_i^{(0)}$— начальные значения варьируемых параметров, предъявляет требование наименьшего суммарного изменения параметров, участвующих в связях. Конечные значения параметров $x_1, x_2, \ldots, x_n$ должны быть найдены из системы уравнений (7.18).

Система уравнений (7.18) при использовании критерия (7.20) примет вид

$$
\begin{gathered}
x_1 - x_1^{(0)} + \lambda_1 \frac{\partial f_1}{\partial x_1} + \lambda_2 \frac{\partial f_2}{\partial x_1} + \ldots + \lambda_m \frac{\partial f_m}{\partial x_1} = 0, \\
x_2 - x_2^{(0)} + \lambda_1 \frac{\partial f_1}{\partial x_2} + \lambda_2 \frac{\partial f_2}{\partial x_2} + \ldots + \lambda_m \frac{\partial f_m}{\partial x_2} = 0, \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
x_n - x_n^{(0)} + \lambda_1 \frac{\partial f_1}{\partial x_n} + \lambda_2 \frac{\partial f_2}{\partial x_n} + \ldots + \lambda_m \frac{\partial f_m}{\partial x_n} = 0, \\
f_1(x_1, x_2, \ldots, x_n) = 0, \\
f_2(x_1, x_2, \ldots, x_n) = 0, \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
f_m(x_1, x_2, \ldots, x_n) = 0.
\end{gathered}
\tag{7.21}
$$

Критерий (7.20) будем использовать при наложении вариационных связей на координаты точек и на любые другие параметры геометрических объектов.

13.8. Вариационный метод определения изменений параметров

Требование минимума суммарного изменения параметров связей приводит к задаче поиска условного экстремума некоторой функции. Покажем, что в более общей постановке эта задача является вариационной и сводится к поиску экстремума некоторого функционала.

Функционал.

Рассмотрим точку, координаты которой участвуют в некоторых уравнениях вариационных связей. Представим, что мы изменили или сами вариационные уравнения, или некоторые их параметры. В результате координаты точки в общем случае должны измениться. Пусть координаты рассматриваемой точки в начальном положении равны $x^{(0)}$ $y^{(0)}$ $z^{(0)}$, а в конечном положении равны x, y, z. В декартовой системе координат квадрат расстояния между начальным и конечным положением точки равен

$$2\psi = (x - x^{(0)})^2 + (y - y^{(0)})^2 + (z - z^{(0)})^2.$$

Для минимизации перемещения точки в декартовой системе координат можно потребовать условного минимума функции ψ. В результате мы придем к системе уравнений (7.21), из которой определим координаты рассматриваемой точки в конечном положении.

В криволинейной системе координат квадрат расстояния между бесконечно близкими положениями рассматриваемой точки определяется формулой

$$(ds)^2 = g_{11}(dx)^2 + g_{22}(dy)^2 + g_{33}(dz)^2 + 2g_{12}\,dx\,dy + 2g_{23}\,dy\,dz + 2g_{13}\,dx\,dz.$$

$g_{ij} = g_{ij}\,(x, y, z)$— ковариантные компоненты метрического тензора. Пусть каждая координата является функцией некоторого параметра t. Тогда функции $x(t), y(t), z(t)$ будут описывать кривую, вдоль которой осуществляется перемещение точки. Величина

$$\left(\frac{ds}{dt}\right)^2 = g_{11}\left(\frac{dx}{dt}\right)^2 + g_{22}\left(\frac{dy}{dt}\right)^2 + g_{33}\left(\frac{dz}{dt}\right)^2 + 2g_{12}\frac{dx}{dt}\frac{dy}{dt} + 2g_{23}\frac{dy}{dt}\frac{dz}{dt} + 2g_{13}\frac{dx}{dt}\frac{dz}{dt}$$

равна квадрату первой производной этой кривой.

Рассмотрим интеграл

$$\Phi = \frac{1}{2}\int_{t_{\min}}^{t_{\max}}\left(g_{11}\left(\frac{dx}{dt}\right)^2 + g_{22}\left(\frac{dy}{dt}\right)^2 + g_{33}\left(\frac{dz}{dt}\right)^2 + 2g_{12}\frac{dx}{dt}\frac{dy}{dt} + 2g_{23}\frac{dy}{dt}\frac{dz}{dt} + 2g_{13}\frac{dx}{dt}\frac{dz}{dt}\right)dt,$$

где t — параметр траектории перемещения точки. Для минимизации перемещения точки в криволинейной системе координат потребуем минимума интеграла Ф. В этом интеграле координаты точки считаются функциями некоторого параметра t. Для начального значения параметра координаты принимают начальные значения: $x(t_{min}) = x^{(0)}$, $y(t_{min}) = y^{(0)}$, $z(t_{min}) = z^{(0)}$. Конечные значения координат $x(t_{max}) = x$, $y(t_{max}) = y$, $z(t_{max}) = z$ являются искомыми. Они должны удовлетворять уравнениям вариационных связей. Значение интеграла Ф зависит от функций x(t), y(t), z(t). Подобные интегралы называются функционалами. Таким образом, требование минимума суммы квадратов изменения параметров в более общей постановке сводится к поиску минимума некоторого функционала. Если варьируемым точкам приписать некоторую массу а параметр t считать временем, то принцип минимума суммы квадратов изменения параметров будет аналогичен принципу наименьшего действия в форме Гамильтона для движения механической системы.

В декартовой прямоугольной системе координат интеграл Ф для одной точки определяется формулой

$$\Phi = \frac{1}{2}\int_{t_{min}}^{t_{max}} \left(\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2 \right) dt.$$

Функционал представляет собой переменную величину, значение которой зависит от выбора одной или нескольких варьируемых функций. Параметры связей в этом функционале выступают в роли варьируемых функций. Начальные значения этих функций известны, а конечные значения должны удовлетворять уравнениям связей. Переход параметров из начального состояния в конечное состояние должен минимизировать некоторый функционал, выступающий в роли критерия такого перехода. Например, требование минимума суммы квадратов перемещения точек (6.1) заменим общим требованием минимума функционала

$$\Phi = \frac{1}{2}\int_{t_{min}}^{t_{max}} \left(\left(\frac{dx_1}{dt}\right)^2 + \left(\frac{dy_1}{dt}\right)^2 + \left(\frac{dz_1}{dt}\right)^2 + \left(\frac{dx_2}{dt}\right)^2 + \left(\frac{dy_2}{dt}\right)^2 + \left(\frac{dz_2}{dt}\right)^2 \right) dt, \quad (8.1)$$

где x_1, y_1, z_1 , x_2, y_2, z_2— координаты точек **p₁** и **p₂** соответственно. Функционал записан в декартовой прямоугольной системе координат. Координаты точек мы представили как функции параметра t, которому можно приписать смысл параметра траекторий перехода точек из исходного состояния в новое состояние при изменении констант уравнений связей. Начальное t_{min}и конечное t_{max} значения параметра t будем считать известными. Для определенности положим t_{min} =0, t_{max}=1. Отличие вариационной постановки (8.1) данной задачи от задачи в постановке (6.1) заключается в следующем. Функция (6.1) предполагает, что точки перемещаются по прямой линии, а в функционале (8.1) такое предположение отсутствует, и перемещение точек может быть произвольным. Функция (6.1) не зависит от пути изменения параметров, тогда как функционал (8.1) зависит от траектории перехода параметров из начального состояния в конечное состояние. Начальное положение точек нам известно.

В своем конечном положении координаты точек должны удовлетворять уравнениям связи. Вариационная постановка дает больше свободы в выборе критерия перехода параметров из начального состояния в конечное состояние. Мы увидим, что в декартовой системе координат условный экстремум функции (6.1) и условный экстремум функционала (8.1) приводят к одной и той же системе уравнений. В криволинейной системе координат это не так. Рассмотрим метод вариаций в общем случае.

Вариация функционала.

Пусть некоторое требование, предъявленное к варьируемым функциям $x_1(t)$, $x_2(t)$,…, $x_n(t)$ определяется экстремумом функционала вида

$$\Phi = \int_0^1 h\left(t,\, x_1,\, x_2,\, \ldots,\, x_n,\, \frac{dx_1}{dt},\, \frac{dx_2}{dt},\, \ldots,\, \frac{dx_n}{dt}\right) dt. \qquad (8.2)$$

Функцию $h=(t, x_1, x_2,\ldots, x_n, x'_1, x'_2,\ldots, x'_n)$будем считать дифференцируемой необходимое число раз по t, по каждой функции $x_1(t)$, $x_2(t)$,…, $x_n(t)$ и по каждой первой производной $x'_1(t)$, $x'_2(t)$,…, $x'_n(t)$ каждой функции по параметру t. В функционале x_1, x_2,…, x_n выступают в роли варьируемых функций некоторого параметра t. Параметром t может служить общий параметр траекторий, описывающих изменение функций $x_1, x_2, \ldots, x_n$. Областью изменения параметра t нами выбран отрезок 0≤t≤1 Значение функционала Ф зависит от функций $x_1(t), x_2(t), \ldots, x_n(t)$. Предположим, что экстремум функционала достигается на дважды дифференцируемых функциях $x_1(t), x_2(t), \ldots, x_n(t)$.

Возьмем совокупность близких к ним функций $x_i(t)+\delta x_i(t), i=1,2,\ldots,n$ и представим их в виде

$$x_i(t, \alpha) = x_i(t) + \alpha\delta x_i(t), \quad 0 \leqslant \alpha \leqslant 1, \quad i = 1, 2, \ldots, n. \tag{8.3}$$

Величины $\delta x_i(t)$ называются вариациями соответствующих функций. Эти функции можно дифференцировать по параметру t

$$\frac{d(\delta x_i(t))}{dt} = \delta x_i{}' \equiv \delta\left(\frac{dx_i}{dt}\right), \quad \frac{d^2(\delta x_i(t))}{dt^2} = \delta x_i{}'' \equiv \delta\left(\frac{d^2 x_i}{dt^2}\right), \quad \ldots$$
$$\ldots, \quad \frac{d^k(\delta x_i(t))}{dt^k} = \delta x_i{}^{(k)} \equiv \delta\left(\frac{d^k x_i}{dt^k}\right), \quad i = 1, 2, \ldots, n.$$

Подставим (8.3) в (8.2) и получим новое значение функционала Ф(α). По предположению эта функция достигает экстремума при α=0. Необходимым условием экстремума функции Ф(α) является равенство нулю ее первой производной Ф′(0)=0. Величина Ф′(0) называется вариацией функционала и обозначается через δФ. Найдем производную функции Ф(α) по α как производную сложной функции и положим в ней α=0. В результате получим

$$\left.\frac{d\Phi(\alpha)}{d\alpha}\right|_{\alpha=0} =$$
$$= \int_0^1 (h_1\delta x_1 + h_2\delta x_2 + \ldots + h_n\delta x_n + h_{1(1)}\delta x_1{}' + h_{2(1)}\delta x_2{}' + \ldots + h_{n(1)}\delta x_n{}')\, dt, \tag{8.4}$$

где

$$h_1 = \frac{\partial h(t,\, x_1,\, x_2,\, \ldots,\, x_n,\, x_1{}',\, x_2{}',\, \ldots,\, x_n{}')}{\partial x_1},$$

$$h_2 = \frac{\partial h(t,\, x_1,\, x_2,\, \ldots,\, x_n,\, x_1{}',\, x_2{}',\, \ldots,\, x_n{}')}{\partial x_2},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$h_n = \frac{\partial h(t,\, x_1,\, x_2,\, \ldots,\, x_n,\, x_1{}',\, x_2{}',\, \ldots,\, x_n{}')}{\partial x_n},$$

$$h_{1(1)} = \frac{\partial h(t,\, x_1,\, x_2,\, \ldots,\, x_n,\, x_1{}',\, x_2{}',\, \ldots,\, x_n{}')}{\partial (dx_1/dt)},$$

$$h_{2(1)} = \frac{\partial h(t,\, x_1,\, x_2,\, \ldots,\, x_n,\, x_1{}',\, x_2{}',\, \ldots,\, x_n{}')}{\partial (dx_2/dt)},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$h_{n(1)} = \frac{\partial h(t,\, x_1,\, x_2,\, \ldots,\, x_n,\, x_1{}',\, x_2{}',\, \ldots,\, x_n{}')}{\partial (dx_n/dt)}.$$

Таким образом, необходимым условием экстремума функционала является равенство нулю его вариации, что выражается формулой

$$\int_0^1 (h_1 \delta x_1 + h_2 \delta x_2 + \ldots + h_n \delta x_n + h_{1(1)} \delta x_1{}' + h_{2(1)} \delta x_2{}' + \ldots + h_{n(1)} \delta x_n{}')\, dt = 0. \tag{8.5}$$

Проинтегрируем по частям вторую половину слагаемых в (8.5)

$$\int_0^1 h_{i(1)} \delta x_i{}'\, dt = \int_0^1 h_{i(1)} \frac{d(\delta x_i)}{dt}\, dt = \int_0^1 \left(\frac{d(h_{i(1)} \delta x_i)}{dt} - \frac{dh_{i(1)}}{dt} \delta x_i \right) dt =$$

$$= h_{i(1)} \delta x_i \big|_0^1 - \int_0^1 \frac{dh_{i(1)}}{dt} \delta x_i\, dt, \quad i = 1,\, 2,\, \ldots,\, n.$$

В результате получим, что необходимое условие экстремума функционала (8.2) выражается равенством

$$\sum_{i=1}^{n} h_{i(1)}\delta x_i\Big|_0^1 + \int_0^1 \sum_{i=1}^{n}\left(\left(h_i - \frac{dh_{i(1)}}{dt}\right)\delta x_i\right)dt = 0. \tag{8.6}$$

Уравнения Эйлера. Если функции $x_1, x_2,\ldots, x_n$ при вариации не изменяют свои значения в начальных и конечных точках, т. е. если $\delta x_i(t_{\min})\equiv \delta x_i(0)=0$ и $x_i(t_{\max}) \equiv \delta x_i(1)=0$, то все слагаемые

$$\sum_{i=1}^{n} h_{i(1)}\delta x_i\Big|_0^1 = 0, \quad i = 1, 2, \ldots, n. \tag{8.7}$$

Это означает, что функции $x_1(t), x_2(t),\ldots, x_n(t)$ и все их вариации проходят через фиксированные граничные точки. При этих условиях для выполнения равенства (8.6) необходимо выполнение равенств

$$\begin{gathered} h_1 - \frac{dh_{1(1)}}{dt} = 0, \\ h_2 - \frac{dh_{2(1)}}{dt} = 0, \\ \cdots\cdots\cdots\cdots\cdots \\ h_n - \frac{dh_{n(1)}}{dt} = 0, \end{gathered} \tag{8.8}$$

так как вариации δx_i произвольны и в общем случае не равны нулю. Дифференциальные уравнения (8.8) называются уравнениями Эйлера. После дифференцирования по параметру t как сложной функции каждое из равенств (8.8) примет вид

$$h_i - \frac{\partial h_{i(1)}}{\partial t} - \sum_{j=1}^{n}\frac{\partial h_{i(1)}}{\partial x_j}\frac{dx_j}{dt} - \sum_{j=1}^{n}\frac{\partial h_{i(1)}}{\partial x_j{}'}\frac{d^2 x_j}{dt^2} = 0, \quad i = 1, 2, \ldots, n,$$

или

$$\frac{\partial h}{\partial x_i} - \frac{\partial^2 h}{\partial x_i' \partial t} - \sum_{j=1}^{n} \frac{\partial^2 h}{\partial x_j \partial x_i'} \frac{dx_j}{dt} - \sum_{j=1}^{n} \frac{\partial^2 h}{\partial x_j' \partial x_i'} \frac{d^2 x_j}{dt^2} = 0,$$
$$i = 1, 2, \ldots, n. \qquad (8.9)$$

Проинтегрировав систему (8.8), получим общее решение для экстремалей, каждая из которых содержит по две константы. Константы могут быть найдены из граничных условий $\delta x_i |^1_0=0$. Заметим, что система (8.8) не всегда имеет решение, а если решение существует, то оно может быть не единственным или не удовлетворять достаточным условиям экстремума. Для функционала (8.1) уравнения (8.9) примут более простой вид:

$$\frac{d^2 x_i}{dt^2} = 0, \quad \frac{d^2 y_i}{dt^2} = 0, \quad \frac{d^2 z_i}{dt^2} = 0, \quad i = 1, 2. \qquad (8.10)$$

Решением системы (8.10) являются линейные функции $x_i(t)=C_{0xi}+ C_{1xi}t$, $y_i(t)=C_{0yi}+ C_{1yi}t$, $z_i(t)=C_{0zi}+ C_{1zi}t$,. Константы $C_{0xi}, C_{1xi}, C_{0yi}, C_{1yi}, C_{0zi}, C_{1zi}$ должны быть найдены из краевых условий (8.7).

В функционале (8.2) функция h зависит от $x_1, x_2,\ldots, x_n$ и их первых производных $x'_1, x'_2,\ldots, x'_n$ по t. В общем случае функция h может зависеть от производных более высокого порядка. Рассмотрим функционал вида

$$\Phi(h) = \int_0^1 h(t, x_1, x_2, \ldots, x_n, x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}, x_1^{(2)}, x_2^{(2)}, \ldots, x_n^{(2)}, \ldots, x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)})\, dt, \qquad (8.11)$$

где $x_1^{(k)}, x_2^{(k)},\ldots, x_n^{(k)}$— производные k-го порядка варьируемых функций $x_1, x_2,\ldots, x_n$ по параметру t.

Подставим вариации (8.3) функций и их производные в (8.11), найдем производную $\Phi'(\alpha)$ при $\alpha=0$ и приравняем ее нулю. В результате получим, что необходимое условие экстремума функционала (8.11) выражается равенством

$$\int_0^1 \sum_{i=1}^{n} (h_i \delta x_i + h_{i(1)} \delta x_i{}^{(1)} + h_{i(2)} \delta x_i{}^{(2)} + \ldots + h_{i(k)} \delta x_i{}^{(k)})\, dt = 0. \tag{8.12}$$

где

$$h_i = \frac{\partial h(t,\, x_1,\, x_2,\, \ldots,\, x_n,\, x_1{}^{(1)},\, x_2{}^{(1)},\, \ldots,\, x_n{}^{(1)},\, \ldots,\, x_1{}^{(k)},\, x_2{}^{(k)},\, \ldots,\, x_n{}^{(k)})}{\partial x_i},$$

$$h_{i(m)} = \frac{\partial h(t,\, x_1,\, x_2,\, \ldots,\, x_n,\, x_1{}^{(1)},\, x_2{}^{(1)},\, \ldots,\, x_n{}^{(1)},\, \ldots,\, x_1{}^{(k)},\, x_2{}^{(k)},\, \ldots,\, x_n{}^{(k)})}{\partial (d^m x_i / dt^m)},$$

$$i = 1,\, 2,\, \ldots,\, n, \quad m = 1,\, 2,\, \ldots,\, k.$$

Проинтегрируем *k* раз по частям слагаемые, в которые входят *k*-е производные вариаций $\delta x_i{}^{(k)}$

$$\int_0^1 h_{i(k)} \delta x_i{}^{(k)}\, dt = h_{i(k)} \delta x_i{}^{(k-1)} \Big|_0^1 - \frac{dh_{i(k)}}{dt} \delta x_i{}^{(k-2)} \Big|_0^1 + \frac{d^2 h_{i(k)}}{dt^2} \delta x_i{}^{(k-3)} \Big|_0^1 - \ldots$$
$$\ldots + (-1)^{k-1} \frac{d^{k-1} h_{i(k)}}{dt^{k-1}} \delta x_i \Big|_0^1 + (-1)^k \int_0^1 \frac{d^k h_{i(k)}}{dt^k} \delta x_i\, dt, \qquad i = 1,\, 2,\, \ldots,\, n. \tag{8.13}$$

В результате получим, что необходимое условие экстремума функционала (8.11) выражается равенством

$$\sum_{i=1}^{n}\left\{ h_{i(k)}\delta x_i^{(k-1)}\Big|_0^1 + \left(h_{i(k-1)} - \frac{dh_{i(k)}}{dt}\right)\delta x_i^{(k-2)}\Big|_0^1 + \right.$$
$$+\left(h_{i(k-2)} - \frac{dh_{i(k-1)}}{dt} + \frac{d^2 h_{i(k)}}{dt^2}\right)\delta x_i^{(k-3)}\Big|_0^1 + \ldots$$
$$\left.\ldots + \left(h_{i(1)} - \frac{dh_{i(2)}}{dt} + \ldots + (-1)^{k-1}\frac{d^{k-1}h_{i(k)}}{dt^{k-1}}\right)\delta x_i\Big|_0^1\right\} +$$
$$+\sum_{i=1}^{n}\int_0^1 \left(h_i - \frac{dh_{i(1)}}{dt} + \frac{d^2 h_{i(2)}}{dt^2} + \ldots + (-1)^k \frac{d^k h_{i(k)}}{dt^k}\right)\delta x_i\, dt = 0,$$
$$i = 1, 2, \ldots, n. \tag{8.14}$$

Если функции $x_1, x_2, \ldots, x_n$ и их производные до (k-1) порядка включительно при вариации не изменяют свои значения в начальных и конечных точках, то

$$h_{i(m)}\delta x_i^{(m-1)}\Big|_0^1 - \frac{dh_{i(m)}}{dt}\delta x_i^{(m-2)}\Big|_0^1 + \frac{d^2 h_{i(m)}}{dt^2}\delta x_i^{(m-3)}\Big|_0^1 - \ldots$$
$$\ldots + (-1)^{m-1}\frac{d^{m-1}h_{i(m)}}{dt^{m-1}}\delta x_i\Big|_0^1 = 0, \quad m = 1, 2, \ldots, k, \quad i = 1, 2, \ldots, n. \tag{8.15}$$

При этих условиях экстремали функционала (8.11) должны удовлетворять уравнениям

$$\begin{gathered}
h_1 - \frac{dh_{1(1)}}{dt} + \frac{d^2 h_{1(2)}}{dt^2} + \ldots + (-1)^k \frac{d^k h_{1(k)}}{dt^k} = 0, \\
h_2 - \frac{dh_{2(1)}}{dt} + \frac{d^2 h_{2(2)}}{dt^2} + \ldots + (-1)^k \frac{d^k h_{2(k)}}{dt^k} = 0, \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
h_n - \frac{dh_{n(1)}}{dt} + \frac{d^2 h_{n(2)}}{dt^2} + \ldots + (-1)^k \frac{d^k h_{n(k)}}{dt^k} = 0,
\end{gathered} \tag{8.16}$$

так как вариации δx_i произвольны и в общем случае не равны нулю. Дифференциальные уравнения системы (8.16) имеют порядок $2k$ и называются уравнениями Эйлера-Пуассона. Общее решение для каждой варьируемой функции, если оно существует, содержит $2k$ произвольных постоянных, которые могут быть определены из условий на границах функций

$$\delta x_i^{(k-1)}\Big|_0^1 = 0, \quad \delta x_i^{(k-2)}\Big|_0^1 = 0, \quad \delta x_i^{(k-3)}\Big|_0^1 = 0, \quad \ldots, \quad \delta x_i\Big|_0^1 = 0, \quad i = 1, 2, \ldots, n.$$

Подвижные границы.

При выводе уравнений Эйлера предполагалось, что значения функций $x_1(t), x_2(t),\ldots, x_n(t)$ в крайних точках не изменяются при варьировании этих функций. В решаемых нами задачах известны только значения функций на одной границе — значения $x_1(t), x_2(t),\ldots, x_n(t)$ при $t= t_{\min} =0$. Значения функций $x_1(t), x_2(t),\ldots, x_n(t)$ при $t= t_{\max} =1$являются искомыми. Рассмотрим задачу определения экстремума функционала (7.8.2) при заданных начальных значениях варьируемых функций и при условии, что конечные значения варьируемых функций $x_1, x_2,\ldots, x_n$ удовлетворяют m уравнениям связей

$$\begin{aligned} &f_1(x_1, x_2, \ldots, x_n)\big|_{t=1} = 0, \\ &f_2(x_1, x_2, \ldots, x_n)\big|_{t=1} = 0, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &f_m(x_1, x_2, \ldots, x_n)\big|_{t=1} = 0. \end{aligned} \tag{8.17}$$

Данная задача называется задачей с подвижными граничными точками. В такой постановке варьируемые функции получают большую свободу, чем в предыдущей постановке. Если на какой-либо совокупности функций достигается экстремум в задаче с подвижными граничными точками, то экстремум должен достигаться и по отношению к классу функций, имеющих те же граничные точки. Следовательно, функции, реализующие экстремум функционала с подвижными границами, должны являться решением системы дифференциальных уравнений Эйлера (8.8). Общее решение уравнений Эйлера содержат $2n$ произвольных постоянных.

Произвольные постоянные выбираются такими, чтобы обеспечить выполнение равенств (8.7). Половина из них может быть определена из условия, что при t=0 вариации искомых функций равны нулю:

$$\delta x_i\Big|_{t=0} = 0 \quad \text{или} \quad x_i(0) = x_i^{(0)}, \quad i = 1, 2, \ldots, n. \tag{8.18}$$

С учетом равенств (8.18) и уравнений Эйлера (8.8) необходимое условие (8.6) экстремума функционала (8.2) сведется к равенству

$$\left(h_{1(1)}\delta x_1 + h_{2(1)}\delta x_2 + \ldots + h_{n(1)}\delta x_n\right)\Big|_{t=1} = 0. \tag{8.19}$$

Если все вариации искомых функций в конечной точке являются независимыми и, соответственно, могут быть отличны от нуля, то для выполнения необходимого условия экстремума функционала должны быть справедливы равенства

$$\frac{\partial h}{\partial(dx_i/dt)}\Bigg|_{t=1} = 0, \quad i = 1, 2, \ldots, n. \tag{8.20}$$

Эти условия называются естественными граничными условиями. Их принимают в отсутствие уравнений связей. При наличии уравнений связей (8.17) не все вариации функции $\delta x_1, \delta x_2, \ldots, \delta x_n$ при t=1 являются независимыми, так как они связаны уравнениями

$$\begin{aligned}
&\frac{\partial f_1}{\partial x_1}\delta x_1 + \frac{\partial f_1}{\partial x_2}\delta x_2 + \ldots + \frac{\partial f_1}{\partial x_n}\delta x_n = 0,\\
&\frac{\partial f_2}{\partial x_1}\delta x_1 + \frac{\partial f_2}{\partial x_2}\delta x_2 + \ldots + \frac{\partial f_2}{\partial x_n}\delta x_n = 0,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\frac{\partial f_m}{\partial x_1}\delta x_1 + \frac{\partial f_m}{\partial x_2}\delta x_2 + \ldots + \frac{\partial f_m}{\partial x_n}\delta x_n = 0,
\end{aligned} \tag{8.21}$$

вытекающими из уравнений связи (8.17). Действительно, если точки $x_i=b_i$ и $x_i=b_i+\alpha(\delta x_i)$, $i=1,2,\ldots,n$, удовлетворяют уравнениям связей, то в пределе $\alpha\to 0$, разность уравнений (8.17) для этих точек даст уравнения (8.21). Независимыми являются только n-m вариаций δx_i . Можно найти m вариаций δx_i из системы уравнений (8.21), подставить их в (8.19) и приравнять нулю коэффициенты при остальных вариациях δx_i, $i=m+1,m+2,\ldots,n$. Но можно использовать метод неопределенных коэффициентов Лагранжа, который делает варьируемые функции равноправными. Умножим равенства (8.21) на неопределенные пока постоянные множители $\lambda_1, \lambda_2,\ldots, \lambda_m$ соответственно и добавим их к равенству (8.19). В результате получим, что для выполнения равенства (8.6) требуется выполнение равенств

$$\left(\frac{\partial h}{\partial(dx_i/dt)}+\lambda_1\frac{\partial f_1}{\partial x_i}+\lambda_2\frac{\partial f_2}{\partial x_i}+\ldots+\lambda_m\frac{\partial f_m}{\partial x_i}\right)\bigg|_{t=1}\delta x_i=0,\quad i=1,\,2,\,\ldots,\,n, \tag{8.22}$$

в конечном состоянии варьируемых функпий. Эти равенства называются общими условиями трансверсальности. Совместно с уравнениями связей (8.17) уравнения (8.22) представляют систему $n+m$ уравнений относительно $x_i|_{t=1}$, $i=1,2,\ldots,n$ и $\lambda_1, \lambda_2,\ldots, \lambda_m$. Они позволяют найти вторую половину произвольных постоянных в общем решении уравнений Эйлера.

В результате мы получили, что необходимое условие экстремума функционала (8.2) с условиями (8.17) на границе, требует выполнения уравнений Эйлера (8.8), равенств нулю вариаций начального положения (8.18), выполнения условий трансверсальности (8.22) и уравнений связей (8.17) в конечном состоянии варьируемых функций (при t=1). Другими словами, при известных начальных условиях и известных уравнениях связей для конечных значений варьируемых функций необходимое условие экстремума функционала $\Phi(h)$ на функциях $x_1(t), x_2(t),\ldots, x_n(t)$ удовлетворяющих уравнениям Эйлера, сводится к выполнению системы уравнений

$$\left(\frac{\partial h}{\partial(dx_1/dt)} + \lambda_1\frac{\partial f_1}{\partial x_1} + \lambda_2\frac{\partial f_2}{\partial x_1} + \ldots + \lambda_m\frac{\partial f_m}{\partial x_1}\right)\Bigg|_{t=1} = 0,$$
$$\left(\frac{\partial h}{\partial(dx_2/dt)} + \lambda_1\frac{\partial f_1}{\partial x_2} + \lambda_2\frac{\partial f_2}{\partial x_2} + \ldots + \lambda_m\frac{\partial f_m}{\partial x_2}\right)\Bigg|_{t=1} = 0,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\left(\frac{\partial h}{\partial(dx_n/dt)} + \lambda_1\frac{\partial f_1}{\partial x_n} + \lambda_2\frac{\partial f_2}{\partial x_n} + \ldots + \lambda_m\frac{\partial f_m}{\partial x_n}\right)\Bigg|_{t=1} = 0,$$
$$f_1(x_1, x_2, \ldots, x_n)\big|_{t=1} = 0,$$
$$f_2(x_1, x_2, \ldots, x_n)\big|_{t=1} = 0,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$f_m(x_1, x_2, \ldots, x_n)\big|_{t=1} = 0. \tag{8.23}$$

Эта система и ее решение симметричны относительно искомых функций. Функционал Ф(h) формулирует критерий поведения варьируемых функций $x_1(t), x_2(t),\ldots, x_n(t)$ при их переходе из начального состояния в состояние, удовлетворяющее наложенным на них вариационным связям (7.8.17).

Связь с экстремумом функции.

Для принятого нами критерия поведения функций $x_1(t), x_2(t),\ldots, x_n(t)$ потребуем наименьшей суммы квадратов изменений функций, которую мы опишем интегралом

$$\Phi = \frac{1}{2}\int_0^t \sum_{i=1}^{n}\left(\frac{dx_i}{dt}\right)^2 dt. \tag{8.24}$$

Подинтегральная функция в данном случае равна половине суммы квадратов производных

$$h=\frac{1}{2}\left(\left(\frac{dx_1}{dt}\right)^2+\left(\frac{dx_2}{dt}\right)^2+\ldots+\left(\frac{dx_n}{dt}\right)^2\right). \qquad (8.25)$$

В соответствии с (8.9) и (8.18) общее решение системы уравнений Эйлера для функции (8.25) имеет линейный вид

$$x_i(t)=tx_i+(1-t){x_i}^{(0)}, \quad i=1,\,2,\,\ldots,\,n, \qquad (8.26)$$

где $x_i^{(0)}= x_i(0)$— начальные значения функций, $x_i= x_i(1)$искомые конечные значения варьируемых функций. Пусть функции $x_1(t)$, $x_2(t),\ldots, x_n(t)$ при t=1 должны удовлетворять уравнениям связей (8.17).

Конечные значения функций должны быть найдены из системы уравнений (8.23), которая после подстановки (8.26) в функцию (8.25) примет вид

$$\begin{array}{c}
x_1-{x_1}^{(0)}+\lambda_1\dfrac{\partial f_1}{\partial x_1}+\lambda_2\dfrac{\partial f_2}{\partial x_1}+\ldots+\lambda_m\dfrac{\partial f_m}{\partial x_1}=0,\\
x_2-{x_2}^{(0)}+\lambda_1\dfrac{\partial f_1}{\partial x_2}+\lambda_2\dfrac{\partial f_2}{\partial x_2}+\ldots+\lambda_m\dfrac{\partial f_m}{\partial x_2}=0,\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
x_n-{x_n}^{(0)}+\lambda_1\dfrac{\partial f_1}{\partial x_n}+\lambda_2\dfrac{\partial f_2}{\partial x_n}+\ldots+\lambda_m\dfrac{\partial f_m}{\partial x_n}=0,\\
f_1(x_1,\,x_2,\,\ldots,\,x_n)=0,\\
f_2(x_1,\,x_2,\,\ldots,\,x_n)=0,\\
\cdots\cdots\cdots\cdots\cdots\cdots\\
f_m(x_1,\,x_2,\,\ldots,\,x_n)=0.
\end{array} \qquad (8.27)$$

Эта система совпадает с системой (7.21). Таким образом, критерии (7.20) и (8.24) эквивалентны, так как при одних и тех же уравнениях связей приводят к одинаковому необходимому условию экстремума.

Вариационная постановка задачи учитывает не только изменение варьируемых функций, но и способ этого изменения. Это играет роль в криволинейных координатах. Вариационная постановка задачи позволяет получить критерий минимума суммы квадратов изменений варьируемых функций в криволинейных координатах. Для двухмерных точек на криволинейной поверхности, когда координаты точек являются параметрами поверхности, необходимо использовать вариационную постановку задачи, так как в общем случае на поверхности невозможно построить двухмерную прямоугольную декартову систему координат для ее параметров.

13.9. Геодезические линии

В прямоугольной декартовой системе координат кратчайшая линия, соединяющая две точки пространства, описывается линейной функцией. Если используется криволинейная система координат, то кратчайшая линия, соединяющая две точки пространства, будет в общем случае описываться нелинейной функцией. Если пространство не является евклидовым, то эта линия не является прямой. Кратчайшая линия, соединяющая две точки пространства, называется геодезической линией. На геодезической линии величина $\int ds$, где ds— дифференциал длины дуги геодезической, принимает экстремальное значение. Уравнения геодезической линии можно найти из равенства нулю вариации $\delta(\int ds)$ ее длины дуги. Мы получим уравнения геодезической линии из равенства нулю вариации функционала

$$\Phi = \frac{1}{2}\int_0^1 \left(\frac{ds}{dt}\right)^2 dt. \qquad (9.1)$$

Пусть u^1, u^2, u^3— криволинейные координаты в трехмерном пространстве. Криволинейные координаты в силу их поведения при преобразованиях координат должны иметь верхний индекс. Рассмотрим поведение точки, координаты которой являются варьируемыми параметрами связей, в криволинейной системе координат. Пусть в исходном состоянии рассматриваемая точка имела координаты $u^i = u^{i(0)}$, i=1,2,3. После изменения уравнений связей или констант в уравнениях связей она займет новое положение в пространстве. Проведем произвольную линию $u^i(t)$ из исходного положения рассматриваемой точки в ее новое положение.

Пусть при t=0 линия проходит через исходную точку с координатами $u^i = u^{i(0)}$, а при t=1 линия проходит через новое положение точки. Найдем уравнения линии, для которой функционал (9.1) принимает минимальное значение. В декартовой прямоугольной системе координат $u^1=x$, $u^2=y$, $u^3=z$, а функционал (9.1) имеет вид

$$\Phi = \frac{1}{2}\int_0^1 \left(\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2\right)dt.$$

В криволинейной системе координат u^1, u^2, u^3 аналогичный функционал получим, подставив (1.10.5) в (9.1),

$$\Phi = \frac{1}{2}\int_0^1 g_{ij}\frac{du^i}{dt}\frac{du^j}{dt}\,dt = \frac{1}{2}\int_0^1 \left(g_{11}\left(\frac{du^1}{dt}\right)^2 + g_{22}\left(\frac{du^2}{dt}\right)^2 + g_{33}\left(\frac{du^3}{dt}\right)^2 + \right.$$
$$\left. +2g_{12}\frac{du^1}{dt}\frac{du^2}{dt} + 2g_{23}\frac{du^2}{dt}\frac{du^3}{dt} + 2g_{13}\frac{du^1}{dt}\frac{du^3}{dt}\right)dt, \quad (9.2)$$

где g_{ij}— ковариантные компоненты метрического тензора. В (9.2) используется соглашение о суммировании по повторяющимся верхним и нижним индексам. Пусть функции $u^1(t)$, $u^2(t)$, $u^3(t)$ дифференцируемы требуемое число раз. Возьмем совокупность близких к ним функций $u^i(t)+\delta u^i(t)$, $i=1,2,3$, где величины $\delta u^i(t)$ называются вариациями соответствующих функций. Эти функции можно дифференцировать по параметру t.

$$\frac{d(\delta u^i)}{dt} = \delta\left(\frac{du^i}{dt}\right), \quad \frac{d^2(\delta u^i)}{dt^2} = \delta\left(\frac{d^2u^i}{dt^2}\right), \quad \frac{d^k(\delta u^i)}{dt^k} = \delta\left(\frac{d^ku^i}{dt^k}\right).$$

Символ δ обозначает переход из какой-нибудь точки искомой линии в точку некоторой другой кривой, которой соответствует то же значение параметра t. Вариация функционала (9.1) выражается формулой

$$\delta\Phi = \frac{1}{2}\int\limits_0^1 \left(\frac{\partial g_{ij}}{\partial u^k}\frac{du^i}{dt}\frac{du^j}{dt}\delta u^k + 2g_{ij}\frac{du^i}{dt}\delta\left(\frac{du^j}{dt}\right)\right) dt, \tag{9.3}$$

где использовалась симметрия метрического тензора $g_{ij}=g_{jij}$. Проинтегрируем по частям вторую половину слагаемых в (9.3) и получим

$$\int\limits_0^1 g_{ij}\frac{du^i}{dt}\delta\left(\frac{du^j}{dt}\right) dt = g_{ij}\frac{du^i}{dt}\delta u^j\Big|_0^1 - \int\limits_0^1 \frac{d\left(g_{ij}(du^i/dt)\right)}{dt}\delta u^j\, dt =$$

$$= g_{ij} = \frac{du^i}{dt}\delta u^j\Big|_0^1 - \int\limits_0^1 \left(\frac{\partial g_{ij}}{\partial u^m}\frac{du^i}{dt}\frac{du^m}{dt} + g_{ij}\frac{d^2u^i}{dt^2}\right)\delta u^j\, dt. \tag{9.4}$$

В силу равенства нулю вариаций $\delta u^j|^1_0=0$ в концевых точках вариация функционала (9.1) выразится формулой

$$\delta\Phi = \frac{1}{2}\int\limits_0^1 \left(\frac{\partial g_{ij}}{\partial u^k}\frac{du^i}{dt}\frac{du^j}{dt} - 2\frac{\partial g_{ik}}{\partial u^m}\frac{du^i}{dt}\frac{du^m}{dt} - 2g_{ik}\frac{d^2u^i}{dt^2}\right)\delta u^k\, dt. \tag{9.5}$$

Необходимое условие экстремума функционала выражается равенством нулю его вариации. Отсюда в силу произвольности выбора δu^k следует, что функционал (9.1) достигает экстремума на кривых $u^i(t)$, i=1,2,3, удовлетворяющих дифференциальным уравнениям

$$\frac{\partial g_{ij}}{\partial u^k}\frac{du^i}{dt}\frac{du^j}{dt} - 2\frac{\partial g_{ik}}{\partial u^m}\frac{du^i}{dt}\frac{du^m}{dt} - 2g_{ik}\frac{d^2u^i}{dt^2} = 0. \tag{9.6}$$

Изменим обозначения индексов (индексы, по которым выполняется суммирование, мы имеем право обозначать любыми буквами) и преобразуем первое и второе слагаемые в левой части равенства (9.6) следующим образом:

$$\frac{\partial g_{ij}}{\partial u^k}\frac{du^i}{dt}\frac{du^j}{dt}-2\frac{\partial g_{ik}}{\partial u^m}\frac{du^i}{dt}\frac{du^m}{dt}\equiv$$
$$\equiv\frac{\partial g_{ij}}{\partial u^k}\frac{du^i}{dt}\frac{du^j}{dt}-\frac{\partial g_{ik}}{\partial u^m}\frac{du^i}{dt}\frac{du^m}{dt}-\frac{\partial g_{ik}}{\partial u^m}\frac{du^i}{dt}\frac{du^m}{dt}\equiv$$
$$\equiv\frac{\partial g_{ij}}{\partial u^k}\frac{du^i}{dt}\frac{du^j}{dt}-\frac{\partial g_{ik}}{\partial u^j}\frac{du^i}{dt}\frac{du^j}{dt}-\frac{\partial g_{jk}}{\partial u^i}\frac{du^j}{dt}\frac{du^i}{dt}=$$
$$=\left(\frac{\partial g_{ij}}{\partial u^k}-\frac{\partial g_{ik}}{\partial u^j}-\frac{\partial g_{jk}}{\partial u^i}\right)\frac{du^j}{dt}\frac{du^i}{dt}=-2\Gamma_{ij,k}\frac{du^j}{dt}\frac{du^i}{dt},$$

где $\Gamma_{ij,k}=1/2\left(\frac{\partial g_{ik}}{\partial u^j}+\frac{\partial g_{jk}}{\partial u^i}-\frac{\partial g_{ij}}{\partial u^k}\right)$ — символы Кристоффеля 1-го рода.

С учетом полученного вариация функционала (9.1) будет равна нулю при выполнении равенства

$$g_{mk}\frac{d^2u^m}{dt^2}+\Gamma_{ij,k}\frac{du^i}{dt}\frac{du^j}{dt}=0.$$

Умножив равенства (9.7) на g^{rk} выполнив суммирование и используя свойство $g^{rk}\, g_{mk}=\delta_k^r$, где δ_k^r — символы Кронекера, получим

$$\frac{d^2u^r}{dt^2}+\Gamma_{ij}{}^r\frac{du^i}{dt}\frac{du^j}{dt}=0. \tag{9.8}$$

где Γ_{ij}^r— символы Кристоффеля 2-го рода. Таким образом, геодезические линии должны удовлетворять уравнениям (9.8). Эти уравнения представляют собой уравнения Эйлера (8.8) для частного случая функционала (9.1), но в общем случае криволинейных координат. Геодезические линии являются кротчайшими линиями, соединяющими две заданные точки. В евклидовом пространстве кротчайшей линией является прямая линия. Если в евклидовом пространстве мы построим криволинейную систему координат то прямые линии в этой системе координат будут описываться уравнениями (9.8).

В частном случае декартовых прямоугольных координат Γ_{ij}^r и уравнениями геодезических линий являются интегралы уравнений d^2u^r/dt^2, которые совпадают с линейными функциями (8.26). Таким образом, если мы описываем геометрические объекты в криволинейной системе координат, то мы должны пользоваться вариационным критерием их поведения. Например, мы строим вариационные связи для двухмерных точек на поверхности, координатами которых являются параметры и и v поверхности. В этом случае для координат двухмерных точек $p_i=[u_i, v_i]^T$,i=1,2,…,n, следует использовать вариационный критерий минимума функционала

$$\Phi = \frac{1}{2}\sum_{i=0}^{n}\int_{0}^{1}\left(g_{11}\frac{du_i}{dt}\frac{du_i}{dt} + 2g_{12}\frac{du_i}{dt}\frac{dv_i}{dt} + g_{22}\frac{dv_i}{dt}\frac{dv_i}{dt}\right)^2 dt \tag{9.9}$$

с подвижными границами, удовлетворяющими уравнениям связей

$$\begin{aligned} &f_1(u_1, v_1, u_2, v_2, \ldots, u_n, v_n) = 0, \\ &f_2(u_1, v_1, u_2, v_2, \ldots, u_n, v_n) = 0, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &f_m(u_1, v_1, u_2, v_2, \ldots, u_n, v_n) = 0, \end{aligned} \tag{9.10}$$

при t=1 . Необходимыми условиями минимума функционала (9.9) является выполнение уравнений (9.8) для координат точек. Точки на поверхности при изменении параметров вариационных связей будут двигаться по геодезическим линиям поверхности.

13.10. Вариационные связи двухмерных точек

Применим метод минимизации изменений связанных параметров для вариационных связей двухмерных геометрических объектов в декартовой прямоугольной системе координат. Применение метода минимизации изменений связанных параметров для вариационных связей трехмерных геометрических объектов аналогично тому, что будет изложено ниже. Уравнения связей для трехмерных геометрических объектов приведены выше и мы не будем повторяться. Варьируемые параметры трехмерных геометрических объектов посредством свободных параметров могут быть связаны с двухмерными геометрическими объектами.

Двухмерные геометрические объекты используются для построения пространственных геометрических объектов. Многие тела строятся на основе плоских контуров путем выдавливания, вращения, движения вдоль заданной линии, плавного соединения нескольких плоских контуров. Плоские кривые и контуры строятся на двухмерных линиях. Для того чтобы перестроить пространственный объект, требуется перестроить двухмерный объект. Большие возможности в этом случае дают вариационные связи, делающие зависимыми параметры из структур данных геометрических объектов.

Для двухмерных объектов многие уравнения вариационных связей имеют тот же векторный вид, что и для пространственных объектов. Отличие от вышеизложенного состоит в том, что двухмерные точки и векторы имеют всего две координаты и, соответственно, каждая вариационная связь будет содержать меньшее число параметров. Рассмотрим некоторые двухмерные вариационные связи.

Для фиксации скалярных параметров, например, радиуса окружности или угла дуги окружности, используются фиксирующие связи типа (2.1). Для закрепления двухмерной точки применяется векторное уравнение (2.2), состоящее из двух скалярных уравнений (2.3).

Размер вдоль координаты.

Размер вдоль одной координаты между двухмерными точками $p_1=[x_1\ y_1]^T$ и $p_2=[x_2\ y_2]^T$ описывается одним из уравнений

$$x_2 - x_1 - d = 0, \quad \text{или} \quad y_2 - y_1 - d = 0, \tag{10.1}$$

где d — требуемый размер. Пусть до постановки размера положение точек описывалось радиус-векторами $p_1^{(0)}=[x_1^{(0)}\ y_1^{(0)}]^T$ и $p_2^{(0)}=[x_2^{(0)}\ y_2^{(0)}]^T$. Для того, чтобы поведение точек при установки размера было симметричным, потребуем, чтобы сумма квадратов перемещений связанных точек

$$\psi = \frac{1}{2}\left(|\boldsymbol{p}_1 - \boldsymbol{p}_1{}^{(0)}|^2 + |\boldsymbol{p}_2 - \boldsymbol{p}_2{}^{(0)}|^2\right) = \\ = \frac{1}{2}\left((x_1 - x_1{}^{(0)})^2 + (y_1 - y_1{}^{(0)})^2 + (x_2 - x_2{}^{(0)})^2 + (y_2 - y_2{}^{(0)})^2\right)$$

(10.2)

была минимальной. Для определенности будем считать, что размер поставлен вдоль координат *x*. Необходимым условием минимума функции (10.2) при условии (10.1) является равенство нулю частных производных по параметрам $x_1, y_1, x_2, y_2, \lambda$ функции

$$F \doteq \psi + \lambda f = \frac{1}{2}\left((x_1 - x_1{}^{(0)})^2 + (y_1 - y_1{}^{(0)})^2 + (x_2 - x_2{}^{(0)})^2 + (y_2 - y_2{}^{(0)})^2\right) + \\ + \lambda(x_2 - x_1 - d). \quad \textbf{(7.10.3)}$$

(10.3)

Искомые координаты точек и параметр λ найдем из системы уравнений

$$\begin{gathered} x_2 - x_1 = d, \\ x_1 - x_1{}^{(0)} - \lambda = 0, \qquad y_1 - y_1{}^{(0)} = 0, \\ x_2 - x_2{}^{(0)} + \lambda = 0, \qquad y_2 - y_2{}^{(0)} = 0. \end{gathered}$$

(10.4)

Данная система уравнений имеет решение

$$\begin{gathered} x_1 = \frac{1}{2}(x_2{}^{(0)} + x_1{}^{(0)} - d), \qquad y_1 = y_1{}^{(0)}, \\ x_2 = \frac{1}{2}(x_2{}^{(0)} + x_1{}^{(0)} + d), \qquad y_2 = y_2{}^{(0)}, \\ \lambda = \frac{1}{2}(x_2{}^{(0)} - x_1{}^{(0)} - d). \end{gathered}$$

(10.5)

Можно доказать, что найденное решение является точкой минимума функции (10.2). Перемещение точек производится вдоль координаты *x*. В данной постановке задачи точки являются равноправными. Уравнения системы и ее решение являются симметричными без использования каких-либо дополнительных ограничений относительно поведения системы точек.

Если одна из точек закреплена, то на нее наложена связь (2.2) и ее координаты не варьируются. Пусть точка p_2 закреплена. Тогда система будет иметь решение

$$x_1 = x_2{}^{(0)} - d, \quad y_1 = y_1{}^{(0)}, \quad x_2 = x_2{}^{(0)}, \quad y_2 = y_2{}^{(0)},$$
$$\lambda = x_2{}^{(0)} - x_1{}^{(0)} - d. \tag{10.0}$$

Размер вдоль направления.

Размер вдоль заданного направления между двухмерными точками $p_1=[x_1 y_1]^T$ и p_2 $[x_2, y_2]^T$ описывается уравнением (3.12) , которое в координатной записи имеет вид

$$(x_2 - x_1)m_x + (y_2 - y_1)m_y - d = 0, \tag{10.7}$$

где $m=[m_x,m_y]^T$— задающий направление двухмерный вектор единичной длины. Найдем положение точек из условия минимума их суммарного перемещения (10.2). В соответствии с методом Лагранжа составим функцию

$$F = \frac{1}{2}\Big((x_1 - x_1{}^{(0)})^2 + (y_1 - y_1{}^{(0)})^2 + (x_2 - x_2{}^{(0)})^2 + (y_2 - y_2{}^{(0)})^2\Big) + \\ + \lambda\big((x_2 - x_1)m_x + (y_2 - y_1)m_y - d\big). \tag{10.8}$$

Искомые координаты точек и параметр λ найдем из равенство нулю частных производных этой функции по параметрам $x_1, y_1, x_2, y_2, \lambda$

$$(x_2 - x_1)m_x + (y_2 - y_1)m_y - d = 0,$$
$$x_1 - x_1{}^{(0)} - \lambda m_x = 0, \qquad y_1 - y_1{}^{(0)} - \lambda m_y = 0,$$
$$x_2 - x_2{}^{(0)} + \lambda m_x = 0, \qquad y_2 - y_2{}^{(0)} + \lambda m_y = 0. \tag{10.9}$$

Данная система уравнений имеет решение

$$x_1 = x_1{}^{(0)} + \frac{m_x}{2}\left((x_2{}^{(0)} - x_1{}^{(0)})m_x + (y_2{}^{(0)} - y_1{}^{(0)})m_y - d\right),$$
$$y_1 = y_1{}^{(0)} + \frac{m_y}{2}\left((x_2{}^{(0)} - x_1{}^{(0)})m_x + (y_2{}^{(0)} - y_1{}^{(0)})m_y - d\right),$$
$$x_2 = x_2{}^{(0)} - \frac{m_x}{2}\left((x_2{}^{(0)} - x_1{}^{(0)})m_x + (y_2{}^{(0)} - y_1{}^{(0)})m_y - d\right),$$
$$y_2 = y_2{}^{(0)} - \frac{m_y}{2}\left((x_2{}^{(0)} - x_1{}^{(0)})m_x + (y_2{}^{(0)} - y_1{}^{(0)})m_y - d\right),$$
$$\lambda = \frac{1}{2}\left((x_2{}^{(0)} - x_1{}^{(0)})m_x + (y_2{}^{(0)} - y_1{}^{(0)})m_y - d\right). \tag{10.10}$$

Перемещение точек производится вдоль вектора **m** на одинаковое расстояние от исходного положения точек. Решение системы уравнений (7.10.9) является симметричным.

Линейный размер.

Линейный размер на плоскости между двумя точками $p_1=[x_1 y_1]^T$ и $p_2\ [x_2, y_2]^T$ описывается уравнением (3.1), которое в координатной записи имеет вид

$$f = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} - d = 0. \tag{10.11}$$

Для симметричного поведения точек при установки размера потребуем, чтобы суммарное перемещение связанных точек было минимальным. Необходимым условием минимума функции (10.2) при условии (10.11) является равенство нулю частных производных по параметрам функции

$$F = \frac{1}{2}\left(|p_1 - p_1{}^{(0)}|^2 + |p_2 - p_2{}^{(0)}|^2\right) + \lambda(|p_2 - p_1| - d) =$$
$$= \frac{1}{2}\left((x_1 - x_1{}^{(0)})^2 + (y_1 - y_1{}^{(0)})^2 + (x_2 - x_2{}^{(0)})^2 + (y_2 - y_2{}^{(0)})^2\right) +$$
$$+\lambda\left(\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} - d\right). \quad (7.$$

(10.12)

Система уравнений, из которой найдется искомое положение точек, аналогична системе (6.3). Решение такой системы уравнений имеет вид

$$x_1 = x_1{}^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(x_1{}^{(0)} - x_2{}^{(0)}),$$
$$y_1 = y_1{}^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(y_1{}^{(0)} - y_2{}^{(0)}),$$
$$x_2 = x_2{}^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(x_2{}^{(0)} - x_1{}^{(0)}),$$
$$y_2 = y_2{}^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(y_2{}^{(0)} - y_1{}^{(0)}),$$
$$\lambda = -\frac{1}{2}(d - d^{(0)}).$$

(10.13)

где $d^{(0)} = \sqrt{(x_2^{(0)} - x_1{}^{(0)})^2 + (y_2^{(0)} - y_1{}^{(0)})^2}$. Перемещение точек производится вдоль прямой линии, проходящей через точки и $p_1^{(0)}$ и $p_2^{(0)}$ на одинаковое расстояние.

Если точки p_1 и p_2входят в структуру данных отрезка, то с помощью связей

$$x_2 = x_1, \quad \text{или} \quad y_2 = y_1,$$

отрезок можно сделать параллельным одной из координатных осей. Пусть требуется, чтобы отрезок был параллельным оси y. Тогда из равенства нулю частных производных функции

$$F = \frac{1}{2}\left((x_1 - {x_1}^{(0)})^2 + (x_2 - {x_2}^{(0)})^2\right) + \lambda(x_2 - x_1) \tag{10.14}$$

получим систему уравнений

$$x_1 - {x_1}^{(0)} - \lambda = 0, \qquad x_2 - {x_2}^{(0)} + \lambda = 0,$$
$$x_2 - x_1 = 0.$$

Ее решение равно

$$x_1 = \frac{1}{2}({x_1}^{(0)} + {x_2}^{(0)}), \qquad x_2 = \frac{1}{2}({x_2}^{(0)} + {x_1}^{(0)}),$$
$$\lambda = \frac{1}{2}({x_2}^{(0)} - {x_1}^{(0)}). \tag{10.15}$$

Симметрия двух точек.

Рассмотрим симметрию точек $p_1=[x_1 y_1]^T$ и p_2 $[x_2, y_2]^T$ относительно прямой линии

$$l(t) = p + ti, \tag{10.16}$$

что показано на рис. 10.1. Прямую будем считать неподвижной, поэтому ее параметры варьироваться не будут.

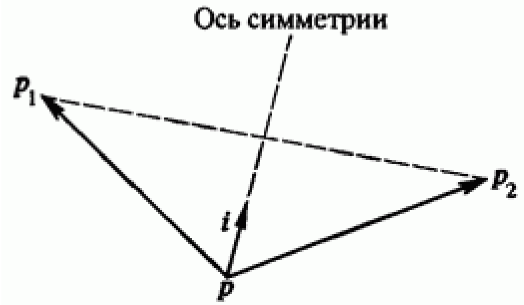


Рис. 10.1. Симметрия двухмерных точек относительно прямой

Симметрия точек относительно прямой линии описывается двумя уравнениями

$$(p_1 - p) \cdot i = (p_2 - p) \cdot i,$$
$$(p_1 - p) \cdot j + (p_2 - p) \cdot j = 0,$$

где вектор $j=[-b\ a\ ]^T$ ортогонален вектору $i=[\ a\ \ b]^T$. Пусть точка p имеет координаты x и y. Составим функцию

$$F = \frac{1}{2}\left((x_1 - x_1{}^{(0)})^2 + (y_1 - y_1{}^{(0)})^2 + (x_2 - x_2{}^{(0)})^2 + (y_2 - y_2{}^{(0)})^2\right) + \\ + \lambda_1\big((x_1 - x_2)a + (y_1 - y_2)b\big) + \lambda_2\big((y_1 + y_2 - 2y)a - (x_1 + x_2 - 2x)b\big). \quad (10.18)$$

Систему уравнений для определения положения точек получим из равенства нулю частных производных функции (10.18)

$$x_1 - x_1^{(0)} + \lambda_1 a - \lambda_2 b = 0,$$
$$y_1 - y_1^{(0)} + \lambda_1 b + \lambda_2 a = 0,$$
$$x_2 - x_2^{(0)} - \lambda_1 a - \lambda_2 b = 0,$$
$$y_2 - y_2^{(0)} - \lambda_1 b + \lambda_2 a = 0,$$
$$(x_1 - x_2)a + (y_1 - y_2)b = 0,$$
$$(y_1 + y_2 - 2y)a - (x_1 + x_2 - 2x)b = 0.$$

Решение этой системы определит новое положение точек.

Если мы хотим, чтобы точка p_1 осталась неподвижной, то введем уравнение

$$\boldsymbol{p}_1 = \boldsymbol{p}_1^{(0)} \quad (10.19)$$

или в (10.18) ее координаты будем считать константами.

Если мы хотим, чтобы точка p_2 осталась неподвижной, а точка p_1 стала ей симметрична, то введем уравнение

$$\boldsymbol{p}_2 = \boldsymbol{p}_2^{(0)} \quad (10.20)$$

или в (10.18) ее координаты будем считать неварьируемыми константами. При закреплении одной из точек решение системы уравнений (10.18) определит положение другой точки.

Угловой размер между тремя точками на плоскости.

Угловой размер на плоскости между тремя точками $p_1=[x_1, y_1]^T,$ p_2 $[x_2, y_2]^T$, $p_3=[x_3, y_3]^T$ описывается уравнением (7.3.28), где

$$s = (x_1 - x_2)(y_3 - y_2) - (y_1 - y_2)(x_3 - x_2),$$

$$c = (x_1 - x_2)(x_3 - x_2) + (y_1 - y_2)(y_3 - y_2). \tag{10.21}$$

Величина s равна произведению длин векторов p_1- p_2 и p_3- p_2 на sinα. Величина c равна произведению длин векторов p_1- p_2 и p_3- p_2 на cosα. Угловой размер на плоскости связывает шесть параметров. Составим функцию

$$F = \frac{1}{2}\left(|p_1 - \boldsymbol{p}_1{}^{(0)}|^2 + |p_2 - \boldsymbol{p}_2{}^{(0)}|^2 + |p_3 - \boldsymbol{p}_3{}^{(0)}|^2\right) + \lambda\left(\operatorname{arctg}\frac{s}{c} - \alpha\right), \tag{10.23}$$

где $p_1^{(0)}=[x_1^{(0)}\ y_1^{(0)}]^T$, $p_2^{(0)}=[x_2^{(0)}\ y_2^{(0)}]^T$, $p_3^{(0)}=[x_3^{(0)}\ y_3^{(0)}]^T$— исходные положения точек. Систему уравнений для определения параметров получим из равенства нулю частных производных функции (10.23)

$$\begin{gathered}
x_1 - x_1{}^{(0)} + \lambda\frac{c(y_3 - y_2) - s(x_3 - x_2)}{s^2 + c^2} = 0,\\
y_1 - y_1{}^{(0)} + \lambda\frac{-c(x_3 - x_2) - s(y_3 - y_2)}{s^2 + c^2} = 0,\\
x_2 - x_2{}^{(0)} + \lambda\frac{c(y_1 - y_3) - s(2x_2 - x_1 - x_3)}{s^2 + c^2} = 0,\\
y_2 - y_2{}^{(0)} + \lambda\frac{c(x_3 - x_1) - s(2y_2 - y_1 - y_3)}{s^2 + c^2} = 0,\\
x_3 - x_3{}^{(0)} + \lambda\frac{-c(y_1 - y_2) - s(x_1 - x_2)}{s^2 + c^2} = 0,\\
y_3 - y_3{}^{(0)} + \lambda\frac{c(x_1 - x_2) - s(y_1 - y_2)}{s^2 + c^2} = 0,\\
\operatorname{arctg}\left(\frac{(x_1 - x_2)(y_3 - y_2) - (x_3 - x_2)(y_1 - y_2)}{(x_1 - x_2)(x_3 - x_2) + (y_1 - y_2)(y_3 - y_2)}\right) = \alpha.
\end{gathered} \tag{10.24}$$

Решение этой системы определит положение точек при заданном угле α.

13.11. Вариационные связи двухмерных линий

Вариационные связи для нескольких двумерных точек рассмотрим в виде связей точек двухмерных кривых. С помощью вариационных связей легко управлять кривыми. Например, отрезки можно сделать ортогональными, параллельными друг другу или осям координат, окружности и сплайны можно делать сделать касательными отрезкам или друг другу.

Пусть даны два отрезка прямых

$$r_1(t_1) = (1 - t_1)p_1 + t_1 p_2,$$
$$r_3(t_3) = (1 - t_3)p_3 + t_3 p_4, \quad (11.1)$$

построенных по точкам $p_1^{(0)}=[x_1^{(0)}\ y_1^{(0)}]^T$, $p_2^{(0)}=[x_2^{(0)}\ y_2^{(0)}]^T$, $p_3^{(0)}=[x_3^{(0)}\ y_3^{(0)}]^T$, $p_4^{(0)}=[x_4^{(0)}\ y_4^{(0)}]^T$. На данные четыре точки можно наложить вариационную связь, определяющую угол между отрезками $r_1(t_1)$ и $r_3(t_3)$Эта связь аналогична угловому размеру и описывается тем же уравнением

$$\operatorname{arctg}\frac{s}{c} = \alpha, \quad (11.3)$$

где α — заданный угол. Синус и косинус угла пропорциональны величинам:

$$s = (x_2 - x_1)(y_4 - y_3) - (y_2 - y_1)(x_4 - x_3),$$
$$c = (x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3). \quad (11.4\text{-}11.5)$$

Уравнение (11.3) связывает восемь параметров.

Ортогональность отрезков.

Как частный случай связи (11.3) может рассматриваться вариационная связь устанавливающая ортогональность отрезков. Данная вариационная связь описывается уравнением

$$(p_2 - p_1) \cdot (p_4 - p_3) = (x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3) = 0. \tag{11.6}$$

Систему уравнений для определения положения точек получим из равенства нулю частных производных функции

$$F = \frac{1}{2}\left(|p_1 - p_1{}^{(0)}|^2 + |p_2 - p_2{}^{(0)}|^2 + |p_3 - p_3{}^{(0)}|^2 + |p_4 - p_4{}^{(0)}|^2\right) + \\ + \lambda\big((x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3)\big), \tag{11.7}$$

где $p_1^{(0)}=[x_1^{(0)}\ y_1^{(0)}]^T$, $p_2^{(0)}=[x_2^{(0)}\ y_2^{(0)}]^T$, $p_3^{(0)}=[x_3^{(0)}\ y_3^{(0)}]^T$, $p_4^{(0)}=[x_4^{(0)} y_4^{(0)}]^T$— исходные положения точек, в общем случае не удовлетворяющие уравнению (11.6).

Параллельность отрезков.

Другим частным случаем связи (11.3) является вариационная связь, устанавливающая параллельность отрезков. Она определяется уравнением

$$(x_2 - x_1)(y_4 - y_3) - (y_2 - y_1)(x_4 - x_3) = 0. \tag{11.8}$$

Пусть в исходном состоянии координаты точек $p_1^{(0)}=[x_1^{(0)}\ y_1^{(0)}]^T$, $p_2^{(0)}=[x_2^{(0)}\ y_2^{(0)}]^T$, $p_3^{(0)}=[x_3^{(0)}\ y_3^{(0)}]^T$, $p_4^{(0)}=[x_4^{(0)} y_4^{(0)}]^T$ не удовлетворяют уравнению (11.7). Применим метод минимизации суммарного изменения параметров. Систему уравнений для определения положения точек получим из равенства нулю частных производных функции

$$F = \frac{1}{2}\left(|p_1 - p_1{}^{(0)}|^2 + |p_2 - p_2{}^{(0)}|^2 + |p_3 - p_3{}^{(0)}|^2 + |p_4 - p_4{}^{(0)}|^2\right) + \\ + \lambda\big((x_2 - x_1)(y_4 - y_3) - (y_2 - y_1)(x_4 - x_3)\big). \quad (11.9)$$

Для простоты предположим, что точки p_2, p_3, p_4 зафиксированы и, следовательно, не изменяют своего положения. Параллельность отрезков (11.1) и (11.2) будет достигнута перемещением точки p_1. Система уравнений будет иметь вид

$$\begin{gathered} x_1 - x_1{}^{(0)} - \lambda(y_4 - y_3) = 0, \\ y_1 - y_1{}^{(0)} + \lambda(x_4 - x_3) = 0, \\ (x_2 - x_1)(y_4 - y_3) - (y_2 - y_1)(x_4 - x_3) = 0. \end{gathered} \quad (11.10)$$

Эта система имеет решение

$$\begin{gathered} x_1 = x_1{}^{(0)} + (y_4{}^{(0)} - y_3{}^{(0)})\frac{h^{(0)}}{d^{(0)}}, \\ y_1 = y_1{}^{(0)} - (x_4{}^{(0)} - x_3{}^{(0)})\frac{h^{(0)}}{d^{(0)}}, \\ \lambda = \frac{h^{(0)}}{d^{(0)}}, \end{gathered} \quad (11.11)$$

где $d^{(0)}=|\, p_4{}^{(0)}- p_3{}^{(0)}|^2$, $h^{(0)}=(x_2{}^{(0)}- x_1{}^{(0)})\,(y_4{}^{(0)}- y_3{}^{(0)})- (y_2{}^{(0)}- y_1{}^{(0)})\,(x_4{}^{(0)}- x_3{}^{(0)})$.

Перемещение точки p_1происходит по нормали к отрезку, проходящему через точки $p_4{}^{(0)}$- $p_3{}^{(0)}$ (рис. 11.1).

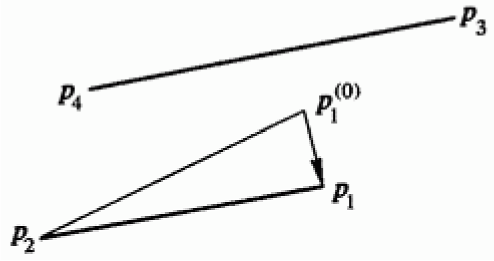


Рис. 11.1. Параллельность отрезков прямой линии

В описанном случае параллельности отрезков первый зависит от второго и подстраивается под него. Рассмотрим, как будет себя вести зависимый отрезок при вращении другого отрезка.

Пусть в исходном состоянии точки p_2 и p_4 закреплены, отрезки параллельны и их точки имеют координаты

$$x_1{}^{(0)} - x_2{}^{(0)} = b,$$
$$y_1{}^{(0)} - y_2{}^{(0)} = 0,$$
$$x_3{}^{(0)} - x_4{}^{(0)} = a,$$
$$y_4{}^{(0)} - y_3{}^{(0)} = 0.$$

Будем вращать точку p_3 вокруг точки p_4 так, чтобы их координаты изменялись по закону

$$x_3{}^{(0)} = x_4{}^{(0)} + a\cos\varphi,$$
$$y_3{}^{(0)} = y_4{}^{(0)} + a\sin\varphi. \qquad (11.12)$$

В соответствии с решением (11.11) точка p_1 будет двигаться вокруг точки p_2 так, что ее координаты будут изменяться по закону

$$x_1 = x_1{}^{(0)} - b\sin^2\varphi = x_2{}^{(0)} + b\cos^2\varphi = x_2{}^{(0)} + r\cos\varphi,$$
$$y_1 = y_1{}^{(0)} + b\sin\varphi\cos\varphi = y_2{}^{(0)} + b\cos\varphi\sin\varphi = y_2{}^{(0)} + r\sin\varphi,$$

где $r=b$ cosφ. Из формул видно, что точка p_1 будет двигаться вокруг точки p_2 по окружности, но с удвоенной частотой по сравнению с частотой вращения точки p_3 вокруг точки p_4

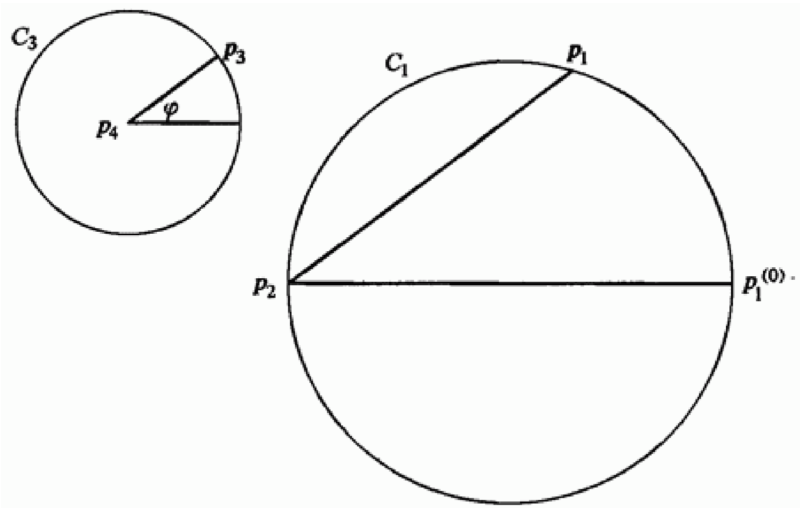


Рис. 11.2. Траектории движения отрезка p_1 p_2 параллельно зависимого от отрезка p_3 p_4 при закрепленных точках p_2, p_4

При изменении угла φ от 0 до π точка p_3 сделает половину оборота по окружности вокруг точки p_4, а точка p_1 сделает полный оборот по окружности вокруг точки p_2. Траектории движения точек p_1, p_3 показаны на рис. 11.2. Точка p_3 движется по окружности C_3, а точка p_1 движется по окружности C_1.

Касание сплайнов.

Рассмотрим вариационные связи касания линий друг друга. Пусть даны две NURBS кривые. Одна из них описывается функцией

$$r_p(t) = \frac{\sum_{i=1}^{n} N_{i,m}(t) w_i p_i}{\sum_{i=1}^{n} N_{i,m}(t) w_i}, \qquad t_{\min} \leqslant t \leqslant t_{\max}, \tag{11.13}$$

и построена на точках p_i, $i=1,2,\ldots,n$ с весами w_i, а другая описывается функцией

$$r_q(u) = \frac{\sum_{j=1}^{k} N_{j,l}(u) z_j q_j}{\sum_{j=1}^{k} N_{j,l}(u) z_j}, \qquad u_{\min} \leqslant u \leqslant u_{\max}, \tag{11.14}$$

и построена на точках q_j, $j=1,2,\ldots,k$ с весами z_j. Найдем две наиболее близкие друг к другу точки $r_p(t_0)$ и $r_q(u_0)$ на этих кривых, одна — на первой, другая — на второй, касательные в которых параллельны.

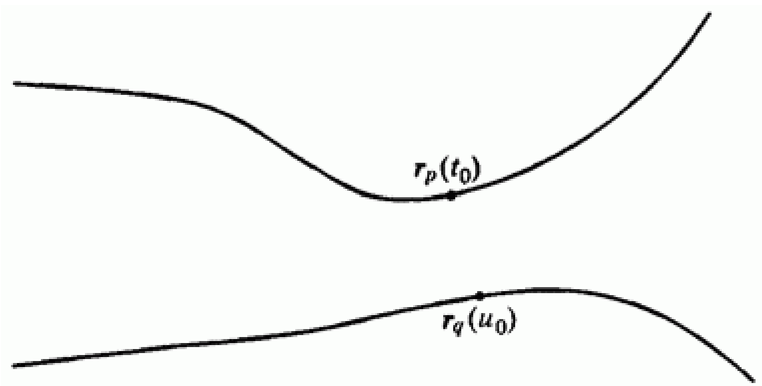


Рис. 11.3. NURBS кривые будут касаться ближайшими точками

Эти точки должны удовлетворять уравнениям имеющим вид

$$\left(r_p(t) - r_q(u)\right) \cdot \frac{dr_p}{dt} = 0,$$
$$\left(r_q(u) - r_p(t)\right) \cdot \frac{dr_q}{du} = 0.$$

Пусть этим точкам соответствуют параметры t_0 и u_0 (рис. 11.3).

Точки на кривых выражаются в виде сумм

$$r_p(t_0) = P_1p_1 + P_2p_2 + \ldots + P_np_n = \sum_{i=1}^{n} P_ip_i,$$

$$r_q(u_0) = Q_1q_1 + Q_2q_2 + \ldots + Q_kq_k = \sum_{j=1}^{k} Q_jq_j, \qquad (11.16\text{-}11.17)$$

где

$$P_i = \frac{N_{i,m}(t_0)w_i}{\sum\limits_{r=1}^{n} N_{r,m}(t_0)w_r}, \qquad Q_j = \frac{N_{j,l}(u_0)z_j}{\sum\limits_{r=1}^{k} N_{r,l}(u_0)z_r}.$$

Пусть точки имеют координаты $p_i=[x_i\ y_i]^T$, $i=1,2,\ldots,n$, $q_j=[a_j\ b_j]^T$, $j=1,2,\ldots,k$. Если бы точки $r_p(t_0)$ и $r_q(u_0)$ совпали, то являлись бы решением задачи касания кривых (11.13) и (11.14). Поэтому в качестве уравнения связи возьмем уравнение

$$|\boldsymbol{r}_p(t_0) - \boldsymbol{r}_q(u_0)| = |(P_1\boldsymbol{p}_1 + P_2\boldsymbol{p}_2 + \ldots + P_n\boldsymbol{p}_n) - (Q_1q_1 + Q_2q_2 + \ldots + Q_kq_k)| =$$
$$= \sqrt{\left(\sum_{i=1}^{n} P_i x_i - \sum_{j=1}^{k} Q_j a_j\right)^2 + \left(\sum_{i=1}^{n} P_i y_i - \sum_{j=1}^{k} Q_j b_j\right)^2} = 0. \quad (7.11.18)$$

(11.18)

Составим функцию суммарного перемещения точек касающихся кривых совместно с уравнением связи (11.18). Эта функция имеет вид

$$F = \frac{1}{2}\Big(|p_1 - p_1{}^{(0)}|^2 + |p_2 - p_2{}^{(0)}|^2 + \ldots + |p_n - p_n{}^{(0)}|^2 +$$
$$+ |q_1 - q_1{}^{(0)}|^2 + |q_2 - q_2{}^{(0)}|^2 + \ldots + |q_k - q_k{}^{(0)}|^2\Big) +$$
$$+ \lambda|\boldsymbol{r}_p(t_0) - \boldsymbol{r}_q(u_0)|.$$

(11.19)

Систему уравнений для определения положения точек получим из равенства нулю частных производных функции (11.19)

$$x_i - x_i{}^{(0)} + \lambda P_i \frac{x_p(t_0) - a_q(u_0)}{\sqrt{\big(x_p(t_0) - a_q(u_0)\big)^2 + \big(y_p(t_0) - b_q(u_0)\big)^2}} = 0, \quad i = 1, 2, \ldots, n,$$
$$y_i - y_i{}^{(0)} + \lambda P_i \frac{y_p(t_0) - b_q(u_0)}{\sqrt{\big(x_p(t_0) - a_q(u_0)\big)^2 + \big(y_p(t_0) - b_q(u_0)\big)^2}} = 0, \quad i = 1, 2, \ldots, n,$$
$$a_j - a_j{}^{(0)} - \lambda Q_j \frac{x_p(t_0) - a_q(u_0)}{\sqrt{\big(x_p(t_0) - a_q(u_0)\big)^2 + \big(y_p(t_0) - b_q(u_0)\big)^2}} = 0, \quad j = 1, 2, \ldots, k,$$
$$b_j - b_j{}^{(0)} - \lambda Q_j \frac{y_p(t_0) - b_q(u_0)}{\sqrt{\big(x_p(t_0) - a_q(u_0)\big)^2 + \big(y_p(t_0) - b_q(u_0)\big)^2}} = 0, \quad j = 1, 2, \ldots, k,$$
$$\sqrt{\big(x_p(t_0) - a_q(u_0)\big)^2 + \big(y_p(t_0) - b_q(u_0)\big)^2} = 0,$$

(11.20)

где

$$x_p(t_0) = P_1 x_1 + P_2 x_2 + \ldots + P_n x_n = \sum_{i=1}^{n} P_i x_i,$$

$$y_p(t_0) = P_1 y_1 + P_2 y_2 + \ldots + P_n y_n = \sum_{i=1}^{n} P_i y_i,$$

$$a_q(u_0) = Q_1 a_1 + Q_2 a_2 + \ldots + Q_k a_k = \sum_{j=1}^{k} Q_j a_j,$$

$$b_q(u_0) = Q_1 b_1 + Q_2 b_2 + \ldots + Q_k b_k = \sum_{j=1}^{k} Q_j b_j.$$

Система уравнений (7.11.20) на (r+1)-й итерации метода Ньютона дает приращения искомых функций, равные:

$$\Delta x_i^{(r+1)} = -\frac{P_i}{s}\left(\sum_{m=1}^{n} P_m x_m^{(r)} - \sum_{m=1}^{k} Q_m a_m^{(r)}\right), \qquad i = 1, 2, \ldots, n,$$

$$\Delta y_i^{(r+1)} = -\frac{P_i}{s}\left(\sum_{m=1}^{n} P_m y_m^{(r)} - \sum_{m=1}^{k} Q_m b_m^{(r)}\right), \qquad i = 1, 2, \ldots, n,$$

$$\Delta a_j^{(r+1)} = \frac{Q_j}{s}\left(\sum_{m=1}^{n} P_m x_m^{(r)} - \sum_{m=1}^{k} Q_m a_m^{(r)}\right), \qquad j = 1, 2, \ldots, k,$$

$$\Delta b_j^{(r+1)} = \frac{Q_j}{s}\left(\sum_{m=1}^{n} P_m y_m^{(r)} - \sum_{m=1}^{k} Q_m b_m^{(r)}\right), \qquad j = 1, 2, \ldots, k.$$

(11.21)

Значение коэффициента λ на текущей итерации равно

$$\lambda = \frac{1}{s}\sqrt{\left(\sum_{i=1}^{n} P_i x_i - \sum_{j=1}^{k} Q_j a_j\right)^2 + \left(\sum_{i=1}^{n} P_i y_i - \sum_{j=1}^{k} Q_j b_j\right)^2} = \\ = \frac{1}{s}\sqrt{\left(x_p(t_0) - a_q(u_0)\right)^2 + \left(y_p(t_0) - b_q(u_0)\right)^2}, \quad (11.22)$$

где $s = \sum_{i=1}^{n} P_i^2 + \sum_{j=1}^{k} Q_j^2$. Пока решение не найдено, коэффициент λ остается не равным нулю.

Перемещение всех точек на каждой итерации происходит параллельно касательным к кривым, проходящим через точки $r_p(t_0)$ и $r_q(u_0)$ на расстояние, пропорциональное их вкладу (коэффициенту P_i или Q_j) в точки касания. Перед началом новой итерации необходимо заново вычислить параметры t_0 и u_0 для наиболее близких друг к другу точек P_i и Q_j на рассматриваемых кривых и коэффициенты для новых ближайших точек $r_p(t_0)$ и $r_q(u_0)$.

Следует заметить, что в искомой точке выполняется равенство (11.18), что обращает в нуль числители и знаменатели последних слагаемых уравнений системы (11.20).

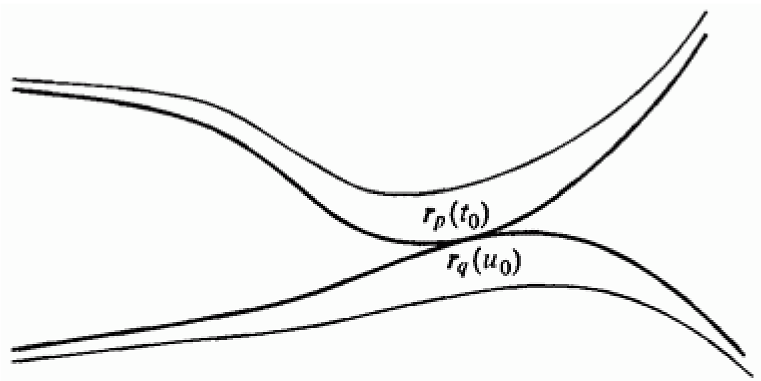


Рис. 11.4. Модификация NURBS кривых при наложении связи касания

Если на очередной r-й итерации выполнилось равенство (11.18), а решение системы нелинейных уравнений еще не закончено, то следует использовать $|\ r_p(t_0)- r_q(u_0)|^{(r-1)}$, $(x_p(t_0)- x_q(u_0))^{(r-1)}$, $(y_p(t_0)- y_q(u_0))^{(r-1)}$, вычисленные на предыдущей итерации. Поведение кривых приведено на рис. 11.4.

Касание отрезка и окружности.

Решим задачу касания в частных случаях. Касание отрезка $r_1(t)$ и окружности $r_0(u)$

$$\begin{aligned} \boldsymbol{r}_1(t) &= (1-t)\boldsymbol{p}_1 + t\boldsymbol{p}_2, \\ \boldsymbol{r}_0(u) &= \boldsymbol{p}_0 + \boldsymbol{i}_x r\cos u + \boldsymbol{i}_y r\sin u, \end{aligned} \tag{11.23}$$

описывается уравнением

$$|(1-t_0)\boldsymbol{p}_1 + t_0\boldsymbol{p}_2 - \boldsymbol{p}_0| = r, \tag{11.24}$$

где $p_1=[x_1\ y_1]^T$ $p_2=[x_2\ y_2]^T$— начальная и конечная точки отрезка, r— радиус окружности, $p_0=[x_0\ y_0]^T$ — центр окружности,

$t_0 = \dfrac{(p_0 - p_1)\bullet(p_2 - p_0)}{(p_2 - p_1)\bullet(p_2 - p_1)}$ —значение параметра отрезка,

соответствующее проекции центра окружности на отрезок. В данном случае радиус окружности также является варьируемым параметром. Параметр t_0, соответствующий проекции центра окружности на отрезок, мы будем вычислять на каждой итерации. Систему уравнений для определения параметров связи получим, приравняв нулю частные производные по параметрам функции

$$\begin{aligned} F = \frac{1}{2}\Big(|\boldsymbol{p}_1 - \boldsymbol{p}_1^{(0)}|^2 + |\boldsymbol{p}_2 - \boldsymbol{p}_2^{(0)}|^2 + |\boldsymbol{p}_0 - \boldsymbol{p}_0^{(0)}|^2 + (r - r^{(0)})^2\Big) + \\ + \lambda\big(|(1-t_0)\boldsymbol{p}_1 + t_0\boldsymbol{p}_2 - \boldsymbol{p}_0| - r\big), \end{aligned} \tag{11.25}$$

где $p_1^{(0)}$, $p_2^{(0)}$, $p_3^{(0)}$— исходные положения точек p_1, p_2, p_3, $r^{(0)}$— исходное значение радиуса окружности.

Система уравнений для определения положения точек имеет вид

$$
\begin{gathered}
x_0 - x_0^{(0)} - \lambda \frac{x_0}{\sqrt{\left((1-t_0)x_1 + t_0x_2 - x_0\right)^2 + \left((1-t_0)y_1 + t_0y_2 - y_0\right)^2}} = 0, \\
y_0 - y_0^{(0)} - \lambda \frac{y_0}{\sqrt{\left((1-t_0)x_1 + t_0x_2 - x_0\right)^2 + \left((1-t_0)y_1 + t_0y_2 - y_0\right)^2}} = 0, \\
x_1 - x_1^{(0)} + \lambda \frac{(1-t_0)x_1}{\sqrt{\left((1-t_0)x_1 + t_0x_2 - x_0\right)^2 + \left((1-t_0)y_1 + t_0y_2 - y_0\right)^2}} = 0, \\
y_1 - y_1^{(0)} + \lambda \frac{(1-t_0)y_1}{\sqrt{\left((1-t_0)x_1 + t_0x_2 - x_0\right)^2 + \left((1-t_0)y_1 + t_0y_2 - y_0\right)^2}} = 0, \\
x_2 - x_2^{(0)} + \lambda \frac{t_0x_2}{\sqrt{\left((1-t_0)x_1 + t_0x_2 - x_0\right)^2 + \left((1-t_0)y_1 + t_0y_2 - y_0\right)^2}} = 0, \\
y_2 - y_2^{(0)} + \lambda \frac{t_0y_2}{\sqrt{\left((1-t_0)x_1 + t_0x_2 - x_0\right)^2 + \left((1-t_0)y_1 + t_0y_2 - y_0\right)^2}} = 0, \\
r - r^{(0)} - \lambda = 0, \\
\sqrt{\left((1-t_0)x_1 + t_0x_2 - x_0\right)^2 + \left((1-t_0)y_1 + t_0y_2 - y_0\right)^2} - r = 0.
\end{gathered}
\tag{11.26}
$$

Система уравнений (11.26) решается итерационно. Перед началом новой итерации необходимо заново вычислить параметр t_0. Поведение отрезка и окружности приведено на рис. 11.5.

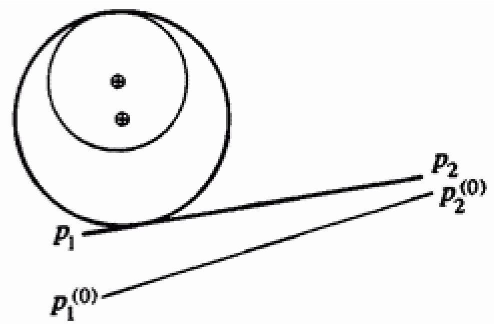


Рис. 11.5. Поведение окружности и отрезка при касании

Касание окружностей.

Касание двух окружностей

$$r_1(t_1) = q_1 + i_x r_1 \cos t_1 + i_y r_1 \sin t_1,$$
$$r_2(t_2) = q_2 + i_x r_2 \cos t_2 + i_y r_2 \sin t_2, \quad (11.27\text{-}11.28)$$

описывается уравнением

$$|q_1 - q_2| = |r_1 \pm r_2|. \quad (11.29)$$

В данном случае радиусы окружностей также являются варьируемыми параметрами. Уравнение (11.29) связывает шесть параметров.

Систему уравнений для определения параметров связи получим, приравняв нулю частные производные по параметрам функции

$$F = \frac{1}{2}\left(|q_1 - q_1{}^{(0)}|^2 + |q_2 - q_2{}^{(0)}|^2 + (r_1 - r_1{}^{(0)})^2 + (r_2 - r_2{}^{(0)})^2\right) + \\ + \lambda(|q_1 - q_2| - |r_1 \pm r_2|),$$
(11.30)

где $q_1{}^{(0)}$, $q_2{}^{(0)}$— исходные положения центров, $r_1{}^{(0)}$, $r_2{}^{(0)}$— исходное значение радиусов окружностей. Знак ± должен быть раскрыт в зависимости от положения ближайших точек окружностей. Эти точки лежат на линии, соединяющей их центры. Пусть окружности расположены вне друг друга. Тогда система уравнений для определения положения центров окружностей и их радиусов имеет вид

$$x_1 - x_1{}^{(0)} + \lambda\frac{x_1 - x_2}{d} = 0, \qquad y_1 - y_1{}^{(0)} + \lambda\frac{y_1 - y_2}{d} = 0,$$
$$x_2 - x_2{}^{(0)} - \lambda\frac{x_1 - x_2}{d} = 0, \qquad y_2 - y_2{}^{(0)} - \lambda\frac{y_1 - y_2}{d} = 0,$$
$$r_1 - r_1{}^{(0)} - \lambda = 0, \quad r_2 - r_2{}^{(0)} - \lambda = 0, \quad |q_1 - q_2| - (r_1 + r_2) = 0,$$

(11.31)

где $d = |q_1 - q_2|$. Эта система имеет решение

$$x_1 = x_1{}^{(0)} - \frac{1}{4}a^{(0)}\frac{x_1{}^{(0)} - x_2{}^{(0)}}{d^{(0)}}, \quad y_1 = y_1{}^{(0)} - \frac{1}{4}a^{(0)}\frac{y_1{}^{(0)} - y_2{}^{(0)}}{d^{(0)}},$$

$$x_2 = x_2{}^{(0)} + \frac{1}{4}a^{(0)}\frac{x_1{}^{(0)} - x_2{}^{(0)}}{d^{(0)}}, \quad y_2 = y_2{}^{(0)} + \frac{1}{4}a^{(0)}\frac{y_1{}^{(0)} - y_2{}^{(0)}}{d^{(0)}},$$

$$r_1 = r_1{}^{(0)} + \frac{1}{4}a^{(0)}, \quad r_2 = r_2{}^{(0)} + \frac{1}{4}a^{(0)}, \quad \lambda = \frac{1}{4}a^{(0)},$$

(11.32)

где $d^{(0)} = |q_1{}^{(0)} - q_2{}^{(0)}|$, $a^{(0)} = |q_1{}^{(0)} - q_2{}^{(0)}| - (r_1{}^{(0)} - r_2{}^{(0)})$. Из решения следует, что радиусы изменяются на одинаковую величину, равную перемещению центров окружностей.

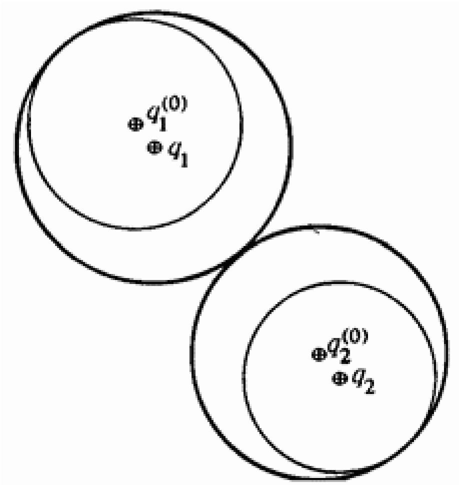


Рис. 11.6. Касание окружностей

Поведение окружностей приведено на рис. 11.6. При расположении окружностей одна внутри другой в уравнениях (11.31) изменяется знак перед одним из радиусов.

13.12. Формирование и решение системы уравнений связей

Вариационные связи являются мощным средством управления геометрическими объектами. Они позволяют редактировать геометрические объекты совместно, когда изменение параметра одного из объектов влечет за собой соответствующие изменения параметров других объектов. Параметром может служить любая величина из структуры данных геометрического объекта или специально введенная переменная. Специально введенные переменные используются для удобства редактирования геометрических объектов с помощью вариационных связей.

С помощью вариационных связей осуществляется управление сразу всеми связанными объектами. Как правило, на одни и те же параметры наложено несколько вариационных связей. Каждая связь представляет собой некоторый вариационный объект, который изменяет параметры геометрических объектов. Различные типы вариационных связей геометрических объектов приведены на рис. 12.1. Каждая связь имеет одно или несколько уравнений и перечень параметров, участвующих в каждом уравнении.

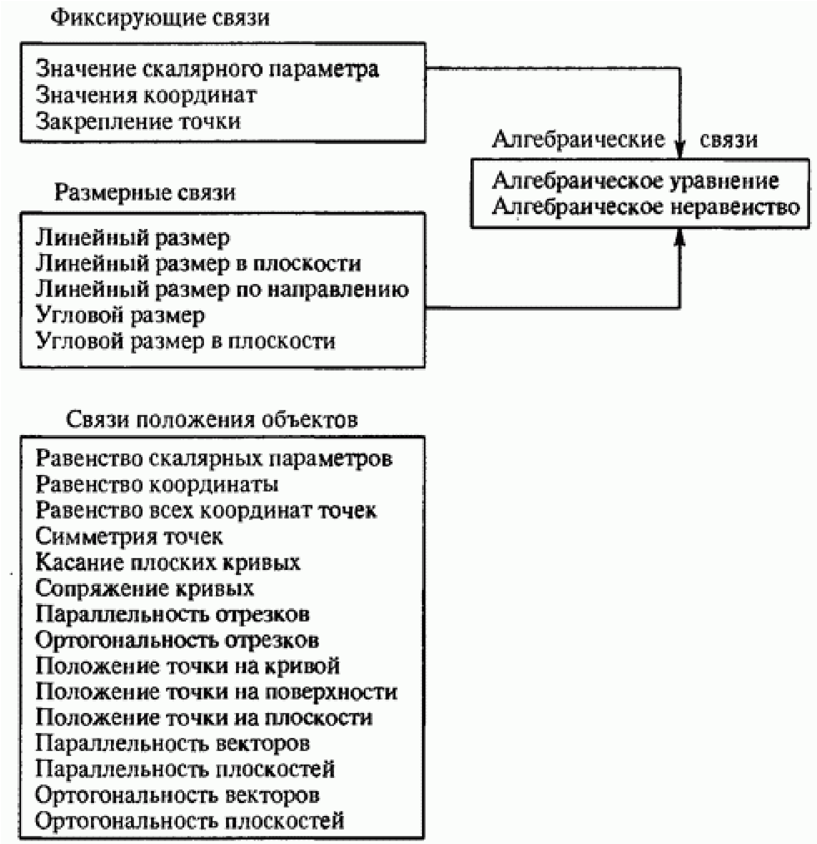


Рис. 12.1. Типы вариационных связей

Нормальным состоянием для вариационных связей является состояние, когда их уравнения выполняются. Будем называть это состояние равновесным. В процессе моделирования приходится модифицировать геометрические объекты или их взаимное положение. Таким образом, вариационные связи выводятся из состояния равновесия.

В результате все или часть уравнений вариационных связей перестает выполняться. Далее следует найти новые значения варьируемых параметров, чтобы уравнения связей опять выполнялись. Для того чтобы связанные параметры удовлетворяли уравнениям связей, требуется сформировать систему уравнений и решить ее. В процессе решения этой системы уравнений необходимо модифицировать все геометрические объекты, параметры которых изменились. Рассмотрим, каким образом удовлетворяются вариационные связи.

Пусть задан набор вариационных связей, приводящий к системе m уравнений

$$
\begin{aligned}
&f_1(x_1, x_2, \ldots, x_n) = 0, \\
&f_2(x_1, x_2, \ldots, x_n) = 0, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&f_m(x_1, x_2, \ldots, x_n) = 0,
\end{aligned}
\qquad (12.1)
$$

для $n \geq m$ изменяемых параметров $x_1, x_2, \ldots, x_n$ геометрических объектов. Число уравнений связей, как правило, меньше числа связанных параметров. Для того, чтобы решить систему уравнений связей, требуется сформулировать критерий поведения связанных геометрических объектов, который бы позволил решить систему уравнений связей. Простым и эффективным критерием поведения параметров является требование минимума суммы квадратов изменений параметров. Этот критерий выражается минимумом функции

$$
\psi(x_1, x_2, \ldots, x_n) = \frac{1}{2}\left((x_1 - x_1{}^{(0)})^2 + (x_2 - x_2{}^{(0)})^2 + \ldots + (x_n - x_n{}^{(0)})^2\right).
$$

(7.12

$$(12.2)$$

Поиск минимума функции (12.2) при условии выполнения связей (12.1) приводит к задаче нахождения минимума функции

$$
\begin{aligned}
F(x_1, x_2, \ldots, x_n) = \frac{1}{2}\left((x_1 - x_1{}^{(0)})^2 + (x_2 - x_2{}^{(0)})^2 + \ldots + (x_n - x_n{}^{(0)})^2\right) + \\
+ \lambda_1 f_1(x_1, x_2, \ldots, x_n) + \lambda_2 f_2(x_1, x_2, \ldots, x_n) + \ldots + \lambda_m f_m(x_1, x_2, \ldots, x_n).
\end{aligned}
\qquad (12.3)
$$

Для выполнения необходимого условия минимума указанной функции требуется наряду с (12.1) выполнение уравнений

$$
\begin{array}{l}
x_1 - x_1{}^{(0)} + \lambda_1 \dfrac{\partial f_1}{\partial x_1} + \lambda_2 \dfrac{\partial f_2}{\partial x_1} + \ldots + \lambda_m \dfrac{\partial f_m}{\partial x_1} = 0, \\
x_2 - x_2{}^{(0)} + \lambda_1 \dfrac{\partial f_1}{\partial x_2} + \lambda_2 \dfrac{\partial f_2}{\partial x_2} + \ldots + \lambda_m \dfrac{\partial f_m}{\partial x_2} = 0, \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
x_n - x_n{}^{(0)} + \lambda_1 \dfrac{\partial f_1}{\partial x_n} + \lambda_2 \dfrac{\partial f_2}{\partial x_n} + \ldots + \lambda_m \dfrac{\partial f_m}{\partial x_n} = 0,
\end{array}
\tag{12.4}
$$

где $\lambda_1, \lambda_2,\ldots, \lambda_m$ подлежащие определению совместно с параметрами $x_1, x_2,\ldots, x_n$ коэффициенты. В конечном итоге мы пришли к системе $n+m$ уравнений (12.1) и (12.4) для $n+m$ неизвестных $x_1, x_2,\ldots, x_n$, $\lambda_1, \lambda_2,\ldots, \lambda_m$. В общем случае эта система решается одним из итерационных методов.

Если на какие-либо из параметров наложены фиксирующие связи (2.1), то их можно исключить из функции (12.2) и при решении общей системы уравнений считать константами.

В большинстве случаев управление данными происходит следующим образом. На некотором этапе система уравнений связей (12.1) выполняется. Далее одному или нескольким параметрам присваивается новое значение, и эти параметры фиксируются. Например, мы переместили некоторую точку геометрического объекта и хотим узнать, как это повлияет на другие варьируемые параметры. Управление данными геометрических объектов может происходить путем изменения констант уравнений, что также выведет систему уравнений связей из состояния равновесия. Например, мы изменили значение некоторого размера. В результате для исходных значений параметров некоторые уравнения связей перестанут выполняться. Необходимо найти новые значения параметров, удовлетворяющие системе уравнений связей, и перестроить геометрические объекты. Система вариационных связей выходят из равновесия также при добавлении к ней новых связей. Например, мы хотим, чтобы некоторые две плоскости геометрической модели были параллельными. При этом рождается новая вариационная связь, которая добавляет в систему уравнений свои уравнения. Нужно решить систему уравнений связей и перестроить геометрические объекты.

На практике решение системы уравнений является самой сложной и тонкой проблемой при наложении вариационных связей. **В разных случаях для решения системы уравнений используются разные подходы: аналитический, конструктивный, численный.**

Аналитический подход используется для двухмерных вариационных связей, когда искомое положение двухмерных геометрических объектов можно найти с помощью «линейки и компаса». В некоторых простых случаях аналитический подход применим и для трехмерных объектов.

При наложении вариационных связей на положения твердых тел относительно друг друга эффективным является конструктивный подход. **Конструктивный подход привлекает теорию графов**. Для системы уравнений вариационных связей строится граф. **Узлами графа являются геометрические объекты, а ребрами графа являются уравнения связей.** На основе графа геометрические объекты и вариационные связи делятся на **кластеры (группы).** Кластер в модели ведет себя, как нечто жесткое целое. **В результате задача сводится к задаче позиционирования геометрических объектов кластеров.** Сначала определяется удовлетворяющее вариационным связям положение геометрических объектов для каждого кластера в отдельности, а затем кластеры последовательно стыкуются между собой. Размещение геометрических объектов внутри одного кластера осуществляется путем решения соответствующей системы уравнений. Стыковка кластеров между собой выполняется на основе анализа степеней свободы. **Стыковка кластеров производится последовательно.** Будем считать, что положение геометрических объектов некоторого кластера задано. Расположим геометрические объекты соседнего кластера так, чтобы удовлетворялись связывающие их уравнения. Тем самым мы состыкуем два кластера. Продолжим стыковку кластеров до тех пор, пока не будут выполняться все уравнения связей. В процессе стыковки кластеров возможно потребуется перебирать различные варианты и численно решать систему уравнений связей нескольких кластеров. Конструктивный подход уменьшает размерность системы уравнений, но его можно использовать для строго определенного набора геометрических объектов и вариационных связей.

Численный подход заключается в решении системы уравнений вариационных связей численными итерационными методами. При использовании этого подхода следует помнить, что вариационные связи могут иметь несколько решений. Численный метод позволяет найти одно из этих решений, и оно может не совпадать с искомым. Найденное решение зависит от начального приближения, с которого начинается итерационный процесс. **Чтобы начальное приближение находилось в области сходимости к искомому решению, следует не допускать сильных отклонений вариационных связей от их равновесных состояний.** Например, если вариационные связи выведены из состояния равновесия путем перемещения некоторой точки геометрического объекта, то решение следует искать не для окончательного положения точки, а для нескольких промежуточных ее положений, постепенно сдвигая точку от исходного положения к конечному. Аналогично, при изменении значения некоторого размера следует последовательно искать решения для нескольких промежуточных значений этого размера, постепенно переходя от старого к новому значению. Численный подход чаще используется для решения двухмерных задач.

При использовании численного подхода желательно, чтобы ненулевые приращения параметров были одного порядка. Например, если одна часть варьируемых параметров является угловыми величинами, другая часть — компонентами векторов, а третья часть параметров — координатами точек, то желательно, чтобы их значения лежали в одних и тех же пределах. Для достижения этого используется нормирование варьируемых параметров.

Численное решение системы уравнений связей может производиться несколько иначе. Используемый критерий поведения геометрических объектов стремится сохранить параметры геометрических объектов ближе к их исходному состоянию. Воспользуемся этим свойством для уменьшения размерности системы уравнений. Попытаемся решить систему уравнений, зафиксировав еще некоторые из варьируемых параметров, кроме тех, которые зафиксированы уравнениями связей. Если нам это удастся, то мы будем придерживаться правила, которое заключается в сохранении объектов как модно ближе к их исходному состоянию, и уменьшим число неизвестных и уравнений в (12.4). Остается выяснить, какие из параметров можно дополнительно зафиксировать. Некоторые из уравнений связей могут быть удовлетворены при исходных значениях параметров. Определим эти уравнения и составим список варьируемых параметров, участвующих в них. Попробуем решить полную систему уравнений, считая параметры этого списка фиксированными. Если это удается, попробуем расширить этот список, если нет, то будем сужать этот список до тех пор, пока не удастся решить полную систему уравнений. Таким образом, нам, возможно, удастся найти новые значения параметров, сохранив некоторым из них исходные значения. Если даже при отсутствии дополнительной фиксации параметров систему уравнений (7.12.1) и (7.12.4) решить не удастся, то следует вернуться к исходному равновесному состоянию геометрических объектов и констант уравнений, при которых уравнения связей удовлетворяются.

В процессе решения уравнения связей изменяют значения параметров и тем самым перестраивают геометрические объекты в соответствии с новыми значениями параметров. Объекты будут перестроены в соответствии с наложенными связями, когда все уравнения полной системы будут удовлетворены с требуемой точностью.

14. Нечеткие пространства

Геометрия нечетких множеств

14.1. Проблема развития и контроля визуального мышления

Визуальные средства компьютерной графики дают основание некоторым авторам утверждать о появлении новой инженерной дисциплины – визуальной науки. Не умаляя актуальность ее становления, ее необходимости для профессионального применения, ее многочисленных достоинств и ее значимости в развитии научной культуры, необходимо объективно оценивать и ее роль в развитии пространственного фактора интеллекта или визуального мышления человека. Исследований на эту тему еще мало, чтобы строить заключения, но тенденция к снижению уровня визуального мышления специалистов заметна. С другой стороны, умения формировать и превращать мысленные образа играет ключевую роль почти во всех сферах деятельности человека. При этом лучшего инструмента для развития визуального мышления, чем геометрия, особенно конструктивная геометрия, пока не найдено. Лучшего метода контроля уровня развитию визуального мышления, чем метод тестирования, тоже пока не создано. Таким образом, существует проблема создания эффективного метода развития и контроля пространственного фактора интеллекта, метода, адаптированного к современным компьютерным технологиям и таким, которые могут быть применены в автоматизированных интеллектуальных системах обучения. Для решения этой проблемы необходимо найти формальную модель визуального мышления и описать ее доступными математическими средствами. Другими словами, ставится задача найти соответствие между геометрическими образами как объектами некоторого реального пространства и мысленных геометрических образов пространства образного мышления. Объекты реального пространства имеют измерения и контролируемые характеристики и могут быть описаны аналитически, изображены графически, преобразованы в другие и т.д. Они могут быть отнесены к ***трем классам объектов*** – объекты-фигуры, объекта-условия (инцидентности, аффинные, метрические, дифференциальные и др.), объекты-преобразования и объекты-алгоритмы. Объекты образного мышления имеют принципиальное свойство исчезать при любой попытке их описания с помощью первых. Т.е. мысленные образы сохраняются только в пространстве воображения. Уместно в этой связи вспомнить Я. Штейнера, который читал свои лекции слушателям в полной темноте, чтобы развить в них пространственное мышление.

Прежде чем предложить один из возможных подходов к решению этой проблемы, рассмотрим простую планиметрическую задачу. Пусть нужно ответить на следующие вопросы, не пользуясь ни аналитическими, ни конструктивными моделями:

1. Каковы координаты центра окружности, проходящего через точки (10, 30), (50, 10) и (60, 100)?

2. Чему равен радиус этой окружности?

3. Пересекает ли эта окружность ось абсцисс? Ось ординат?

Совсем очевидно, что существует только один способ решения задачи – представить декартову систему координат и эту окружность мысленно. Можно придумать еще много подобных задач, но становится очевидно, что от человека даже с очень развитым визуальным мышлением нельзя получить определенных и четких ответов. Ответы на эти и другие вопросы будут принадлежать к классу нечетких, т.е. таких: «Центр окружности находится приблизительно в точке (50, 55)», «Радиус окружности – около 45», «Окружность, возможно, пересекает ось ординат, а ось абсцисс, скорее всего, нет» и т.п. Человек с неразвитым визуальным мышлением ответить на эти вопросы вообще не сможет. Отсюда следует вывод: *в пространстве образного мышления человек оперирует нечеткими образами*. Таким образом, мы приходим к выводу о возможностях использования нечетких геометрических образов как формальных моделей мысленных образов и процесса визуального мышления. Для реализации этой возможности необходимо:

- определить понятие нечеткой точки, нечеткой фигуры, нечеткого условия, нечеткого преобразования и нечеткого алгоритма;

- определить понятие сложности визуального решения задачи;

- выстроить иерархию задач по сложности;

- предложить алгоритм генерации нечетких образов для каждой задачи;

- создать возможность адаптации алгоритма к различным уровням визуального мышления.

Теория нечетких множеств и нечеткой логики широко используется при решении разных слабоформализованных задач. Согласно теореме, доказанной Б. Коско в 1983 году, любая математическая система может быть апроксимована системой, которая основана на нечеткой логике. Исходя из этой универсальности отмечается значительное расширение математических основ и прикладной части теории нечетких множеств. Понятие нечеткости позволяет «удвоить» математику. Заменяя обычные множества нечеткими, можно каждому математическому термину поставить в соответствие нечеткий аналог. Рассматривают, например, нечеткие классификации, упорядочения, логики, теоремы, алгоритмы, правила принятия решений и т.п. .

Как мы уже отмечали, процесс развития теории нечетких множеств начался с появлением в 1965 году в журнале «Информация и управления» статьи американского ученого Лотфи А. Заде, специалиста по теории управления сложными системами, где впервые появилось понятие нечеткого множества.

В 1970-и года были развиты понятия лингвистической переменной. Э. Мамдани сформулировал основные идеи нечетких регуляторов. В 1978 г. Заде предложил вариант вычисления неопределенностей, которое опирается на неаддитивную меру возможности, т.е. на интерпретацию нечеткого множества как функции распределения возможностей. В 1979 г. он же ввел теорию приближенных размышлений.

Наиболее значимыми из работ в области развития теории нечетких множеств есть публикации Заде, Дюбуа, Прада, Сугено и др. Исторически первыми появились нечеткие экспертные системы. Исследованиям в этой области посвященные работы А.Н. Борисова, Д.А. Поспелова, А. Кофмана и др.

За период существования теории нечетких множеств этой теме посвященные тысячи книг и статей, появилось новое направление в математической кибернетике - теория нечеткости, выходит международный журнал «Нечеткие множества и системы», проводятся конференции.

В области технологий можно отметить появление принципиально новых средств слабоструктуризированной, фрагментарной, неполной, нечеткой информации. К таким средствам можно отнести технологию информационного мониторинга сложных проблем и процессов. Использование такого рода инструментария показало их эффективность и огромный потенциал для задач бизнеса, политологии, социологии и т.п. Появился новый класс адаптивных нечетких моделей. Исследованием в этой области посвященные работы Ф. Херрери, О. Границы, Р. Янга, В.В. Круглова и др.

Однако в данное время не существует общепризнанной геометрической теории нечетких множеств: изображений нечетких точек, прямых, пространств, функций, а также способов решения метрических и позиционных задач, геометрических прикладных алгоритмов в вычислительной геометрии и т.д. В литературе встречаются некоторые примеры изображения нечетких объектов. Так нечеткие множества можно изобразить с помощью кругов Эйлера с размытыми границами. Ниспадающая к краям насыщенность тона передает значение функции принадлежности - чем она больше, тем тон более насыщен.

А.И. Орлов приводит пример нечеткого аналога теоремы о том, что три медианы треугольника пересекаются в одной точке. Она звучит таким образом. Пусть *AB, BP, CA* - приблизительно прямые линии, которые образуют приблизительно треугольник с вершинами *A, B, C.* Пусть *L, M, N* - приблизительно середины сторон треугольника. Тогда приблизительно прямые линии – приблизительно медианы - образуют приблизительно треугольник, который более или менее мал в сравнении с треугольником *ABC.*

Эта формулировка становится понятной только после того, как будет определено содержание слов «приблизительно» и «более или менее». Вот как можно уточнить понятие «приблизительно отрезок *АВ»:* под ним можно понимать кое-какую кривую линию, которая проходит через точки *А, В,* такую, что расстояние (в обычном смысле) от любой точки кривой к отрезку *АВ* имела относительно длины *АВ*.

Приведенную теорему можно еще более «размыть», если сказать, что приблизительно прямые линии образуют приблизительно треугольник, вершины которого находятся приблизительно в точках *А, В, С*.

14.2. Основные понятия и определения

Остановимся еще раз на понятии нечеткости и нечеткого множества. Нечеткое множество является формализацией нечеткой информации, необходимой для построения математических моделей. В основе этого понятия лежит представление о том, что элементы, которые составляют данное множество, обладающие общим свойством, могут иметь это свойство в разной степени и, следовательно, принадлежать к данному множеству с разной степенью. При таком подходе высказывания типа "такой-то элемент принадлежит данному множеству" теряют смысл, поскольку необходимо указать "насколько сильно" или с какой степенью конкретный элемент удовлетворяет свойствам данного множества. Имея в виде геометрическую или конструктивную сторону проблемы, отметим, что целью данного модуля является изучение именно конструктивных свойств тех геометрических образов, которые могут быть с некоторой степенью сопоставлены с нечеткими множествами. Учитывая тот факт, что нечеткие множества изучаются и используются в основном алгоритмическими методами, получим возможность изучать их и отношение между ними конструктивно.

Не претендуя на общность утверждения, выдвигается гипотеза, что конструктивные модели систем могут быть реализованы, если будут получены геометрические образы нечетких объектов, нечетких соответствий между ними и нечеткие условия их определения.

В этой связи рассмотрим некоторые нечеткие геометрические образы. Естественно, что любой нечеткий геометрический образ можно определить как множество четких точек или как множество четких геометрических объектов такой же структуры, в определенном смысле близких к идеальному. Исключение составляет объект - нечеткая точка, которая определяется как множество точек пространства, каждой из которых приписано некоторое значение функции принадлежности этому множеству. Не определяя здесь понятие нечеткого числа, можно утверждать, что нечеткая точка есть упорядоченное множество чисел (параметров), среди которых хотя бы одно есть нечетким. Одним из геометрических образов нечеткой точки может быть отсек конической или пирамидальной поверхности соответствующей размерности.

Сложнее определить геометрический образ, который отвечает понятию нечеткой прямой. Во-первых, нечеткую прямую можно определить как линейное множество нечетких точек, которая зависит от одного четко определенного параметра. Во-вторых, нечеткая прямая может быть представлена как множество четких точек, которая зависит от одного нечеткого параметра. В-третьих, нечеткая прямая может быть множеством нечетких точек, которые линейно зависят от одного нечеткого параметра. Все эти определения являются частными случаями. В-четвертых, нечеткая прямая есть множество четких прямых, каждой из которых приписано значение функции принадлежности данному множеству. Во всех этих случаях выходят разные геометрические образы. Очевидно, что такие определения можно обобщить на линейные пространства любой размерности, поскольку упомянутые выше системы чаще всего есть многопараметрическими.

Нечетким соответствием (преобразованием) назовем некоторое правило или алгоритм, который четко определенному геометрическому объекту - прообразу ставит в соответствие один или несколько нечетко определенных образов. Если соответствие определяется своими параметрами, то оно будет нечетким, если хотя бы один его параметр является нечетким числом. Если соответствие определятся заданием необходимого числа прообразов и образов, то оно будет нечетким, если хотя бы один из заданных образов или прообразов является нечетким.

Нечеткие условия, как и четкие, можно разделить на условия полной и неполной инцидентности, аффинные условия и условия метрические. Если в простейшее условие принадлежности точки некоторому геометрическому объекту ввести условие нечеткости, то получим семь вариантов нечеткого условия полной инцидентности. То же самое относится к условию пересечения (неполной инцидентности), аффинным и метрическим условиям. Нечеткие условия можно определить с помощью соответствующей функции, которая принимает значение равное единице, в случае выполнения данного условия, и значение равное нулю, при ее невыполнении.

Все это можно сформулировать в терминах теории параметризации: любой многопараметрический объект будет нечетким, если хотя бы один из его параметров описывается нечетким числом.

Воспользуемся двумя известными понятиями - понятием многомерного аффинного пространства и понятиям расстояния между множествами. Первое является множеством всех элементов ($a_1,\dots, a_n$), которые называют точками, причем все координаты $a_i \in R$, $i=1, \dots, n$. Предположим, что какая-нибудь одна координата является нечетким числом $\tilde{a}_i$. Это означает, что нам неизвестно ее точное значение, можно только сказать, что значение координаты приблизительно равно такому-то числу или координата принимает значение «возле» такого-то числа. Другими словами, для каждого значения $a_i \in R$, $i=1, \dots, n$ существует значение функции принадлежности, которая формализует нечеткие высказывания «приблизительно», «возле» и т.п. Легко представить себе, что все координаты точки являются нечеткими. Тогда множество всех элементов ($\tilde{a}_1 \dots, \tilde{a}_n$), называемых по аналогии нечеткими точками, образует нечеткое пространство.

Определение 1. Нечеткой k-мерной точкой является множество ($a_1,\dots, a_n$) чисел, из которых k, $1 \le k \le n$, любых, являются нечеткими.

Если $k = 0$, то такая точка является четкой. Если $k=n$, то такую точку назовем нечеткой точкой и будем обозначать *А*. Во всех других случаях нечеткую точку будем обозначать $\tilde{A}$, не указывая, если в этом нет необходимости, какие именно координаты являются нечеткими.

Определение 2. Нечетким k-мерным пространством является множество всех элементов ($a_1,\dots, a_n$), среди которых хотя бы один элемент является нечеткой k -мерной точкой.

Нечеткое k -мерное аффиное пространство будем обозначать $\tilde{\mathbf{A}}^k$.

Предложение 1. Нечеткая k-мерная точка выделяет в k-мерном пространстве нечеткое k -мерное подпространство.

Доказательство не требуется, так как каждая нечеткая координата точки задается нечетким числом, для которого всегда будет определена числовая ось, множество определений и множество значений функции принадлежности. Поэтому пространство, натянутое на множество нечетких осей, будет нечетким пространством. Пусть будут заданы две точки

$$\tilde{A}^{p+1}\ (a_1, \dots, \tilde{a}_i, \tilde{a}_{i+1}, \dots, \tilde{a}_{i+p}, a_{i+p+1}, \dots, a_n)$$

и

$$\tilde{B}^{q+1}\ (b_1,\dots, \tilde{b}_j, \tilde{b}_{j+1},\dots, \tilde{b}_{j+q}, b_{j+q+1},\dots, b_n),$$

причем $\tilde{A}^{p+1} \in \tilde{\mathbf{A}}^n$, $\widetilde{B}^{q+1} \in \tilde{\mathbf{A}}^n$. Тогда в n-мерном пространстве эти две нечеткие точки выделят нечеткое подпространство $\tilde{\mathbf{A}}^{p+q-n+2}$. Доведение Доказательство не требуется.

Определение 3. Две нечеткие точки являются тождественными или совпадают, если все их координаты - четкие и нечеткие совпадают.

Другими словами, расстояние между двумя совпадающими нечеткими точами равно нулю.

Определение 4. Две нечеткие точки являются различными, если они отличаются значением хотя бы одной - четкой или нечеткой координатой.

В классической теории множеств непустое подмножество A из универсального множества X однозначно определяется характеристическим функционалом

$$I_A(x) = \begin{cases} 1, x \in A \\ 0, x \notin A. \end{cases}$$

Т.е. подмножество A определяется как совокупность объектов, которые имеют некоторое общее свойство, наличие или отсутствие которого в любом элементе множества задается характеристическим функционалом.

Для нечеткого подмножества, которое является расширением понятия множества в классическом смысле, на пространстве объектов $X=\{x\}$ вводится уже не функционал, а характеристическая функция, которая задает для всех элементов степень наличия в них некоторого свойства, по которому они относятся к подмножеству A. Эта характеристическая функция традиционно носит название функция принадлежности. Функция принадлежности $\mu_{A:}\ X \to [0,1]$ ставит в соответствие каждому элементу подмножества число $\mu_A(x)$ из интервала [0, 1] и является непрерывной одномодальной функцией.

Определение 5. Нечеткой k-мерной точкой является выпуклое подмножество множества $X_1 \times ... \times X_n$, для каждой точки которого существует значение функции принадлежности. В точках, которые лежат на границе подмножества функция принадлежности равно нулю. Значение функции принадлежности, равное единице, существует только в одной точке подмножества, называемого ядром нечеткой точки.

Напомним, что фигурой называется любое множество точек. Поэтому нечеткая точка – это фигура, на основе которой можно построить все другие нечеткие образа. В этой связи дадим еще несколько определений.

Интервальной точою $\overline{X}$ назовем *n*-плексное множество $(\overline{x}_1,...,\overline{x}_n)$ точек, в котором $x_i \in \overline{x}_1 = [\ x_{i,1},\ x_{i,1}]$, $x_{i,j} \in R$. К изображению интервальных точек есть два естественных подхода. Первый заключается в представлении интервальной точки в виде прямоугольной области $\overline{X}_1 \times ... \times \overline{X}_n$ (прямоугольного *n*-плекса). Второй – в представлении интервальной точки в виде гиперсферы (сферического *n*-плекса). На рис.1 изображены интервальные точки для случая *n* = 2.

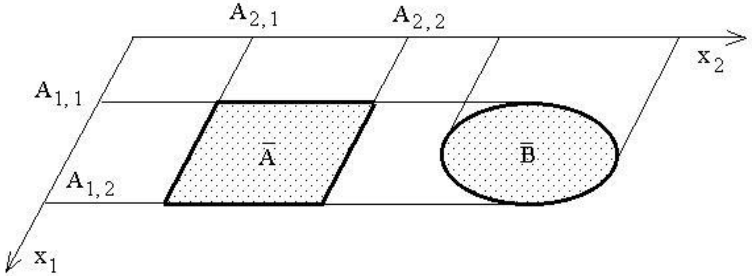


Рис. 1. Изображение интервальных точек для *n*=2

Нечеткой точкой $\widetilde{X}$ назовем *n*-плексное множество $(\widetilde{x}_1,...,\widetilde{x}_n)$ точек, которое удовлетворяет следующим условиям:

$$\widetilde{X} = \{ X ; X \in X_1 \times ... \times X_n \},$$
$$\forall X \exists \mu_{\overline{X}}(X) \in [0,1],$$

$\mu_{\overline{X}} = 0$ для всех точек $X \in \overline{X}$ и на границе интервальной точки $\overline{X}$;

$\mu_{\overline{X}} = 1$ только для одной точки области $\overline{X}$, которую называют ядром.

Пространство, в котором вместе с четкими точками существуют интервальные и нечеткие точки, назовем *n*-мерным континуумом. Можно принять как аксиому предложение, что любая точка континуума может быть ядром некоторой нечеткой точки и что для любого заданного ядра нечеткая точка единственная. К изображению нечетких точек есть тоже два подхода. Учитывая то, что функция принадлежности нечеткой точки имеет П-подобную (кругоподобную) форму, которая в простейшем виде может быть треугольной, первый подход заключается в представлении нечеткой точки в виде носителя – прямоугольной пирамиды с основанием $X_1 \times \ldots \times X_n$ и единичной высотой. Второй - в представлении нечеткой точки в виде конуса единичной высоты с основанием (носителем) в виде гиперсферы (рис. 2).

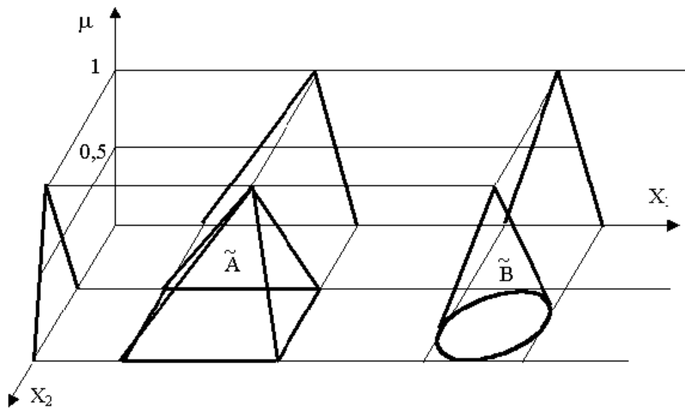


Рис. 2. Изображение прямоугольных и круговых нечетких точек двумерного континуума

Важным свойством нечеткой точки, которое отличает ее от любого другого выпуклого подмножества, является возможность описания ее с помощью альфа-уровней. Каждому альфа-уровню соответствует постоянное значение функции принадлежности. Каждый альфа-уровень является тоже выпуклым подмножеством. Нулевой альфа-уровень совпадает с границей нечеткой точки, а единичный - с ядром. Следовательно, нечеткая точка является естественная формализованная модель нечетких лингвистических понятий «возле» или «приблизительно».

Для любых двух точек континуума можно постулировать наличие двух видов отношений: четкого и нечеткого.

$$R = \{=, \neq\}, \widetilde{R} = \{=, \neq, \approx\}.$$

Т.е.

$$\forall A, \forall B : (x_a = x_b \wedge y_a = y_b) \Leftrightarrow A = B$$

$$\forall A, \forall B : (x_a \neq x_b \vee y_a \neq y_b) \Leftrightarrow A \neq B$$

$$\forall A, \forall B\ \exists \widetilde{A}, \exists \widetilde{B} : (A \in \widetilde{B} \vee B \in \widetilde{A}) \Leftrightarrow A \approx B$$

$$\forall \widetilde{A}, \forall \widetilde{B}, \forall X : (\mu_{\widetilde{A}}(X) = \mu_{\widetilde{B}}(X) \Leftrightarrow \widetilde{A} = \widetilde{B}$$

$$\forall \widetilde{A}, \forall \widetilde{B}, \forall X : (\mu_{\widetilde{A}}(X) \neq \mu_{\widetilde{B}}(X) \Leftrightarrow \widetilde{A} \neq \widetilde{B}.$$

Для отношения $A \cong B$ можно определить функцию принадлежности следующим образом

$$\mu_{A \approx B} = min\,\{\mu_{\widetilde{A}}(B), \mu_{\widetilde{B}}(A)\}.$$

Отношение $\widetilde{A} \cong \widetilde{B}$ рассматривать не будем, так как оно есть нечетким в квадрате.

Отношение $\widetilde{A} \neq \widetilde{B}$ порождает новый вопрос: до какой степени эти две нечеткие точки различны? Степень их различия обозначим символом δ. Тогда

$$\delta = \begin{cases} 0 : \forall X \Rightarrow \mu_{\widetilde{A}}(X) = \mu_{\widetilde{B}}(X) \\ ]0,1[: \exists X \in A, \mu_{\widetilde{B}}(X) \neq 0 \\ 1 : \forall X \in \widetilde{A}, \forall Y \in \widetilde{B} \Rightarrow \mu_{\widetilde{A}}(Y) = \mu_{\widetilde{B}}(X) = 0. \end{cases}$$

Степень их совпадения в этом случае равна 1-δ.

Естественно, что любой нечеткий геометрический образ можно определить как множество четких точек или как множество четких геометрических объектов такой же структуры, в определенном смысле близких к идеальному. Исключение составляет объект - нечеткая точка, которая определяется как множество точек пространства, каждой из которых приписано некоторое значение функции принадлежности этому множеству. Не определяя здесь понятие нечеткого числа, можно утверждать, что нечеткая точка есть упорядоченое множество чисел (параметров), среди которых хотя бы одно есть нечетким.

14.3. Нечеткие прямые и отношения

Предположим, что задано две нечеткие точки $\tilde{A}$, $\tilde{B}$, $\tilde{A} \neq \tilde{B}$. Они должны определить нечеткое множество, которое можно назвать нечетким отрезком. В общем случае две нечеткие точки определяют два нечетких отрезка. Первый получается как множество всех точек, которые принадлежат выпуклой оболочке, натянутой на заданные нечеткие точки.

Определение 6. Нечетким отрезком называется множество всех четких отрезков, концы которых с некоторой степенью принадлежат заданным нечетким точкам:

$$\left\{[AB] \mid A \in \tilde{A}, B \in \tilde{B}, 0 \le \mu_{\tilde{A}}(A) \le 1, 0 \le \mu_{\tilde{B}}(B) \le 1\right\}.$$

Пусть для любых двух точек значения $\mu_{\tilde{A}}(A), \mu_{\tilde{B}}(B)$ известны. Тогда значение функции принадлежности четкого отрезка нечеткому можно определить по формуле

$$\mu_{[\tilde{A}\tilde{B}]}([AB]) = \min\left\{\mu_{\tilde{A}}(A), \mu_{\tilde{B}}(B)\right\}.$$

Определение 7. Носителем нечеткого отрезка является четкое подмножество пространства, элементы которого имеют ненулевые степени принадлежности:

$$\sup\ [\tilde{A}\tilde{B}] = \left\{[AB] \middle| \mu_{[\tilde{A}\tilde{B}]}[AB] > 0\right\}.$$

Границей носителя нечеткого отрезка есть выпуклая оболочка множества $\tilde{A} \cup \tilde{B}$.

Определение 8. Ядром нечеткого отрезка является четкое подмножество пространства, элементы которого имеют степени принадлежности равные единице:

$$\text{core}\ [\tilde{A}\tilde{B}] = \left\{[AB] \middle| \mu_{[\tilde{A}\tilde{B}]}[AB] = 1\right\}.$$

Т.е. ядром нечеткого отрезка является четкий отрезок, который является выпуклой оболочкой двух точек - ядер соответствующих нечетких точек.

Рассмотрим вопрос о том, какими могут быть нечеткие отрезки. Это зависит от того, какими преобразованиями f пространства нечеткая точка $\tilde{A}$ может быть преобразована в нечеткую точку $\tilde{B}$. Пусть $\tilde{B} = f(\tilde{A})$. Тогда две нечеткие точки $\tilde{A} \neq \tilde{B}$ могут быть:

а) конгруэнтными, которые связаны параллельным переносом;

б) гомотетичными, которые связаны гомотетиею;

в) проективными;
г) общего вида, связанные взаимно-однозначными нелинейными преобразованиями.

Для первых трех случаев будем считать обязательным выполнение условия:

$$\forall \left(A \in \widetilde{A}, B \in \widetilde{B}\right) \mu_{\widetilde{A}}(A) = \mu_{\widetilde{B}}(B), \text{ если } B = f(A).$$

Конгруэнтные нечеткие точки связаны соотношениями

$$\widetilde{b}_i = \widetilde{a}_i + t_i, i = \overline{1, n}.$$

Здесь $\widetilde{a}_i$, $\widetilde{b}_i$ - проекции носителя на соответствующую ось координат.

Гомотетичные нечеткие точки связаны соотношениями

$$\widetilde{b}_i = k_i \cdot \widetilde{a}_i + t_i, i = \overline{1, n}.$$

Проективные нечеткие точки связаны соотношениями

$$\widetilde{b}_i = \frac{k_{1,i} \cdot \widetilde{a}_i + t_{1,i}}{k_{2,i} \cdot \widetilde{a}_i + t_{2,i}}, i = \overline{1, n}.$$

Поэтому нечеткие отрезки могут быть отрезками с конгруэнтными концами, отрезками с гомотетичными концами, отрезками с проективными концами и отрезками общего вида. Кроме этих отрезков учтем нечеткий отрезок, у которого один конец есть четкая точка. Во всех случаях нечеткий отрезок есть однопараметрическое линейное множество нечетких точек, которое обладает альфами-уровнями - выпуклыми подмножествами множества нечетких точек.

Введем теперь в континуум четкую прямую a. Появляются отношения принадлежности или инцидентности:

$$A \in a, A \notin a, A \tilde{\in} a, a \supset \widetilde{A}.$$

Если первые два понятны, то вторые два требуют объяснения. Так

$$\forall A, \forall a, \exists \widetilde{A} : (\exists X \in \widetilde{A}, X \in a) \Leftrightarrow (\widetilde{A} \cap a \neq 0) \Leftrightarrow A \tilde{\in} a,$$

$$(A \tilde{\in} a) \equiv (a \supset \widetilde{A}).$$

Степень принадлежности

$$\mu_{\widetilde{A}}(A \tilde{\in} a) = max\, \mu_{\widetilde{A}}(X \in a).$$

Определим нечеткую прямую. Пусть дано две нечетких точки $\widetilde{A}, \widetilde{B}$. Тогда нечеткая прямая есть множество всех точек, которые приблизительно линейно зависят от несовпадающих точек $X \in \widetilde{A}, Y \in \widetilde{B}$. Другое определение нечеткой прямой может быть следующим. Нечеткая прямая есть множество всех прямых, которые проходят через данные нечеткие точки. Исследование свойств нечеткой прямой заняло бы много места, поэтому ограничимся следующими утверждениями:

1) функция принадлежности для нечеткой прямой может быть представлена в виде альфа-уровней;

2) функция принадлежности является кусочно-квадратической, если функции принадлежности образующих точек являются кусочно-линейными;

3) ядром нечеткой прямой является единственная четкая прямая, которая инцидентна ядрам образующих ее точек;

4) граница нечеткой прямой определяется четырьмя опорными прямыми образующих ее точек.

Нечеткая прямая общего вида изображена на рис.3.

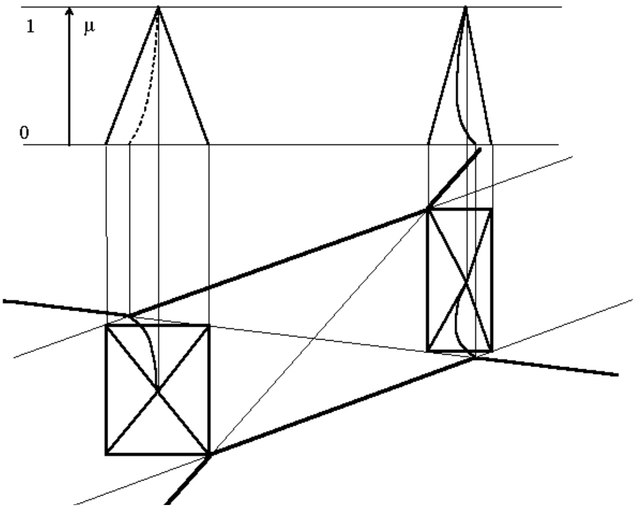


Рис.3. Нечеткая прямая общего вида

Рассмотрим следующие отношения:

$$A \in \widetilde{a}, A \notin \widetilde{a}, \widetilde{A} \subset \widetilde{a}, b \subset \widetilde{a}, b \not\subset \widetilde{a}, b \cong a, \widetilde{b} = \widetilde{a}, \widetilde{b} \subset \widetilde{a}.$$

Очевидно, что

$$\forall A, \forall \tilde{a} : \mu_{\tilde{a}}(A) \neq 0 \Leftrightarrow A \in \tilde{a},$$

$$\forall A, \forall \tilde{a} : \mu_{\tilde{a}}(A) = 0 \Leftrightarrow A \notin \tilde{a}.$$

Однако

$$\forall A, \forall \tilde{a}, \forall X : \mu_{\tilde{A}}(X) > 0, \mu_{\tilde{A}}(X) \leq \mu_{\tilde{a}}(X) \Leftrightarrow \tilde{A} \subset \tilde{a},$$
$$\forall A, \forall \tilde{a}, \forall X : \mu_{\tilde{A}}(X) \neq \mu_{\tilde{a}}(X) \Leftrightarrow \tilde{A} \not\subset \tilde{a}.$$

Что касается отношений четких и нечетких прямых, то

$$\forall \tilde{a}, \forall b, \forall X \in b : \mu_{\tilde{a}}(X) > 0 \Leftrightarrow b \subset \tilde{a},$$
$$\forall \tilde{a}, \forall b : \exists Y \in b \wedge \mu_{\tilde{a}}(Y) = 0 \Leftrightarrow b \not\subset \tilde{a}.$$

Нечеткое отношение равенства четких прямых («приблизительно совпадают») определим так:

$$\forall a, \forall b, \forall X \in b : \exists \tilde{X} \supset X, \exists Y \in a, \mu_{\tilde{X}}(Y) > 0 \Leftrightarrow a \approx b.$$

И, наконец,

$$\forall \tilde{a}, \forall \tilde{b}, \forall X : \mu_{\tilde{b}}(X) \leq \mu_{\tilde{a}}(X) \Leftrightarrow \tilde{b} \subset \tilde{a},$$
$$\forall \tilde{a}, \forall \tilde{b}, \forall X : \mu_{\tilde{b}}(X) = \mu_{\tilde{a}}(X) \Leftrightarrow \tilde{b} = \tilde{a}.$$

Рассмотрим формализацию понятий «возле», «приблизительно» и других, которые имеют такой же смысл. Интуитивно понятно, что аффинное пространство есть как бы удвоенным, поскольку в нем на равных правах присутствуют четкие и нечеткие точки. Не принимая во внимание конкретный вид функции принадлежности для нечетких точек, учтем только, что для данного универсуума X_i функции принадлежности для каждой точки должны быть однотипными. Для разных универсуумов функции принадлежности могут быть разными. Охарактеризуем каждую функцию принадлежности значением в ядре нечеткой точки и значениями нижней и верхней границы носителя нечеткой точки:

$$\widetilde{A}_i \in X_i,\ \widetilde{A}_i = \left\{\text{core}\ \widetilde{A}_i, \left|\text{core}\ \widetilde{A}_i - \sup_-\ \widetilde{A}_i\right|, \left|\sup_+\ \widetilde{A}_i - \text{core}\ \widetilde{A}_i\right|\right\}.$$

Другими словами, любая нечеткая точка в первом приближении считается точкой *(R — L)*-типа. Правая *R* и левая *L* функции - это невозрастающие функции на множестве неотрицательных действительных чисел. В простейшем случае они могут быть:

1) постоянными на всем универсууме X_i ;
2) линейными;
3) квадратичными (параболическими).

Какими они есть в каждом конкретном случае, решают эксперты. Например, если сравнить понятие «возле единицы» и «около 10^6», которые принадлежат одному универсууму, то понятно, что эти понятия не равны. Действительно, первое характеризуется нечетким числом {1, 0.1, 0.1}, а второе — нечетким числом {10, 5×10, 5×10}. Промежуточные понятия могут подчиняться линейной или квадратичной зависимости. Использовать зависимости более высоких степеней теоретически можно, а практически - нецелесообразно, поскольку тяжело представить нечеткое понятие, у которого правая и левая функции в середине диапазона принимали бы большие значения, чем на концах.

В случае множества независимых универсуумов такие зависимости строятся по каждому универсууму. В случае, когда понятие «возле» строится на множестве взаимно зависимых универсуумів, левые и правые функции каждой нечеткой точки есть функции с множеством аргументов.

Предположим, что рассматривается множество таких точек на области параметров, которая ограничена полиэдром. Представим модель искомого понятия как образа полиэдров в изопараметрических соответствиях. Любая точка, которая характеризует понятие «возле», может быть представленная векторной функцией $\bar{r} = \bar{r}(\bar{x}_1,...,\bar{x}_n,t)$, где сумма $\bar{x}_i$ равна единице. При *t=const* будет получено множество точек изопараметрического гиперпересечения. Образы полиэдров представляют собой отсеки криволинейной *(n*—2) поверхности, которые обладают такими же топологическими свойствами, как и отсек *(n* — 2)-плоскости. Но в законе их образования принимают участие криволинейные симплексы. В частности, можно предположить, что вершины образа и прообраза совпадают. Криволинейный *(n* — *k* — 1) - симплекс можно считать образующим, а криволинейные *(k* — 1) -

симплексы - направляющими, причем вершины первого описывают вторые. Справедливым будет и обратное утверждение.

Построение образа опишем в простейшем варианте. Во-первых, предположим, что криволинейные $(n — k — 1)$ - симплексы построены и имеют вид

$$\bar{r}_i(0, ..., 0, \bar{u}_i, 0, ..., 0, \bar{u}_{k+1}, ..., \bar{u}_n).$$

Криволинейные $(k— 1)$ -симплексы также построены в виде

$$\bar{r}_j(\bar{u}_1, ..., \bar{u}_k, 0, ..., 0, \bar{u}_j, 0, ..., 0).$$

Во-вторых, предположим, что закон образования линейный. Тогда образ можно представить как сумму $R_1 + R_2 — R_3$, где R_1 - $(n - 2)$ - мерная поверхность с образующими, которые являются линейными $(n — k — 1)$-симплексами, и направляющими $\bar{r}_i$; R_2-$(n—2)$-мерная криволинейная поверхность с образующими, которые являются линейными $(k — 1)$-симплексами, и направляющими $\bar{r}_i$; R_3 - отсек $(n — 2)$ -плоскости.

Это означает, что

$$R_1 = \sum_{j=k+1}^{n} u_j \cdot r_j, R_2 = \sum_{i=1}^{k} u_i \cdot r_i,$$

$$R_3 = \sum_{i=1}^{k} u_i (\sum_{j=k+1}^{n-k} u_j \cdot r(0,...,0,u_i,0,...,0,u_j,0,...,0))$$

Обобщение в направлении усложнения закона образования очевидно.

$$R_1 = \sum_{j=k+1}^{n} f_j(u_{k+1},..., u_n) \cdot r_j,$$

$$R_2 = \sum_{i=1}^{k} f_i(u_1,..., u_k) \cdot r_i,$$

$$R_3 = \sum_{i=1}^{k} f_i(u_1,..., u_k)(\sum_{j=k+1}^{n-k} f_j(u_{k+1},..., u_n) \cdot r),$$

где

$$f_i(\bar{u}_1, ..., \bar{u}_k) = 1$$

при

$$\overline{u}_i = 1,\ f_i(\overline{u}_1, ..., \overline{u}_k) = 0$$

при

$$\overline{u}_i = 0,\ f_j(\overline{u}_{k+1}, ..., \overline{u}_n) = 1$$

при

$$\overline{u}_j = 1,\ f_j(\overline{u}_{k+1}, ..., \overline{u}_n) = 0$$

при

$$\overline{u}_j = 0.$$

Между двумя симплексами - линейным и нелинейным всегда существует некоторая симплексное соответствие, которое переводит каждую вершину одного в соответствующую вершину другого. Построение такого соответствия осуществляется просто и потому здесь достаточно учесть ее существование. Соответствие между всеми точками пространства, ограниченное линейным симплексом, и всеми точками пространства, ограниченное нелинейным симплексом, есть изопараметрическим, взаимно однозначным для данной области пространства и нелинейным. Его построение основывается на следующем предложении.

Изопараметрическре отображение точек линейного n-симплекса на точки нелинейного n-симплекса без учета закона образования осуществляется векторными функциями вида

$$r = \sum_{j=1}^{n} (-1)^{j+1} \left(\sum_{i=1}^{\binom{n+1}{n-k+1}} r_i^{(n-j)} \right),$$

где r есть вектор-функция точек нелинейного симплекса, которые принадлежат $(n — j)$ -мерной нелинейной грани.

Доказательство основывается на том факте, что вектор любой точки, которая принадлежит $(n — j)$ -мерной нелинейной грани, должен входить в r только один раз. При этом r_i можно не учитывать, так как конкретный вид вектор-функции r_i не играет никакой роли. Поэтому доказательство будет корректным, даже если r_i описывает $(n — j)$-плоскость. Пусть n-симплекс имеет $n + 1$ вершин, а $(n — j)$-плоскость определяется $n — j + 1$ вершиной. Разность между ними равна j точкам. Очевидно, что ответ на вопрос сколько раз $(n — j)$-плоскость будет учтена в r равносильна ответу на вопрос сколько плоскостей, размерности большей и равной $n — j$, будет определяться $(n — j)$-плоскостью и вершинами j, которые остались. Легко показать, что гиперплоскость будет определяться j — 1 точами и $(n — j)$ -плоскостью, $(n$ — 2) -плоскость будет определяться j — 2 точками и $(n — j)$ -плоскостью и так далее. Следовательно, искомое число будет результатом суммы

$$(-1)^{a+1}\sum_{a=1}^{j}\binom{j}{j-a},$$

которая равна единице при любых значениях j.

Определение 9. Нечеткая прямая есть множество таких четких отрезков, которые приблизительно принадлежат некоторой прямой, и таких, что конец каждой предыдущей и начало следующего образцования совпадают.

Определение 10. Непрерывное множество нечетких точек, для которого существует такая четкая прямая, каждая точка которой принадлежит какой-либо нечеткой точке этого множества, называется нечеткой прямой.

Это определение можно распространить на любую нечеткую плоскую фигуру. Например, таким образом.

Определение 11. Нечеткой плоской фигурой (квадратом, прямоугольником, треугольником, кругом и т.д.) называется непрерывное множество нечетких точек, для которого существует такая четкая фигура, каждая точка которого принадлежит какой-нибудь нечеткой точке этого множества.

Теперь ясно, как на основе этого определения сформулировать понятие нечеткого тетраэдра, нечеткого k-симплекса, нечеткого полиэдра и политопа и др.

Определение 12. Нечетким k-симплексом называется непрерывное множество нечетких точек, для которого существует такой k -симплекс, каждая точка которого принадлежит какой-нибудь нечеткой точке этого множества.

Это определение можно обобщить на k -мерную нечеткую плоскость.

Определение 13. Нечеткой k-мерной плоскостью называется непрерывное множество нечетких точек, для которого существует такая четкая k -плоскость, каждая точка которой принадлежит какой-нибудь нечеткой точке этого множества.

Определение 14. Нечеткой k-мерной поверхностью называется непрерывное множество нечетких точек, для которого существует такая четкая k-поверхность, каждая точка которой принадлежит какой-нибудь нечеткой точке этого множества.

14.4. Основные условия

Рассмотрим сначала нечеткие условия принадлежности.

Точка пространства принадлежит данной нечеткой точке, если функция принадлежности в ней, которая определена относительно данной нечеткой точки, не равна нулю. Точка C принадлежит нечеткому отрезку [$\tilde{A}\tilde{B}$], если $\mu_{[\tilde{A}\tilde{B}]}(C) \neq 0$. Можно утверждать, что $C \in$ core [$\tilde{A}\tilde{B}$] .

Определение 15. Две и более точки пространства приблизительно совпадают, если они принадлежат одной и той же нечеткой точке.

Другими словами, две и более точки пространства приблизительно совпадают, если обнаружится такая нечеткая точка, относительно которой функции принадлежности всех данных точек не равны нулю.

Рассмотрим несколько задач.

Пусть в данном пространстве формализовано понятия «возле», т.е. в любом месте пространства может быть построена нечеткая точка $\tilde{A}$. Это означает, что определен носитель нечеткой точки и функции принадлежности. Ядром может быть лютая точка пространства. Пусть задано множество B_1, B_2,..., B_m точек пространства - туча точек, для которой известно, что они приблизительно совпадают. Нужно определить степень их совпадения.

Поскольку они приблизительно совпадают, то все они должны принадлежать заданной нечеткой точке $\tilde{A}$. Однако таких нечетких точек существует бесконечное множество, размерность которого совпадает с размерностью заданного пространства. Шевелением ядра нечеткой точки $\tilde{A}$ можно найти такое ее положение, при котором будет выполнено условие

$$\min\{\mu_{\tilde{A}}(B_i)\} \to \max, i = \overline{1, m}.$$

После этого понятия «приблизительно совпадают» для заданных точек B_1, B_2,..., B_m может быть выраженное числом - значением функции принадлежности

$$\mu_{\tilde{A}}\left(\{B_1, B_2, ..., B_m\} \subset \tilde{A}\right) = \frac{1}{m}\sum_{i=1}^{m}\mu_{\tilde{A}}(B_i).$$

Другая задача. Пусть задано множество точек B_1, B_2,..., B_m пространства, для которых неизвестно, что они приблизительно совпадают. Пусть в пространстве определена нечеткая точка. Нужно определить те точки множества, которые приблизительно совпадают и степень их совпадения.

Шевелением ядра нечеткой точки добиваемся выполнения двух условий:

число точек $s \to \max$ при

$$1 \le s \le m, \{B_1, B_2, ..., B_s\} \subset \widetilde{A},$$

$$\min\{\mu_{\widetilde{A}}(B_i)\} \to \max, i = \overline{1, s}.$$

После этого понятие «приблизительно совпадают» может быть выражено значениям функции принадлежности

$$\mu_{\widetilde{A}}\left(\{B_1, B_2, ..., B_s\} \subset \widetilde{A}\right) = \frac{1}{s}\sum_{i=1}^{s}\mu_{\widetilde{A}}(B_i).$$

В дальнейшем слова «приблизительно совпадают», «приблизительно принадлежит» и т.п. будем опускать и пользоваться просто сроками «принадлежит» и другими, помня о том, что имеется в виду нечеткая принадлежность.

Определение 16. Точка принадлежит нечеткой прямой, если она принадлежит какой-нибудь нечеткой точке этой прямой.

Может произойти так, что точка будет принадлежать сразу нескольким нечетким точкам данной нечеткой прямой. Выбор значения функции принадлежности точки прямой по условиям максимума из функций принадлежности этим точкам будет неверный. Тогда выбрав две ближайшие нечеткие точки из тех, которым принадлежит данная точка, и, учитывая то, что они образуют нечеткий отрезок, значение функции принадлежности прямой определяется по значению функции принадлежности этому отрезку. Это согласуется с определением 16, поскольку на нечетком отрезке всегда найдется такая нечеткая точка, которой и будет принадлежать данная точка.

Определение 17. Точка принадлежит нечеткой *k*-плоскости, если она принадлежит какой-нибудь нечеткой точке этой *k* -плоскости.

Чтобы определить значение функции принадлежности в этом случае, необходимо из множества нечетких точек, которые образуют нечеткую *k*-плоскость, выделить *k* ближайших и найти функцию принадлежности данной точки относительно *k*-мерного отсека данной нечеткой плоскости - *k*-плоскости.

Определение 18. Нечеткая *k* -плоскость принадлежит нечеткой *m* -плоскости, $k<m$, если каждая нечеткая точка *k* -плоскости принадлежит нечеткой *m* -плоскости.

Другими словами, если в нечеткой *k* -плоскости обнаружится такая четкая *k* -плоскость, которая будет принадлежать некоторой четкой *m* -

плоскости, которая принадлежит нечеткой k-плоскости, то условие принадлежности будет выполнено. Естественно, значение функции принадлежности не будет равно единице.

Напомним, что в каждой точке пространства понятие «возле» определено единственным образом. Поэтому, если ядра нечетких подпространств взаимно принадлежат друг другу, тот и сами подпространства принадлежат друг другу.

Рассмотрим еще несколько задач.

Пусть на нечеткой плоскости $\widetilde{A}^2$ задано k нечетких точек $\widetilde{A}_i$, $i = \overline{1,k}$. Найти образ - нечеткую прямую или нечеткую конику, которые определяют эти точки. Найти значение функции принадлежности образа заданном нечеткими точками.

Напомним, что любые две несовпадающие нечеткие точки задают множество отрезков или нечеткий отрезок. В этом множестве только в одного отрезка значения функции принадлежности будет равно единице, а именно у того, который задан ядрами нечетких точек. Три и более нечеткие точки могут не определять прямую, если они заданы в довольно общем положении. Но могут и определять. Критерием этого могут служить следующие правила.

Предположим, что все заданные нечеткие точки упорядочены по возрастанию какой-нибудь координаты их ядер. Пусть первая точка имеет минимальное значение этой координаты, а k-я точка - максимальное. Тогда, если все промежуточные нечеткие точки с той или другой, но не нулевой, степенью принадлежности относятся к нечеткому отрезку, концами которого является первая и k-я нечеткие точки, то все нечеткие точки могут принадлежать нечеткой прямой. Другими словами, может существовать такая четкая прямая, относительно которой все нечеткие точки имеют ненулевые степени принадлежности.

Обозначим ω_i границу нечеткой точки $\widetilde{A}_i$ в просторные $\tilde{\mathbf{A}}^2$. Обозначим ext $\omega_i(\widetilde{A}_1 \widetilde{A}_k)$ внешнюю границу нечеткой точки $\widetilde{A}_i$ относительно нечеткого отрезка $\widetilde{A}_1 \widetilde{A}_k$, а int $\omega_i(\widetilde{A}_1 \widetilde{A}_k)$— внутреннюю границу нечеткой точки $\widetilde{A}_i$ относительно нечеткого отрезка $(\widetilde{A}_1 \widetilde{A}_k)$. Если среди отрезков

$$A_1A_k,\ A_1 \in \text{ext}\ \omega_1(\widetilde{A}_1\widetilde{A}_k),\ A_k \in \text{ext}\ \omega_k(\widetilde{A}_1\widetilde{A}_k)$$

обнаружится такой отрезок, относительно которого все другие нечеткие точки будут иметь ненулевую степень принадлежности, то этот отрезок определит искомую прямую.

Естественно, что такие четкие прямые образуют множество или нечеткую прямую (рис. 4).

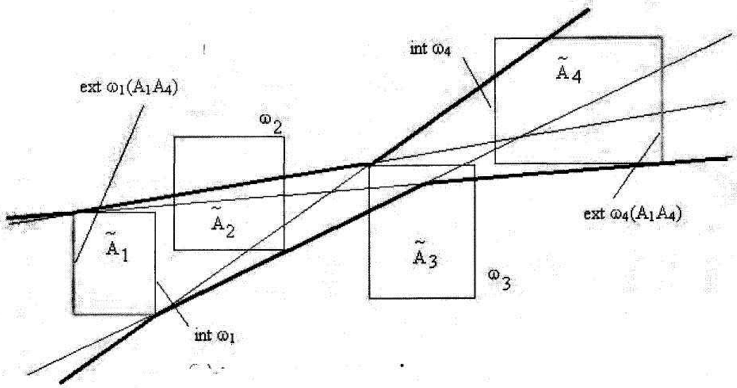


Рис. 4. Нечеткая прямая, которая определена четырьмя нечеткими точками

Возникает задача построения такой прямой, которая имеет максимальное значение функции принадлежности данному множеству нечетких точек. Прежде, чем привести алгоритм решения, опишем две вспомогательных задачи.

Первая. Даны нечеткая точка $\widetilde{A} \in \tilde{\mathbf{A}}^2$ и четкая прямая $a \subset \tilde{\mathbf{A}}^2$. Если функция их взаимной принадлежности не равна нулю и не равна единице, найти ее значение. Для решения необходимо определить конкретный вид нечеткой точки. Пусть, например, нечеткая точка будет *(R — L)* -типа. Напомним, что нечеткая точка будет точкой *(R — L)* -типа, если все ее координаты будут нечеткими числами *(R — L)*-типа. Пусть $a \bigcap \widetilde{A} \neq \varnothing$. Тогда значение функции взаимной принадлежности определяется по формуле

$$\mu_{\tilde{\mathrm{A}}^2}(\widetilde{A} \subset a) = \frac{\delta}{\alpha_y},$$

в которой вместо α_y может быть β_y, α_x, β_x в зависимости от расположения прямой и точки (рис. 5).

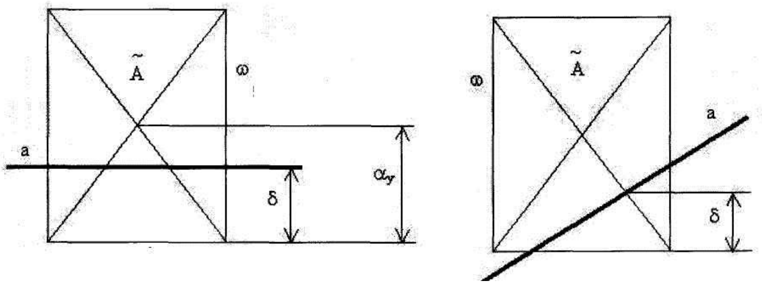


Рис.5. Определение степени принадлежности нечеткой точки и прямой

Вторая. Дано три нечетких точки $\widetilde{A} \in \widetilde{\mathrm{A}}^2$, $\widetilde{B} \in \widetilde{\mathrm{A}}^2$, $\widetilde{C} \in \widetilde{\mathrm{A}}^2$ и прямая $a \subset$ $\tilde{\mathbf{A}}^2$. Функции взаимной принадлежности трех точек прямой не равны нулю и не равны единицы, поскольку точки заданы в общем положении. Найти такое положение прямой, при котором

$$\mu_{\widetilde{\mathrm{A}}^2}(\widetilde{A} \subset a) = \mu_{\widetilde{\mathrm{A}}^2}(\widetilde{B} \subset a) = \mu_{\widetilde{\mathrm{A}}^2}(\widetilde{C} \subset a) = \mu_{\widetilde{\mathrm{A}}^2}(\{\widetilde{A}, \widetilde{B}, \widetilde{C}\} \subset a).$$

Задача имеет точное решение. Алгоритм может быть следующий.

1. Анализ расположения точек $\widetilde{A} \in \tilde{\mathbf{A}}^2$ и $\widetilde{B} \in \tilde{\mathbf{A}}^2$. Смысл здесь в следующем. Две нечеткие точки (для простоты приблизительно $(R — L)$ -типа) в просторанстве $\tilde{\mathbf{A}}^2 \times \mu$ имеют по четыре ребра каждая:

$$\lfloor A(\alpha_x, \alpha_y)_A \rfloor, \lfloor A(\alpha_x, \beta_y)_A \rfloor, \lfloor A(\beta_x, \alpha_y)_A \rfloor, \lfloor A(\beta_x, \beta_y)_A \rfloor,$$

$$[B(\alpha_x, \alpha_y)_B], [B(\alpha_x, \beta_y)_B], [B(\beta_x, \alpha_y)_B], [B(\beta_x, \beta_y)_B].$$

Вместо того, чтобы учитывать все шестнадцать пар ребер, можно исключить по два ребра из каждой точки и рассматривать только четыре пары. Это можно сделать разными способами, которые не будем описывать. В результате будут полученные четыре нечетких отрезка, которые определят четыре нечеткие прямые.

2. Определяются точки пересечения каждой из этих прямых с аналогичными ребрами нечеткой точки $\widetilde{C} \in \tilde{\mathbf{A}}^2$.

3. Из тех точек пересечения, которые находятся в пространстве $\tilde{\mathbf{A}}^2 \times \mu$, выбирается точка с максимальным значением функции принадлежности. Поскольку все нечеткие прямые, которые принимают участие в расчете, представляют собой множество прямых уровня в пространстве $\tilde{\mathbf{A}}^2 \times \mu$, то найденное максимальное значение функции принадлежности будет одинаковым для трех заданных точек.

Теперь возвратимся к первоначальной задаче с k заданными нечеткими точками. Алгоритм ее решения может быть таким.

1. Рассматривается множество отрезков

$$A_1 A_k\,,\; A_1 \in ext\;\omega_1(\widetilde{A}_1 \widetilde{A}_k),\; A_k \in ext\;\omega_k(\widetilde{A}_1 \widetilde{A}_k).$$

2. Для каждого отрезка вычисляются значения функции принадлежности всех промежуточных нечетких точек.

3. По условию

$$\min \mu_{\widetilde{A}^2}\left(\widetilde{A}_i \subset \left[\widetilde{A}_1 \widetilde{A}_k\right]\right) \to \max$$

образовывается итерационный алгоритм, в результате которого будет найдена искомая прямая. Рассмотрим условия пересечения.

Определение 19. Если для нечеткой k-плоскости $\widetilde{\alpha}$ и нечеткой m-плоскости $\widetilde{\beta}$ нечеткого пространства $\tilde{\mathbf{A}}^n$ выполняется условие $k+m-n=0$, то их пересечением является нечеткое множество с

$$\max \mu_{\widetilde{A}}(\widetilde{\alpha} \cap \widetilde{\beta}) = 1 .$$

Естественно, что

$$\widetilde{\alpha} \cap \widetilde{\beta} = \mathrm{core}\,\widetilde{\alpha} \cap \mathrm{core}\,\widetilde{\beta}\,.$$

Этот случай не представляет теоретического интереса, кроме определения параметров функции принадлежности для нечеткого множества пересечения.

Определение 20. Если для нечеткой k - плоскости $\widetilde{\alpha}$ и нечеткой m-плоскости $\widetilde{\beta}$ нечеткого пространства $\tilde{\mathbf{A}}^n$ выполняется условие $k + m - n<0$, то их пересечением может быть нечеткое множество с

$$\mu_{\widetilde{A}}(\widetilde{\alpha} \cap \widetilde{\beta}) < 1$$

Естественно, что

$$\mathrm{core}\,\widetilde{\alpha} \cap \mathrm{core}\,\widetilde{\beta} = \varnothing\,.$$

Другими словами, два нечетких подпространства пересекаются, если существует точка, которая принадлежит им с разной, не нулевой степенью принадлежности.

Возникает следующая вычислительная позиционная задача. Дано два нечетких подпространства: k-плоскость $\widetilde{\alpha}$ и m-плоскость $\widetilde{\beta}$ нечеткого пространства $\tilde{\mathbf{A}}^n$ и условие $k + m - n < 0$. Определить существование их пересечения и значение функции принадлежности, которое будет преднамеренно меньше единице.

Алгоритм решения задачи.

1. Определяется линия *MN* кратчайшего расстояния между заданными подпространствами, (рис. 6).

$$M \in \operatorname{core} \widetilde{\alpha}, N \in \operatorname{core} \widetilde{\beta}$$

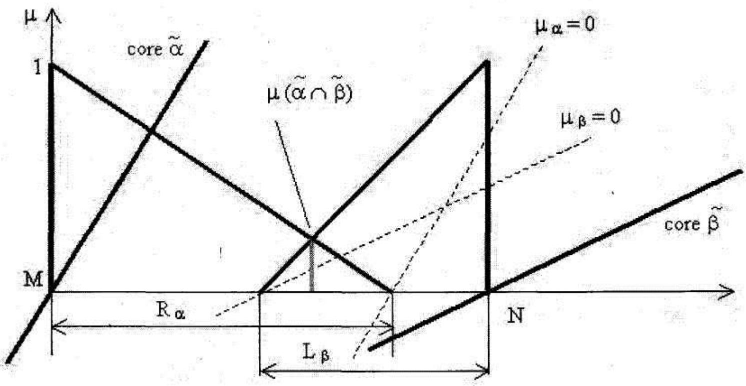


Рис. 6. Определение значения функции принадлежности для условия пересечения

2. Если

$$|MN| > R_{\widetilde{\alpha}}(MN) + L_{\widetilde{\beta}}(MN),$$

это пересечение существует.

3. Допуская линейные функции принадлежности, можно получить следующее выражение для значения функции принадлежности

$$\mu_{\widetilde{A}^n}(\widetilde{\alpha} \cap \widetilde{\beta}) = \frac{R_{\widetilde{\alpha}}(MN) + L_{\widetilde{\beta}}(MN) - |MN|}{R_{\widetilde{\alpha}}(MN) + L_{\widetilde{\beta}}(MN)}.$$

14.5. Метрические задачи

Основными метрическими задачами являются задачи, в которых определяются длины, углы, площади, объемы и так далее, при условии нечеткости исходных данных.

Согласно определению 4, если две нечеткие точки раздичны, то между ними существует расстояние. Напомним определение расстояния.

Если Ω - некоторое множество и если существует отображение $\Omega \times \Omega \to R^+$, то величина $d(x, y)$ является расстоянием в Ω, если при $x, y, z \in \Omega$ выполняются следующие условия:

1) $d(x,x)=0$;
2) $d(x,y)= d(y,x)$;
3) $d(x,z)\leq d(x,y)+d(y,z)$.

Существуют следующие известные расстояния между нечеткими множествами:

1) расстояние Хеммтнга для конечного множества

$$d(A, B) = \sum_{i=1}^{n} \left| \mu_A(x_i) - \mu_B(x_i) \right|;$$

2) относительное расстояние Хемминга для конечного множества

$$d(A, B) = \frac{1}{n}\sum_{i=1}^{n} \left| \mu_A(x_i) - \mu_B(x_i) \right|;$$

3) расстояние Евклида для конечного множества

$$d(A, B) = \sqrt{\sum_{i=1}^{n} \left(\mu_A(x_i) - \mu_B(x_i)\right)^2};$$

4) относительное расстояние Евклида для конечного множества

$$d(A, B) = \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n} \left(\mu_A(x_i) - \mu_B(x_i)\right)^2}.$$

Кроме них существуют и другие расстояния, которые можно найти в специальной литературе.

Рассмотрим возможность применения этих расстояний для определения длины нечеткого отрезка $\widetilde{A}\widetilde{B}$ в различных случаях. Пусть $\widetilde{A} \cap \widetilde{B} \neq \varnothing$. Если $\widetilde{A} = \widetilde{B}$, то очевидно, что всегда $d(\widetilde{A}, \widetilde{B})=0$. Если $\widetilde{A} \neq \widetilde{B}$, то расстояние при уменьшении множества $\widetilde{A} \cap \widetilde{B} \neq \varnothing$ стремится к постоянному значению. Например, для расстояния Евклида будет

$$d(\widetilde{A}, \widetilde{B}) \to \sqrt{\sum_{i=1}^{n} \left(\mu_{\widetilde{A}}(x_i)\right)^2 + \left(\mu_{\widetilde{B}}(x_i)\right)^2},$$

где $x_i \in \widetilde{A}\widetilde{B}$. В случае $\widetilde{A} \cap \widetilde{B} \neq \varnothing$ расстояние будет равно указанной величине, независимо от положения нечетких точек на нечеткой плоскости. Следовательно, ни одно из приведенных расстояний нельзя использовать для определения длины нечеткого отрезка.

С другой стороны, можно определить длину нечеткого отрезка в метрике Кантора в виде

$$d(\widetilde{A}, \widetilde{B}) = \inf \{d(A, B)\},$$

где $A \in \widetilde{A}$, $B \in \widetilde{B}$ или метрике Хаусдорфа

$$d(\widetilde{A}, \widetilde{B}) = \max \{\max \{d(A \in \widetilde{A}, \widetilde{B}\}, \max \{d(\widetilde{A}, B \in \widetilde{B}\}\}.$$

Однако и в этом случае существует сложность, которую покажем на примере. Пусть будут дано три нечетких точки $\widetilde{A}$, $\widetilde{B}$, $\widetilde{C}$ $\in$ $\widetilde{A}^2$ (рис. 7):

Предположим, что

$$x_a - L_{a,x} - x_b - R_{b,x} = x_c - L_{c,x} - x_a - R_{a,x},$$
$$y_b - L_{b,y} - y_a - R_{a,y} = y_c - L_{c,y} - y_a - R_{a,y}.$$

Тогда становится очевидным, что

$$\inf \{d(A, B)\} = \inf \{d(A, C)\}.$$

$$\widetilde{A} = (A(x_a, y_a), R_{a,x}, R_{a,y}, L_{a,x}, L_{a,y}),$$
$$\widetilde{B} = (B(x_b, y_b), R_{b,x}, R_{b,y}, L_{b,x}, L_{b,y}),$$
$$\widetilde{C} = (C(x_c, y_c), R_{c,x}, R_{c,y}, L_{c,x}, L_{c,y}).$$

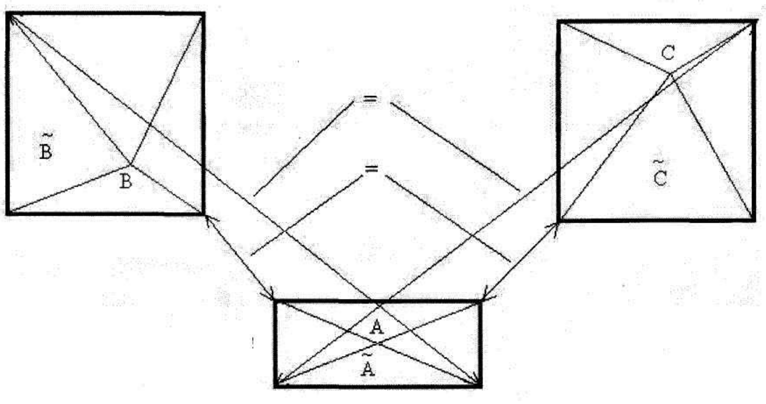


Рис. 7. К определению расстояния между нечеткими точками

Предположим также, что

$$x_a + R_{a,x} - x_b + L_{b,x} = x_c + R_{c,x} - x_a + L_{a,x},$$
$$y_b + R_{b,y} - y_a + L_{a,y} = y_c + R_{c,y} - y_a + L_{a,y}.$$

Тогда выполняется и второе равенство

$$\max\left\{\max\left\{d(A \in \widetilde{A}, \widetilde{B}\right\}, \max\left\{d(\widetilde{A}, B \in \widetilde{B}\right\}\right\} =$$
$$= \max\left\{\max\left\{d(A \in \widetilde{A}, \widetilde{C}\right\}, \max\left\{d(\widetilde{A}, C \in \widetilde{C}\right\}\right\}.$$

Но предположим, что

$$R_{b,x} < R_{c,x},\ L_{b,x} > L_{c,x},\ R_{b,y} > R_{c,y},\ L_{b,y} < L_{c,y},$$
$$R_{b,x} + L_{b,x} = R_{c,x} + L_{c,x},\ R_{b,y} + L_{b,y} = R_{c,y} + L_{c,y}.$$

Тогда можно утверждать, что в силу соотношения

$$d(A, B) < d(A, C)$$

имеет место утверждение

$$d(\widetilde{A}, \widetilde{B}) < d(\widetilde{A}, \widetilde{C}).$$

Следовательно, приведенных способов вычисления расстояния между двумя нечеткими точками недостаточно. Другими словами, такими способами определить длину нечеткого отрезка нельзя. Больше того, для двух нечетких отрезков, которые имеют равные длины в метрике Кантора, метрике Хаусдорфа и имеют равные длины их ядер, нельзя утверждать, что расстояния между их концами равны.

Следовательно, рассматривать следует случай, когда $\widetilde{A} \cap \widetilde{B} \neq \varnothing$.

Определение 21. Под длиной нечеткого отрезка $\widetilde{A}\widetilde{B}$, заданного нечеткими точками «возле *A»* и «близко *B*», будем понимать нечеткое число $l = |\widetilde{A}\widetilde{B}|$, которое определено понятиям «возле $|\widetilde{A}\widetilde{B}|$».

Можно утверждать следующее:

1) $\operatorname{core}\widetilde{l} = \left|\operatorname{core}\widetilde{A}\ \operatorname{core}\widetilde{B}\right|$;

2) $\sup\widetilde{l} = \max\left\{d(A \in \widetilde{A}, B \in \widetilde{B})\right\} - \inf\left\{d(A \in \widetilde{A}, B \in \widetilde{B})\right\}$;

3) если представить понятие «возле $|\widetilde{A}\widetilde{B}|$» как число *(R — L)* -типа, то

$$L_{\widetilde{A}\widetilde{B}} = \text{core}\,\widetilde{l} - \inf\left\{d(A \in \widetilde{A}, B \in \widetilde{B})\right\},$$

$$R_{\widetilde{A}\widetilde{B}} = \max\left\{d(A \in \widetilde{A}, B \in \widetilde{B}\right\} - \text{core}\,\widetilde{l}.$$

Однако при общем расположении нечетких концов отрезка, т.е. если отрезок общего вида, функция принадлежности понятия «возле $|\widetilde{A}\widetilde{B}|$» будет более высокой степени, чем функция принадлежности понятий «возле $\widetilde{A}$» и «близко $\widetilde{B}$». Например, если последние являются кусочно-линейными, то функция принадлежности понятия «возле $|\widetilde{A}\widetilde{B}|$» будет нелинейной. В этом случае получить конкретный вид функции принадлежности для понятия «возле $|\widetilde{A}\widetilde{B}|$» довольно легко.

Пусть

$$\widetilde{A}, \widetilde{B} \in \widetilde{A}^2,\ \widetilde{A} \cap \widetilde{B} = \varnothing,\ x_a + R_{a,x} < x_b - L_{b,x},\ y_a + R_{a,y} < y_b - L_{b,y},$$

$$A_{++} = A(x_a + R_{a,x}, y_a + R_{a,y}),\ B_{--} = B(x_b - L_{b,x}, y_b - L_{b,y}).$$

Используя понятие α-уровня, можно записать

$$d_\alpha(\widetilde{A}, \widetilde{B}) = \sqrt{\left(x_{a,\alpha} - x_{b,\alpha}\right)^2 + \left(y_{a,\alpha} - y_{b,\alpha}\right)^2},$$

где

$$x_{a,\alpha} = (1-\alpha)(x_a + R_{a,x}) + \alpha \cdot x_a$$

и так далее.

Таким образом, получив функции принадлежности для длин двух различных нечетких отрезков, мы можем их сравнить по приведенным выше формулах. Обобщение на многомерное пространство, очевидно, не вызывает трудностей.

Определение 22. Под площадью нечеткого треугольника $\widetilde{A}\widetilde{B}\widetilde{C}$, заданного нечеткими точками «возле $\widetilde{A}$», «возле $\widetilde{B}$» и «возле $\widetilde{C}$»,

$$\widetilde{A} \cap \widetilde{B} \cap \widetilde{C} = \varnothing$$

будем понимать нечеткое число

$$\widetilde{s} = f(\widetilde{A}\widetilde{B}\widetilde{C}),$$

которое определено понятием «приблизительно $\widetilde{s}$». Можно утверждать следующее:

1) $\text{core}\,\widetilde{s} = f\left(\text{core}\,\widetilde{A}, \text{core}\,\widetilde{B}, \text{core}\,\widetilde{C}\right);$

2) $\operatorname{core}\tilde{s} + R_{\tilde{A}\tilde{B}\tilde{C}} = \max\{f(A, B, C)\}$,

где

$A \in \left(\operatorname{ext}\omega_{\tilde{A}}(\tilde{A}\tilde{B}) \cap \operatorname{ext}\omega_{\tilde{A}}(\tilde{A}\tilde{C})\right)$, $B \in \left(\operatorname{ext}\omega_{\tilde{B}}(\tilde{A}\tilde{B}) \cap \operatorname{ext}\omega_{\tilde{B}}(\tilde{B}\tilde{C})\right)$,

$C \in \left(\operatorname{ext}\omega_{\tilde{C}}(\tilde{A}\tilde{C}) \cap \operatorname{ext}\omega_{\tilde{C}}(\tilde{B}\tilde{C})\right)$;

3) $\operatorname{core}\tilde{s} - L_{\tilde{A}\tilde{B}\tilde{C}} = \inf\{f(A, B, C)\}$,

где

$A \in \left(\operatorname{int}\omega_{\tilde{A}}(\tilde{A}\tilde{B}) \cap \operatorname{int}\omega_{\tilde{A}}(\tilde{A}\tilde{C})\right)$, $B \in \left(\operatorname{int}\omega_{\tilde{B}}(\tilde{A}\tilde{B}) \cap \operatorname{int}\omega_{\tilde{B}}(\tilde{B}\tilde{C})\right)$,

$C \in \left(\operatorname{int}\omega_{\tilde{C}}(\tilde{A}\tilde{C}) \cap \operatorname{int}\omega_{\tilde{C}}(\tilde{B}\tilde{C})\right)$.

Это определение с очевидностью обобщается на определение объема нечеткого тетраэдра, гиперобъема нечеткого многомерного симплекса.

Но три точки на плоскости могут определять не только треугольник, но и круг.

Определение 23. Под площадью нечеткого круга

$$\tilde{r}(\tilde{A}\tilde{B}\tilde{C}),$$

заданного нечеткими точками «возле $\tilde{A}$», «возле $\tilde{B}$» и «возле $\tilde{C}$»,

$$\tilde{A} \cap \tilde{B} \cap \tilde{C} = \varnothing,$$

будем понимать нечеткое число

$$\tilde{s} = f(\tilde{A}\tilde{B}\tilde{C}),$$

которое определено понятием «приблизительно $\tilde{s}$ ».

Можно утверждать следующее:

1) $\operatorname{core}\tilde{s} = f\left(\operatorname{core}\tilde{A}, \operatorname{core}\tilde{B}, \operatorname{core}\tilde{C}\right)$;

2) $\operatorname{core}\tilde{s} + R_{\tilde{A}\tilde{B}\tilde{C}} = \max\{f(A, B, C)\}$,

где

$A \in \left(\operatorname{ext}\omega_{\tilde{A}}(\tilde{A}\tilde{B}) \cap \operatorname{ext}\omega_{\tilde{A}}(\tilde{A}\tilde{C})\right)$, $B \in \left(\operatorname{ext}\omega_{\tilde{B}}(\tilde{A}\tilde{B}) \cap \operatorname{ext}\omega_{\tilde{B}}(\tilde{B}\tilde{C})\right)$,

$C \in \left(\operatorname{ext}\omega_{\tilde{C}}(\tilde{A}\tilde{C}) \cap \operatorname{ext}\omega_{\tilde{C}}(\tilde{B}\tilde{C})\right)$;

$$\left(\left(\operatorname{ext}\omega_{\tilde{A}}\left(\tilde{A}\tilde{B}\right)\cap\operatorname{ext}\omega_{\tilde{A}}\left(\tilde{A}\tilde{C}\right)\right)\cap r(A,B,C)\right)/A=\varnothing,$$

$$\left(\left(\operatorname{ext}\omega_{\tilde{B}}\left(\tilde{A}\tilde{B}\right)\cap\operatorname{ext}\omega_{\tilde{B}}\left(\tilde{B}\tilde{C}\right)\right)\cap r(A,B,C)\right)/B=\varnothing,$$

$$\left(\left(\operatorname{ext}\omega_{\tilde{C}}\left(\tilde{A}\tilde{C}\right)\cap\operatorname{ext}\omega_{\tilde{C}}\left(\tilde{B}\tilde{C}\right)\right)\cap r(A,B,C)\right)/C=\varnothing;$$

3) $\operatorname{core}\tilde{s}-L_{\tilde{A}\tilde{B}\tilde{C}}=\inf\{f(A,B,C)\},$

где ,

$A\in\left(\operatorname{int}\omega_{\tilde{A}}\left(\tilde{A}\tilde{B}\right)\cap\operatorname{int}\omega_{\tilde{A}}\left(\tilde{A}\tilde{C}\right)\right)$, $B\in\left(\operatorname{int}\omega_{\tilde{B}}\left(\tilde{A}\tilde{B}\right)\cap\operatorname{int}\omega_{\tilde{B}}\left(\tilde{B}\tilde{C}\right)\right)$,

$C\in\left(\operatorname{int}\omega_{\tilde{C}}\left(\tilde{A}\tilde{C}\right)\cap\operatorname{int}\omega_{\tilde{C}}\left(\tilde{B}\tilde{C}\right)\right);$

$$\left(\left(\operatorname{int}\omega_{\tilde{A}}\left(\tilde{A}\tilde{B}\right)\cap\operatorname{int}\omega_{\tilde{A}}\left(\tilde{A}\tilde{C}\right)\right)\cap r(A,B,C)\right)/A=\varnothing,$$

$$\left(\left(\operatorname{int}\omega_{\tilde{B}}\left(\tilde{A}\tilde{B}\right)\cap\operatorname{int}\omega_{\tilde{B}}\left(\tilde{B}\tilde{C}\right)\right)\cap r(A,B,C)\right)/B=\varnothing,$$

$$\left(\left(\operatorname{int}\omega_{\tilde{C}}\left(\tilde{A}\tilde{C}\right)\cap\operatorname{int}\omega_{\tilde{C}}\left(\tilde{B}\tilde{C}\right)\right)\cap r(A,B,C)\right)/C=\varnothing.$$

Точно так же можно построить обобщение на определение объема нечеткой сферы, гиперсферы. Кроме этого можно определить длину нечеткого круга или его радиус, если иметь в виду вид функции f, которая принимает участие в определении. В общем случае от нечеткого определителя нечеткой фигуры всегда можно перейти к нечетким параметрам этой фигуры. Далее, учитывая то, что нечеткие элементы определителя всегда могут быть разбиты на α-равные, можно, применяя приведенные определения для α - уровней, построить функции принадлежности нечетких параметров.

В качестве примера рассмотрим следующую задачу.

На нечеткой плоскости определено понятия «возле», постояно на всей плоскости. Даны четкие различные точки A, B, C, D, E, F общего положения. Определить, с какой степенью принадлежности множество из трех отрезков, концами которых являются данные точки, может быть отнесено множество нечетких треугольников. Определить площадь полученного нечеткого треугольника.

Алгоритм решения задачи может быть следующим.

1) Анализируются положение заданных точек. Целью анализа является выбор трех пар, точки которых больше всего близко

расположены дуг к другу. Пусть такими парами будут пары *AB, CD, EF*.

2) Допуская, что пара *AB* образует приблизительно вершину треугольника, т.е. принадлежит нечеткой точке *M* , находим значение функции принадлежности

$$\mu_{\widetilde{M}}(A) = \mu_{\widetilde{M}}(B) \to \max .$$

Это условие осуществимо и достигается шевелением заданной нечеткой точки. Более того, если нечеткая точка *M* задана нечеткими числами *(R — L)* -типа, то такое условие будет всегда осуществимым для четырех четких точек.

Аналогично вводим нечеткие точки *N, P* и их шевелением находим значение функций принадлежности

$$\mu_{\widetilde{N}}(C) = \mu_{\widetilde{N}}(D) \to \max$$

и

$$\mu_{\widetilde{P}}(E) = \mu_{\widetilde{P}}(F) \to \max .$$

Предположим, что

$$\widetilde{M} \cap \widetilde{N} \cap \widetilde{P} = \varnothing$$

(рис. 8). Если это не так, то множество треугольников будет включать в себя треугольники, которые вырождены в отрезки с их соответствующими степенями принадлежности.

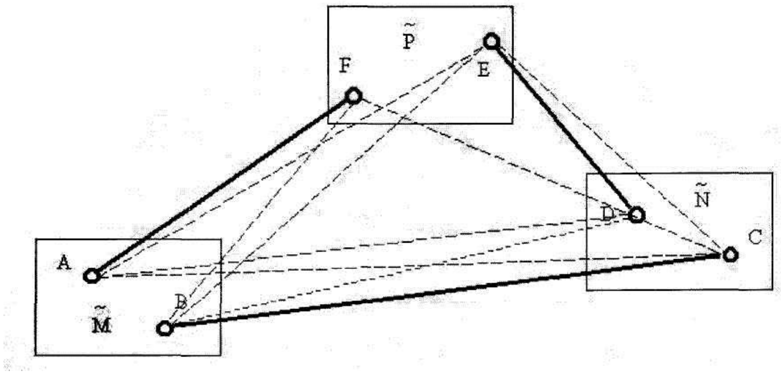


Рис. 8. К определению параметров нечеткого треугольника

Таким образом, данное множество образует нечеткую фигуру, степень принадлежности которой множеству треугольников равна

$$\mu_{\tilde{M}\tilde{N}\tilde{P}}(A, B, C, D, E, F) = \min\left\{\mu_{\tilde{M}}(A), \mu_{\tilde{N}}(C), \mu_{\tilde{P}}(E)\right\}.$$

Заметим, что полученный нечеткий треугольник не единственный, так как его стороны могут быть образованы другими соединениями точек. Они показаны на рис. 6.8 штриховыми линиями. Но значение функции принадлежности при любом соединении точек не изменится.

3) Для определения его площади рассмотрим α - уровень, который отвечает значению

$$\mu_{\tilde{M}\tilde{N}\tilde{P}}(A, B, C, D, E, F) = \min\left\{\mu_{\tilde{M}}(A), \mu_{\tilde{N}}(C), \mu_{\tilde{P}}(E)\right\}.$$

и найдем нечеткое число так, как описано выше.

Введем понятие сепаратности носителей двух нечетких точек - концов нечеткого отрезка $\tilde{A}\tilde{B}$.

Определение 24. Под сепаратностью носителей двух нечетких точек $\tilde{A}, \tilde{B} \in \tilde{A}^2$ будем понимать число

$$\operatorname{sep}(\sup\tilde{A}, \sup\tilde{B}) = |AB'| + |AB''| - \left|\operatorname{int}\omega_A(\tilde{A}\tilde{B})\right| - \left|\operatorname{int}\omega_B(\tilde{A}\tilde{B})\right|,$$

где $|AB'|$, $|AB''|$ - длины двух четких отрезков прямых нечеткого отрезка $\tilde{A}\tilde{B}$, для которых выполняются условия

$$\mu_{\tilde{A}\tilde{B}}(AB') = \mu_{\tilde{A}\tilde{B}}(AB'') = 0, \quad AB' \cap AB'' \in \tilde{A}\tilde{B}, |AB'| = \max, |AB''| = \max .$$

Можно доказать, что sep $(\sup\tilde{A}$, $\sup\tilde{B}$ $)>0$. Но если две нечеткие точки касаются, т.е. имеют хотя бы одну общую точку с нулевым значением функции принадлежности, то сепаратность их носителей становится равной нулю. Однако это не означает, что сепаратность самих нечетких точек тоже будет равна нулю.

Определим сепаратность α-уровня следующим образом. Под сепаратностью α -уровня двух нечетких точек $\tilde{A}, \tilde{B} \in \tilde{A}^2$ будем понимать число

$$\operatorname{sep}_\alpha(\sup_\alpha\tilde{A}, \sup_\alpha\tilde{B}) = |AB'|_\alpha + |AB''|_\alpha - \left|\operatorname{int}\omega_A(\tilde{A}\tilde{B})\right|_\alpha - \left|\operatorname{int}\omega_B(\tilde{A}\tilde{B})\right|_\alpha,$$

где $|AB'|_\alpha$, $|AB''|_\alpha$ - длины двух четких отрезков прямых нечеткого отрезка $\tilde{A}\tilde{B}$, для которых выполняются условия

$$\mu_{\tilde{A}\tilde{B}}(AB') = \mu_{\tilde{A}\tilde{B}}(AB'') = \alpha, \quad AB' \cap AB'' \in \tilde{A}\tilde{B}, |AB'|_\alpha = \max, |AB''|_\alpha = \max .$$

Теперь сепаратность двух нечетких точек $\tilde{A}, \tilde{B} \in \tilde{A}^2$ можно найти по формулам:

$$\operatorname{sep}(\tilde{A}, \tilde{B}) = \sum_{\alpha=1}^{n} \operatorname{sep}_{\alpha}(\tilde{A}, \tilde{B});$$

$$\operatorname{sep}(\tilde{A}, \tilde{B}) = \frac{1}{n} \sum_{\alpha=1}^{n} \operatorname{sep}_{\alpha}(\tilde{A}, \tilde{B});$$

$$\operatorname{sep}(\tilde{A}, \tilde{B}) = \sqrt{\sum_{\alpha=1}^{n} \operatorname{sep}_{\alpha}^{2}(\tilde{A}, \tilde{B})};$$

$$\operatorname{sep}(\tilde{A}, \tilde{B}) = \frac{1}{\sqrt{n}} \sqrt{\sum_{\alpha=1}^{n} \operatorname{sep}_{\alpha}^{2}(\tilde{A}, \tilde{B})}.$$

Возвратимся еще раз к некоторым нечетким образам, которые упоминались выше. Известно, что угол образуется двумя отрезками с общим концом. Нечеткий угол образуется двумя нечеткими отрезками с общим нечетким концом. Пусть у двух нечетких отрезках $\tilde{A}\tilde{B}$, $\tilde{C}\tilde{D}$ существует общая нечеткая точка $\tilde{A} \cap \tilde{C}$. Тогда образуется нечеткий угол, вершина которого принадлежит нечеткой точке $\tilde{A} \cap \tilde{C}$. При этом его максимальное значение образуется двумя внешними опорными отрезками, общий конец которых лежит на границе

$$\operatorname{int} \omega_{\tilde{A}\cap\tilde{C}} \left(\tilde{A} \cap \tilde{C}, \tilde{B}\right) \cap \operatorname{int} \omega_{\tilde{A}\cap\tilde{C}} \left(\tilde{A} \cap \tilde{C}, \tilde{D}\right)$$
$$\operatorname{ext} \omega_{\tilde{A}\cap\tilde{C}} \left(\tilde{A} \cap \tilde{C}, \tilde{B}\right) \cap \operatorname{ext} \omega_{\tilde{A}\cap\tilde{C}} \left(\tilde{A} \cap \tilde{C}, \tilde{D}\right).$$

Его минимальное значение образуется двумя внутренними опорными отрезками, общий конец которых лежит на границе

$$\operatorname{ext} \omega_{\tilde{A}\cap\tilde{C}} \left(\tilde{A} \cap \tilde{C}, \tilde{B}\right) \cap \operatorname{ext} \omega_{\tilde{A}\cap\tilde{C}} \left(\tilde{A} \cap \tilde{C}, \tilde{D}\right).$$

14.6. Нечеткие условия

Рассмотрим более общие условия, чем условия инцидентности - условия параллельности, перпендикулярности, касания.

Пусть на нечеткой плоскости будут заданы две четкие прямой. Определим для них понятие «приблизительно параллельные», для чего на одной из прямых выберем произвольно нечеткий отрезок $\tilde{A}\tilde{B}$. Если расстояние между точкой A и точкой core $\tilde{B}$ довольно большое, то прямая, которая определена точками A и $B \in \tilde{B}$, будет приблизительно параллельна другой прямой. Обозначим значение функции принадлежности двух прямых множества параллельных прямых как *v*. Можем записать

$$\nu_{\tilde{A}^2}\left(AB \;||\; ...\right) = 1 - \frac{h\left(\operatorname{int}\, \omega_{\tilde{B}}\;(A\tilde{B})\right) \cdot a}{\left|A\operatorname{core}\tilde{B}\right|}.$$

Здесь $h(\operatorname{int}\ \omega_{\tilde{B}}\,(A\tilde{B}))$ - длина хорды, которая соединяет концы внутренней границы нечеткой точки $\tilde{B}$ на отрезке $A\tilde{B}$, a - положительное действительное число.

Пусть будет заданы четкая прямая и нечеткий отрезок $\tilde{A}\tilde{B}$. Они будут приблизительно параллельны, если расстояния от концов нечеткого отрезка к прямой приблизительно равны, а сами концы достаточно отдалены друг от друга. Это можно выразить условием, что ядро нечеткого отрезка параллельно заданной прямой и sep $(\tilde{A}, \tilde{B})$»$h(\operatorname{int}\omega_{\tilde{A}\tilde{B}}\ (\tilde{A}\,\tilde{B}))$. Значение функции принадлежности множеству параллельных прямых можно выразить формулой

$$\nu_{\tilde{A}^2}\left(AB \;||\; ...\right) = 1 - \frac{\left(h\left(\operatorname{int}\, \omega_{\tilde{A}}\;(\tilde{A}\tilde{B})\right) + h\left(\operatorname{int}\, \omega_{\tilde{B}}\;(\tilde{A}\tilde{B})\right)\right) \cdot a}{\left|\operatorname{core}\, \tilde{A}\tilde{B}\right|}.$$

Два нечетких отрезка параллельны, если они гомотетичны. Другими словами, два нечетких отрезка параллельны, если для любого четкого отрезка с определенной степенью принадлежности одного нечеткого отрезка в другом нечетком отрезке определится параллельный ему четкий отрезок с такой же степенью принадлежности.

Пусть даны два нечетких отрезка

$$\tilde{A}\tilde{B},\ \tilde{C}\tilde{D} \in \tilde{A}^2.$$

Они будут параллельны при условии, ято для любого отрезка $MN \in \tilde{A}\tilde{B}$ определится такой параллельный ему отрезок $PQ \in \tilde{C}\tilde{D}$, что

$$\mu_{\tilde{A}}(M) = \mu_{\tilde{B}}(N) = \mu_{\tilde{C}}(P) = \mu_{\tilde{D}}(Q).$$

Другими словами значение функции принадлежности этих двух отрезков множества параллельных нечетких отрезков равно единице.

Два нечетких отрезка приблизительно параллельны, если для любого четкого отрезка, который принадлежит одному из них, определится четкий отрезок, который принадлежит другому из них, приблизительно параллельный первому. Если два нечетких отрезка приблизительно параллельны, то наибольшее значение функции принадлежности по условию параллельности будет относиться к отрезкам, которые пересекают ядра в одном направлении, т.е. к отрезкам, которые лежат на внутренних, приблизительно одинаково направленных опорных прямых нечетких концов данных отрезков. Меньше всего значение функции принадлежности по условию параллельности будет относиться к отрезкам внутренних, противоположно направленных опорных прямых.

Две нечеткие прямые приблизительно параллельны, если образующие их нечеткие отрезки приблизительно параллельны. Две нечеткие прямые параллельны, если между образующими их нечеткими отрезками существует взаимно однозначное соответствие и соответствующие нечеткие отрезки параллельны. Можно сказать, что две нечеткие прямые приблизительно параллельны, если расстояние между их соответствующими нечеткими точками (или сепаратность) остается приблизительно постоянной.

Эти определения легко обобщаются на большие размерности. Вообще тема нечетких условий в многомерных нечетких пространствах требует отдельного подробного исследования. Ограничимся здесь только некоторыми примерами.

Четкая прямая и четкое линейное подпространство (не уточняя размерности) нечеткого пространства приблизительно параллельны, если в подпространстве обнаружится прямая, приблизительно параллельная данной. Два четких линейных подпространства разной размерности, которые находятся в нечетком пространстве, приблизительно параллельны, если приблизительно придерживается их признак параллельности. Признаков параллельности в этом случае можно сформулировать несколько. Конкретное число зависит от размерности подпространств.

Два четких отрезка в нечетком пространстве являются приблизительно перпендикулярными, если один из них приблизительно параллельный перпендикуляру к другому.

Два нечетки отрезка являются перпендикулярными, если для каждого четкого отрезка, который принадлежит одному нечеткому с некоторым значением функции принадлежности, в другом нечетком отрезке обнаружится четкий отрезок, перпендикулярный первому и который имеет такое же значение функции принадлежности. Два перпендикулярных нечетких отрезка связаны друг с другом преобразованиями, которые представляет собой композицию гомотетии и поворота на прямой угол.

Два нечетких отрезка являются приблизительно перпендикулярными, если после поворота одного из них на прямой угол они оказываются приблизительно параллельными.

Две нечеткие прямые приблизительно перпендикулярны, если после поворота одной из них на прямой угол, они оказываются приблизительно параллельными. Две нечеткие приблизительно перпендикулярные прямые перпендикулярны, если они образованы взаимно перпендикулярными нечеткими отрезками.

Пусть даны два нечетких отрезка

$$\tilde{A}\tilde{B},\ \tilde{C}\tilde{D} \in \tilde{A}^2.$$

Они будут перпендикулярны при условии, что для любого отрезка $MN \in \tilde{A}\tilde{B}$ обнаружится такой перпендикулярный ему отрезок $PQ \in \tilde{C}\tilde{D}$, что

$$\mu_{\tilde{A}}(M) = \mu_{\tilde{B}}(N) = \mu_{\tilde{C}}(P) = \mu_{\tilde{D}}(Q).$$

Другими словами значения функции принадлежности этих двух отрезков множеству перпендикулярных нечетких отрезков равно единице.

14.7. Нечеткая ортогональность прямых

Для трех разных точек A,B,C можно ввести понятие ортогональности прямых AB и BC . Ортогональнисть эквивалентна равенству

$$|AB|^2 + |BC|^2 = |AC|^2$$

Нечеткая ортогональнсть может быть определена приблизительным равенством

$$AB|^2 + |BC|^2 \approx |AC|^2$$

или

$$AB|^2 + |BC|^2 - |AC|^2 = \text{"около } 0\text{"}.$$

Определим отношение ортогональности:

$$a \,\tilde{\perp}\, b, a \perp \tilde{b}, \tilde{a} \perp \tilde{b}$$

таким образом:

$$\forall a, \forall b, \forall A \in a, \forall B \in b, \exists C = a \cap b : |AB|^2 + |AC|^2 \approx |AC|^2 \Rightarrow a \,\tilde{\perp}\, b$$

$$\forall a, \forall \tilde{b}, \exists\ b \subset \tilde{b} : a \perp b \Rightarrow a \perp \tilde{b}\ ;\ \mu(a \perp \tilde{b}) = \mu_{\tilde{b}}\ (b),$$

$$\forall \tilde{a}, \forall \tilde{b}, \forall a \subset \tilde{a}, \exists\ b \subset \tilde{b} : a \perp b \Rightarrow \tilde{a} \perp \tilde{b}\ ;$$

$$\mu(\tilde{a} \perp \tilde{b}) = \min\left\{\mu_{\tilde{a} \cup \tilde{b}}(a \perp b)\right\},$$

Степень ортогональности означает значение функции принадлежности, с которым две приблизительно перпендикулярные прямые относятся к множеству всех пар взаимно перпендикулярных прямых.

14.8. Нечеткая параллельность

Сформулируем понятие параллельности для четких и нечетких прямых. Как известно, в евклидовой геометрии параллельность двух прямых выражается постоянством расстояния от точки, произвольно выбранной на одной прямой, к другой прямой. Нечеткая параллельность будет определяться образцовым постоянством этого расстояния. Тогда можно определить следующие понятия:

$$a \,|\tilde{|}\, b,\ a \,||\, \tilde{b}, \tilde{a} \,||\, \tilde{b}\ .$$

$$\forall a, \forall\ b, \forall\ X \in\ a, \forall\ Y \in\ a : d(X,Y) \to \infty,\ d(X,b) \cong\ d(Y,b) \Leftrightarrow a \,|\tilde{|}\, b,$$

$$\forall a, \forall \tilde{b}, \exists b \subset \tilde{b} : a \| b \Leftrightarrow a \,\tilde{\|}\, b,$$

$$\forall \tilde{a}, \forall \tilde{b}, \forall a \subset \tilde{a}, \exists b \subset \tilde{b} : a \| b \Leftrightarrow \tilde{a} \| \tilde{b} .,$$

Первое понятие можно сформулировать иначе

$$\forall a, \forall b, \exists \tilde{b} \supset b: a \| \tilde{b} \Leftrightarrow a \,\tilde{\|}\, b.$$

Для второго и третьего понятия можно формально определить степень параллельности

$$\mu(a \| \tilde{b}) = \mu_{\tilde{b}}(b) \, \mu(\tilde{a} \| \tilde{b}) = \min\left\{\mu_{\tilde{a} \cup \tilde{b}}(a \| b)\right\},$$

которая показывает, в какой степени две приблизительно параллельные прямые могут быть отнесены множеству всех пар параллельных прямых.

Можно утверждать, что если две прямые приблизительно перпендикулярны третий, то они приблизительно параллельны.

14.9. Некоторые нечеткие фигуры

Образное мышление оперирует с более сложными фигурами, чем точки и прямые. Рассмотрим такие фигуры как нечеткие треугольники, нечеткие прямоугольники, нечеткие окружности. Сразу следует отметить, что для каждой фигуры можно дать несколько определений. Нечетким треугольником мы назовем нечеткую фигуру, которая состоит из трех нечетких отрезков, которые не принадлежат одной нечеткой прямой и которые имеют приблизительно общие нечеткие концы.
Нечетким треугольником можно назвать множество треугольников, вершины которых принадлежат трем разным нечетким точкам.
Нечетким прямоугольником можно назвать фигуру, которая состоит из четырех нечетких отрезков, из которых смежные отрезки приблизительно перпендикулярные, а противоположные - приблизительно равные.
Нечеткой окружностью можно назвать множество точек, которые

находятся приблизительно на одном расстоянии от данной точки. Или - множество точек, которые находятся на данном расстоянии от данной нечеткой точки. Или - множество окружностей, для которых приблизительно задан центр или приблизительно определен радиус. По большому счету все эти определения нечеткой окружности эквивалентны. Исследование свойств нечетких фигур не является сейчас нашей задачей. Поэтому на этом остановимся.

Нечеткое касание прямых и окружностей

Сформулируем еще отношение нечеткого касания для прямой и окружности и для двух окружностей, которые часто встречается в задачах. А именно:

~ - прямая приблизительно касается даной окружности;
~ - две окружности приблизительно касаются друг друга;
~ - нечеткая прямая касается окружности или нечеткая окружность касается прямой;
~ -нечеткая прямая касается нечеткой окружности или две нечетких окружности касаются друг друга.

Обозначим отношение четкого и нечеткого касания $T, \tilde{T}$. Тогда

$$\forall a, \forall\ \alpha, \exists\ \{A,B\}:\{A,B\}= a \cap\ a,\ d(A,B)=0 \Rightarrow\ aTa\ ,$$

$$\forall a, \forall\ \alpha, \exists\ \{A,B\}:\{A,B\}= a \cap\ a,\ d(A,B)\approx 0 \Rightarrow\ a\tilde{T}\ \ a\ ,$$

или

$$\forall a, \forall \beta, \exists A{:}A \in a, \exists B{:}B \in\ \beta, \exists\ \tilde{A}{:}A=core\,\tilde{A},\ \mu_{\tilde{A}}\ (B)\neq 0 \Rightarrow\ a\tilde{T}\ \ \beta,$$

$$\forall \alpha, \forall\ \beta, \exists A{:}A \in \alpha, \exists B{:}B \in \beta, \exists\ \tilde{A}{:}A=core\,\tilde{A},\ \mu_{\tilde{A}}\ (B)\neq 0 \Rightarrow \alpha\ \tilde{T}\ \beta,$$

$$\forall\ \tilde{a}\ , \forall\ \beta, \exists\ \ B{:}B \in\ \beta,\ \mu_{\tilde{a}}\ \ (B)\neq 0 \Rightarrow \tilde{a}\ \ T\ \ \beta,$$

$\forall\ a, \forall\ \tilde{\beta}\ , \exists\ A{:}A \in\ a, \mu_{\tilde{\beta}}\ (A) \neq 0 \Rightarrow a\ T\ \tilde{\beta}\ ,$

$\forall\ \tilde{a}, \forall\ \tilde{\beta}\ , \exists\ A{:}\ \mu_{\tilde{a}}\ (A) \neq 0, \mu_{\tilde{\beta}}\ (A) \neq 0 \Rightarrow \tilde{a}\ T\ \tilde{\beta}\ ,$

$\forall\ \tilde{\alpha}, \forall\ \tilde{\beta}\ , \exists\ A{:}\ \mu_{\tilde{a}}\ (A) \neq 0, \mu_{\tilde{\beta}}\ (A) \neq 0. \Rightarrow \tilde{\alpha}\ T\ \tilde{\beta}$

14.10. Нечеткие преобразования

Следующими объектами, без которых невозможно обойтись при моделировании мыслемых геометрических образов, есть преобразования. Любое преобразование может быть определено как множество упорядоченных пар точек. Нечеткое преобразование можно определить как множество нечетких упорядоченных пар точек или упорядоченных нечетких пар точек или упорядоченных пар нечетких точек. Исследование нечетких преобразований континуума есть самостоятельная большая задача, которое не является задачей данной работы. Поэтому из всего множества преобразований рассмотрим только преобразование ортогональной и аффинной группы, а именно: перенос, поворот, осевую и центральную симметрии, гомотетию и подобие. Как показывает опыт, эти преобразования можно представить мысленно.
Нечеткий перенос двумерного континуума можно определить следующим образом:

$$\forall\ \tilde{A}, \forall\ \tilde{B}\ , \forall\ \tilde{C}\ , \exists\ \tilde{C}\ : \overline{C\tilde{C}} \approx \overline{\tilde{A}\tilde{B}}\,.$$

Для нечеткого переноса справедлива теорема (сформулируем ее без доказательства):

$$\forall\ \tilde{A}, \forall\ \tilde{B}\ , \exists\ \overline{A\tilde{B}''}\ {:}A{=}core\,\tilde{A}\ , \tilde{B}\ {=}f(\ \tilde{B}''\ ), \overline{A\tilde{B}''}\ = \overline{\tilde{A}\tilde{B}}\ .$$

Поэтому задание любой нечеткой точки эквивалентно заданию нечеткого переноса, определенного нечетким вектором, начало которого совпадает с началом системы координат. Поворот плоскости определяется центром и углом поворота при постоянстве радиуса поворота. При нечетком повороте любые из этих параметров или все они могут быть нечеткими. Можно сформулировать и доказать теорему о том, что если все параметры некоторого нечеткого поворота являются нечеткими, то существует поворот, эквивалентный заданному, у которого нечетким есть только один параметр. Центральная симметрия есть поворот на 180 градусов. Нечеткая центральная симметрия может быть определена аналогично:

$$\forall\ \tilde{A}, \forall\ \tilde{B}\ , \exists\ \tilde{A}'\ : \overline{AB}\ \approx \overline{B\tilde{A}'}\ .$$

Определим нечеткую осевую симметрию:

$$\forall\ A, \forall\ s\ , \exists\ A'\ : d(A,s) \approx d(A',s), AA'\ \tilde{\perp}\ s.$$

Аналогично предыдущему, можно доказать теорему, что нечеткая осевая симметрия, которая определена таким образом, эквивалентна четкой осевой симметрии относительно нечеткой оси, параметры нечеткости которой есть функции параметров нечеткости осевой симметрии. Т.е.

$$\forall\ A, \forall\ s\ , \exists\ A'\ : d(A,s) \approx d(A',s), AA'\ \tilde{\perp}\ s, \Rightarrow \exists\ \tilde{A}'\ : \tilde{A}'\ = \tilde{f}_s\ (A),$$

$$\forall\ A, \forall\ \tilde{f}_s\ (A),\ \exists\ \tilde{s}\ : \tilde{f}_s\ (A) = \ f_{\tilde{s}}(A).$$

Гомотетия есть преобразования, которые сохраняют параллельность отрезков и изменяют их длину в k раз, т.е.

$$\forall\ A, \forall\ B, \forall\ O,\ \forall\, k, \exists\ (A',B'):|O\,A'|=k|O\,A|,\ |OB'|=k|OB|.$$

Нечеткую гомотетию можно определить так

$$\forall\ A, \forall\ B, \forall\ O,\ \forall\, k, \exists\ (\tilde{A}'\ , \tilde{B}'\ ):|O\,\tilde{A}'\ |\approx k|O\,A|,\ |O\,\tilde{B}'\ |\approx k|OB|,$$

что эквивалентно

$$\forall\ A, \forall\, B, \forall\, O, \forall\ \tilde{k}\ , \exists\, (\tilde{A}'\ , \tilde{B}'\ ):|O\,\tilde{A}'\ |\approx \tilde{k}\ |O\,A|,\ |O\,\tilde{B}'\ |\approx \tilde{k}\ |OB|,$$

или

$$\forall\ A, \forall\ B, \forall\ \tilde{O}, \forall\ k, \exists\,(\tilde{A}',\ \tilde{B}'):|\tilde{O}\ \tilde{A}'\ | \approx k|\tilde{O}\ A|,\ |\tilde{O}\ \tilde{B}'\ | \approx k|\tilde{O}\ B|.$$

Нечеткое подобие рассматривать не будем, так как оно является произведением любого нечеткого движения на гомотетию или любого четкого движения на нечеткую гомотетию.

Мысленные операции и мера сложности алгоритма

Если геометрический алгоритм есть последовательность геометрических операций (построений), необходимых для получения решения, то нечеткий алгоритм есть последовательность операций, из которых хотя бы одна операция есть нечеткой. В случае мысленного решения какой-нибудь задачи все подумкові операции являются нечеткими. Мы выделяем только две мысленные геометрические операции, реализованны в двумерном континууме: фиксация точки как результата пересечения двух линий, фиксация линии, заданной некоторым множеством своих параметров.
Как известно, простотой или мерой сложности геометрического алгоритма называется число элементарных геометрических операций, выполняемых в процессе решения. Мерой сложности мысленного алгоритма назовем максимальное число геометрических образов, которые необходимо одновременно удерживать в воображении при мысленном выполнении какого-нибудь шага алгоритма решения задачи.

14.11. Экспериментальное определение числовых параметров нечеткости

Для того, чтобы реализовать все описанные объекты нам необходимо определить числовые параметры нечеткости для нечетких точек двумерного континуума. Это можно сделать экспериментально при принятии некоторых ограничений и допущений, естественных для человека с нормально развитым визуальным мышлением. К таким людям мы относим инженеров-конструкторов и художников. Задачи, которые предлагались выполнить испытуемым, были следующие:

1) отметить на бумаге две горизонтально расположенные точки, расстояние между которыми попадает в интервал [10, 100], [200, 600], [700, 1000] мм. При этом указывались четкие разностные расстояния, например 5,4 мм и т.д. Предлагалось по 10 расстояний из каждого интервала;

2) то же самое, но точки должны быть расположены вертикально;

3) указать интервал, который отвечает, по их мнению, понятию «около X», где X - любое целое случайное число из интервала [10, 1000];

4) обвести в виде окружности и в виде прямоугольника область, которая отвечает, по их мнению, понятию «около точки».

Упомянутыми выше допущениями и ограничениями были:

1) точность мысленной фиксации точки в двумерном континууме составляет ± 1 мм;
2) указанные в задачах понятия зависят от размеров континуума и потому условный размер континуума принят
10 ? 10, 100 ? 100, 1000 ? 1000 мм;
3) из поля зрения испытанных были изъятые всякие предметы, по которым так или иначе можно было бы оценить указанные расстояния.

После обработки множества ответов, были сделаны следующие выводы:

1) при черчении отрезков указанной длины часть испытуемых допускала стойкие отклонения в меньшую сторону, а часть - в большую;

2) отклонение в ответах для указанного континуума можно считать пропорциональными абсолютной величине X. Для длин из интервала [10, 100], после удаления очевидных выбросов, точность составила ± 5 мм, а для длин из интервала [700, 1000] - ± 50 мм;

3) для указанного размера континуума разности в определении длин по горизонтали и по вертикалу не выявлено;

4) понятие «около X» можно представить в виде нечеткого числа (R - L)-типа, равного (0,05X - 1, X, 0,05X + 1) с кусочно-линейной функцией принадлежности;

5) понятие «около точки», которое представленно графически, не зависит от положения точки в указанном континууме и может быть изображено в виде нечеткой точки с квадратным или круговым основанием, приблизительно равными по площади. Для указанного континуума сторона квадрата или радиус окружности может быть принята равным 1/20 ... 1/100 длины стороны континуума;

6) аналитическая и графическая модели понятия «около точки» не совпадают. Следовательно, нечеткие модели аналитического и визуального мышлений не эквивалентны.

14.12. Общий алгоритм развития и контроля визуального мышления

Теперь мы можем подойти к описанию общего алгоритма развития и контроля визуального мышления в автоматизированной интеллектуальной учебной системе. Представим его в виде блок-схемы (рис. 6.9). В блоке выбора текста задачи случайным образом генерируется задача на заданную или выбранную тему. Выбор уровня нечеткости позволяет установить параметры нечеткой точки, которые наиболее удобны для того, кого учат. Выбор уровня сложности позволяет установить исходные данные в общем или частном положении в зависимости от степени подготовки того, кого учат. Выбор способа решения означает, что задача будет решаться с начала до конца или будет решаться поэтапно. В первом случае генерируется минимально необходимое для решение число исходных графических образов и всевозможные скрытые графические образы, которые возникают в процессе решения. Каждый скрытый графический образ сопровождается областью реагирования. Область реагирования на ответ представляет собой скрытую область черчения - нечеткую фигуру (точку, отрезок и т.д.), которая реагирует на щелчок кнопки мыши или на нажатие клавиши при подведении в эту область указателя. Визуально (мысленно) решив задачу и указав на экране возможное положение результирующего образа, тот, кого учат, получает реакцию программы в виде принятия ответа, если она верна, и непринятия ее в противном случае. Ответ того, кого учат, в виде указания на экране точки или нескольких точек анализируется по степени принадлежности скрытым нечетким образам. В случае их принадлежности вычисляется степень принадлежности, которая интерпретируется как точность визуального решения задачи. В случае их непринадлежности скрытым нечетким образам предлагается перейти к поэтапному решению этой же задачи. Поэтапное решение отличается только тем, что общая задача разбивается на подзадачи 1-го уровня сложности, которые выдаются в виде вопросов, которые требуют однозначного ответа. После правильного ответа, т.е. если указанная тем, кого учат, точка принадлежит скрытому нечеткому образу, генерируется видимый графический образ, который отвечает ядру нечеткого образа. Следующий этап генерируется с учетом найденного на предыдущем этапе образа.

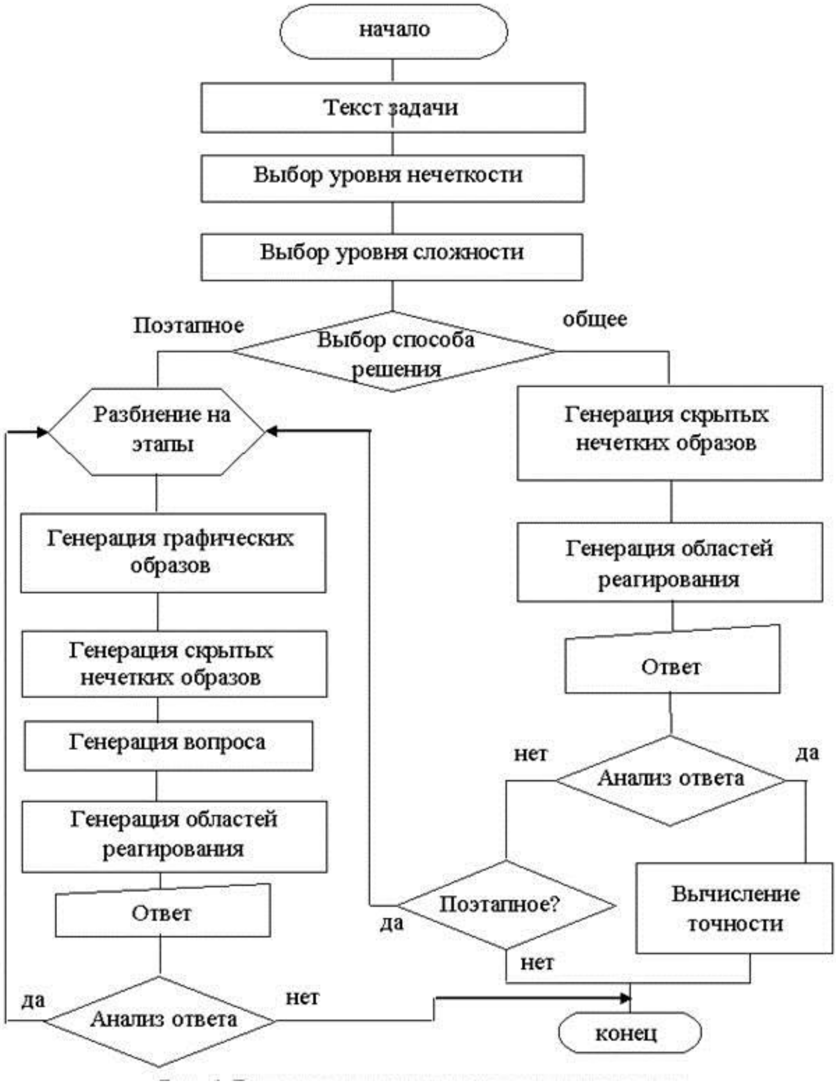


Рис. 6.9. Блок-схема алгоритма автоматизированного контроля визуального решения задач

14.13. Примеры решения задач

Рассмотрим несколько задач разного уровня сложности и их решение. Все задачи разбиты на следующие темы:

1. Планиметрические на построение для развития визуального мышления.

2. Стереометрические на построение для развития пространственного воображения.

3. Ортогонально-проективные на построение для развития мышления в области комплексного черчения.

4. Аксонометрические проективные на построение для развития пространственного воображения.

5. Перспективно-проективные для развития пространственного воображения.

Планиметрические задачи 1-го уровня сложности составили тест для проверки уровня развития визуального мышления того, кого учат,. Полученные в результате данные могут быть использованы в блоке установки параметров нечеткости. К планиметрическим задачам 1-го уровня сложности можно отнести, например, такие:

1) На данной оси с указанным единичным отрезком указать точку, координата которого равна X. Вместо X указывается конкретное число, например 5,4 или 67,3 и т.д.

2) На концах данного отрезка сосредоточены массы *A* кг и *B* кг. Указать центр тяжести отрезка.

Планиметрические задачи 2-го уровня сложности могут быть, например, такими:

1) В системе двух осей с указанными единичными отрезками указать точку с такими-то координатами.

В этой задаче оси могут быть расположены под любым углом и единичные отрезки могут быть разными. Второй уровень сложности определяется необходимостью фиксировать в визуальной памяти две прямые, которые проходят через визуально представленные точки визуально параллельно соответствующим осям. Ответом является визуально представленная точка их пересечения.

2) Дано три точки и в каждой указана сосредоточенная в ней масса. Указать центр тяжести треугольника.

Задача, в которой будут заданы N точек с массами и в которой потребуется указать их центр тяжести, тоже относится к задачам 2-го уровня сложности. Это связано тем, что в визуальной памяти на каждом шаге решения нужно удерживать только одну точку - уже визуально найденный центр тяжести некоторого числа точек и отрезок, который соединяет ее со следующей точкой.

Стереометрическую задачу четвертого уровня сложности рассмотрим подробно. Условие задачи следующее. Дана проекция тетраэдра и дано три точки *A, B, C* на трех его ребрах *a, b, c*. Указать точку пересечения ребра *d* с плоскостью *ABC* (рис. 6.10,а). Естественно, что изображение исходных данных – тетраэдра и точек на его ребрах генерируется случайным образом, т.е. точки могут быть расположены каждый раз в другом месте или на других ребрах. Точку пересечения каждый раз необходимо указывать на одном з трех ребер, которые остались. Если допустить, что выбрано общее решение задачи, то будут сгенерированы скрытые нечеткие образы, т.е. нечеткая точка $\widetilde{D}$ на ребре *d* с ядром, которое отвечает точному решению, и ее область реагирования (рис. 6.10,б). Если допустить, что выбрано поэтапное решение, то, опуская детали диалога, будет предложено указать точку пересечения прямой *AB* с ребром *f* и будут сгенерированы соответствующие скрытые нечеткие образы: нечеткая прямая $\widetilde{A}\widetilde{B} : core\widetilde{A}\widetilde{B} = AB$, нечеткая прямая $\widetilde{f} : core\ \widetilde{f} = f$, нечеткое множество $\widetilde{F} = \widetilde{A}\widetilde{B} \bigcap \widetilde{f}$ (рис.10, в).

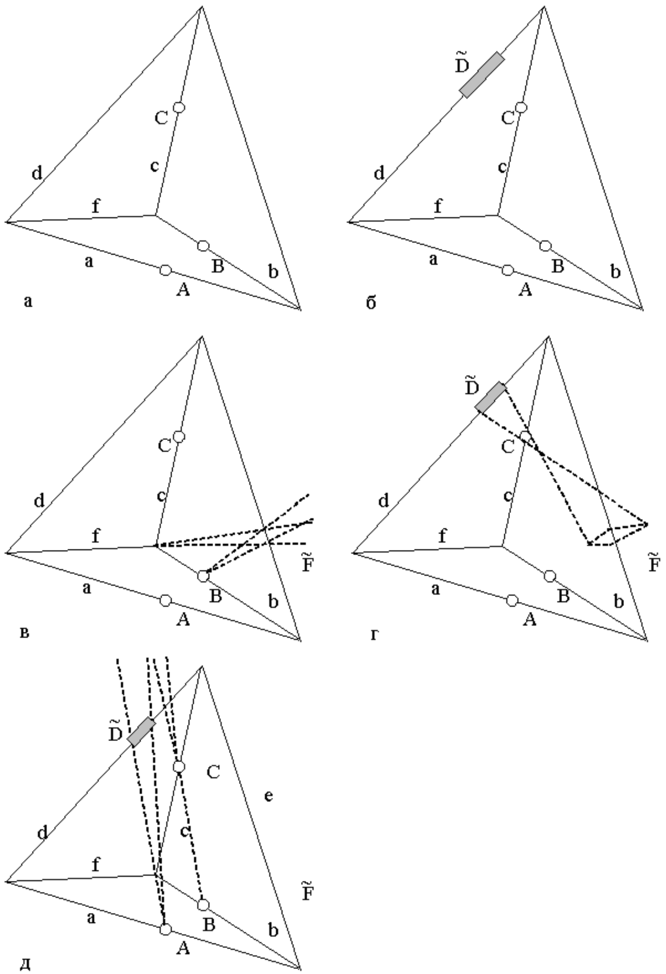


Рис. 10. Поэтапное решение задач с нечеткими образами

Дальше будет предложено представить прямую *FC* и указать ее пересечение с прямой *d*. На чертеже будут сгенерированы скрытые нечеткие образы – нечеткая прямая $\widetilde{FC}$ и нечеткая точка $\widetilde{D} = \widetilde{FC} \bigcap d$ (рис.10, г).

Можно предположить, что будет выбрано другое решение. Например, если решающий задачу захочет найти точку пересечения прямых *BC* и *e*, то будет сгенерирована нечеткая точка $BC \bigcap e = \widetilde{E}$, которая может оказаться вне поля черчения (на рисунке не показана). После этого будет сгенерирована нечеткая прямая $\widetilde{E}A$ и нечеткая точка $\widetilde{D} = \widetilde{E}A \bigcap d$ (рис. 10, д).

Литература

1. Алберг Дж., Нильсон Э., Уолш Дж. Теория сплайнов и ее приложения.—М.: Мир, 1972 (Ahlberg J.H., Nilson E.N., Walsh J.L. The Theory of Splines and their Applications.—New York: Academic Press, 1967).

2. Березин И.С., Жидков Н.П. Методы вычислений. Т. 1,2.—М.: Наука, 1962, 1966.

3. Будак Б. М., Фомин С.В. Кратные интегралы и ряды.—М.: Наука, 1967.

4. Гардан П., Люка М. Машинная графика и автоматизация конструирования.— М.: Мир, 1987 (Techniques Graphiques Interactives et C.A.O./par Michel Lucas et Yvon Gar dan.—Prance: Hermes Publishing, 1983).

5. Гильберт Д., Кон-Фоссен С. Наглядная геометрия.—М.: Наука, 1981 (Hilbert D., Cohn-Vossen S. Anschauliche Geometrie.—Berlin: 1932).

6. Де Бор К. Практическое руководство по сплайнам.—М.: Радио и связь, 1985 (De Boor C. A Practical Guide to Splines.—Berlin: Springer, 1879).

7. Демидович Б.П., Марон И.А. Основы вычислительной математики.—М.: Наука, 1970.

8. Дубровин Б.А., Новиков С.П., Фоменко А.Т. Современная геометрия.— М.: Наука, 1986.

9. Ильин В. А., Позняк Э.Г. Аналитическая геометрия.—М.: Наука, 1981.

10. Ильин В. А., Позняк Э.Г. Линейная алгебра.—М.: Наука, 1984.

11. Калиткин Н.Н. Численные методы.—М.: Наука, 1978.

12. Курош А.Г. Курс высшей алгебры.—М.: Наука, 1971.

13. Лаврентьев М. А., Шабат Б.В. Методы теории
функций комплексного переменного.—М.: Наука, 1987.

14. Ласло М. Вычислительная геометрия и компьютерная графика на С++ К—М.: Бином, 1997 (Laszlo M. J. Computational Geometry and Computer Graphics in C++.—Prentice Hall, 1996).

15. Лойцянский Л.Г., Лурье А.И. Курс теоретической механики. Т. 1,2.—М.: Наука, 1982, 1983.

16. Препарата Ф., Шеймос М. Вычислительная геометрия.—М.: Мир, 1989 (Preparata F. P., Sham os M. Computational Geometry: An Introduction.— New York, Berlin, Tokyo: Springer-Verlag, 1985).

17. Рашевский П. К. Курс дифференциальной геометрии.—М-Л.: Физматгиз, 1950.

18. Рашевский П.К. Риманова геометрия и тензорный анализ.—М.: Наука, 1967.

19. Роджерс Д., Адамс Дж. Математические основы машинной графики.—М.: Машиностроение, 1980 (Rogers D.F., Adams J. A. Mathematical Elements for Computer Graphics.—McGrow-Hill, 1976).

20. Розенфельд А. Распознавание и обработка изображений с помощью вычислительных машин.—М.: Мир, 1972.

21. Самарский А. А., Гулин А.В. Численные методы.—М.: Наука, 1989.

22. Сокольников И. С. Тензорный анализ.—М.: Наука, 1971.

23. Стечкин С.Б., Субботин Ю.Н. Сплайны в вычислительной математике.—М.: Наука, 1976.

24. Тихомиров Ю. Программирование трехмерной графики.—СПб.: BHV-Санкт-Петербург, 1998.

25. Фаворин М.В. Моменты инерции тел.—М.: Машиностроение, 1977.

26. Hilbert D., Grundziige einer allgemeincn Theorie der linearen Integralgleichungen, Lpz.- B., 1912; N. Y., 1953;
27 Besicovitch A. S., Almost periodic functions, Camb., 1932;
28. von Neumann J., "Math. Ann.", 1929, Bd 102, S. 49-131;
29. Riesz P., "Acta Sci. Math. Szeged", 1930, v. 5, № 1, p. 23-54;
[5] Дьедонне Ж., Основы современного анализа, пер. с англ., М., 1964;
30. Бурбаки Н., Топологические векторные пространства, пер. с франц., М., 1959;
31. Ахиезер Н. И., Глазман И. М., Теория линейных операторов в гильбертовом пространстве, 2 изд., М., 1966;
32. Carleman T., Sur les equations integrates singulieres a noyau reel et symetrique, Uppsala, 1923;
[9] Ахиезер Н. И., Классическая проблема моментов и некоторые вопросы анализа, связанные с нею, М., 1961;
33. Данфорд Н., Шварц Д ж., Линейные операторы, пер. с англ., М., 1962;
34. Рисе Ф., Секефальви-Надь Б., Лекции по функциональному анализу, пер. с франц., М., 1954;
35. Наймарк М. А., Линейные дифференциальные операторы, 2 изд., М., 1969;
36. Stone M., Linear transformation in Hilbert space and their applications to analisis, N. Y., 1932;
37. Колмогоров А. Н., Фомин С. В., Элементы теории функций и функционального анализа, 3 изд., М., 1972.
38. Дэй М. М., Нормированные линейные пространства, пер. с англ., М., 1961;
39. Бурбаки Н., Топологические векторные пространства, пер. с франц., М., 1959.