

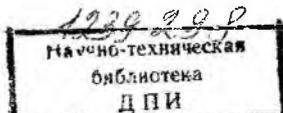
Ю. С. Завьялов,
В. А. Леус,
В. А. Скороспелов

**СПЛАЙНЫ
В ИНЖЕНЕРНОЙ
ГЕОМЕТРИИ**

« МАШИНОСТРОЕНИЕ »

517 ББК 32. 816
3-13 3-13
УДК 681.332.51:621.81.001.2

Рецензент канд. техн. наук В. С. Полозов



Завьялов Ю. С. и др.

З-13 Слайды в инженерной геометрии/Ю. С. Завьялов, В. А. Леус, В. А. Скорospelов. — М.: Машиностроение, 1985. — 224 с., ил.

(В пер.: 1 р. 20 к.

Изложены математические основы обработки геометрической информации на ЭВМ при проектировании и технологической подготовке производства машиностроительных деталей различной формы. Приведены алгоритмы и изложены принципы организации автоматизированной системы обработки данных.

Книга предназначена для инженеров-конструкторов и технологов, а также математиков-программистов, работающих в области автоматизации обработки данных.

3 1502000000-009 9-85
038 (01)-85

ББК 32.816
6Ф0.1

ПРЕДИСЛОВИЕ

Методы математического моделирования сложных геометрических объектов сплайнами широко применяются в инженерной практике различных областей техники. Однако обобщающего труда, где были бы изложены теоретические основы аппроксимации кривых и поверхностей сплайнами, а также вопросы их использования для решения конкретных задач, в мировой литературе нет. Настоящую книгу следует рассматривать как попытку восполнить этот пробел.

Содержание книги отражает двадцатилетний опыт работы авторов по применению сплайновых методов для проектирования и технологической подготовки производства в машиностроении. В книге представлены основы теории аппроксимации полиномиальными сплайнами. Главное внимание уделено кубическим сплайнам одной и двух переменных, которые, с одной стороны, достаточно просты по своей численной реализации на ЭВМ, а с другой — полностью обеспечивают потребности инженерной геометрии.

Центральное понятие теории приближенных вычислений — близость объектов — для кривых и поверхностей определить труднее, чем для функций. Общим вопросам этого плана и конкретным способам моделирования кривых и поверхностей сплайнами посвящены две главы книги, в которых наряду с кубическими сплайнами рассмотрен один тип нелинейных сплайнов. Этот математический аппарат является базой для решения основных задач инженерной геометрии, таких как расчет касательных и нормалей, длин линий, площадей и объемов, расстояний между объектами и их взаимопересечений, построение разверток и квазиразверток поверхностей на плоскость. В книге отражена методика расчета траекторий режущего инструмента при обработке сложных поверхностей на станках с ЧПУ, затронуты вопросы оптимизации обработки сложных деталей. Одна глава книги посвящена проблеме численной визуализации трехмерных геометрических объектов. Речь идет о представлении информационных моделей в привычном человеческому глазу виде. Предлагаемые методы не зависят от способа описания кривых и поверхностей, но их численная реализация наиболее экономична при использовании сплайновых аппроксимаций.

В последней главе указаны пути создания автоматизированной системы для обработки геометрической информации, дано краткое

описание программного обеспечения и его применения в конкретных предметных областях.

Разные части книги написаны с неодинаковой подробностью изложения. Более детально освещены вопросы, ранее затрагивавшиеся только в журнальных статьях, и менее подробно те, которые нашли отражение в монографиях. Так, по теории сплайн-функций читателю рекомендуются работы [1, 4, 8, 21, 29]. В целом для чтения книги достаточно математической подготовки в пределах общих вузовских курсов по инженерным специальностям. Но отдельные параграфы, например п. 2 гл. II, требуют больших усилий для понимания, впрочем, при первом чтении их можно опустить.

Структурой и содержанием предлагаемой книги авторам хотелось подчеркнуть специфику геометрических вычислений и несводимость этой ветви прикладной математики к традиционным разделам численного анализа. Насколько это удалось — судить читателю.

В список литературы включены источники, непосредственно использованные авторами при написании книги, а также те, которые близко примыкают по содержанию к данной книге, расширяют и углубляют рассмотренные в ней вопросы. Весьма полная библиография зарубежных работ по применению сплайнов в инженерной геометрии приведена в монографии [26].

В процессе работы над книгой авторы имели полезные обсуждения отдельных вопросов со своими коллегами по Институту математики СО АН СССР: А. А. Капланом, В. Л. Мирошниченко, В. А. Топоноговым и другими. В книге использованы опубликованные ранее в изданиях института результаты по применению сплайн-функций Г. В. Вайсберга, В. В. Вершинина, В. М. Гамидова, Н. Н. Павлова, П. А. Турук, Б. М. Шумилова; по теории графов Э. Х. Гимади, В. А. Перепелицы, А. И. Сердюкова; по технологии машиностроения А. В. Кана, Ф. А. Колсановой, А. А. Пономаренко, Г. В. Солодовниковой.

ВВЕДЕНИЕ

Современное состояние инженерной геометрии. На всех этапах создания новых изделий — от проектирования до изготовления, приходится решать разнообразные геометрические задачи. В одних областях техники эти задачи играют подчиненную роль, в других — функциональные качества изделия решающим образом зависят от внешних форм отдельных узлов и взаимной их компоновки. Особенно важны задачи формообразования в проектировании аэро- и гидродинамических обводов агрегатов летательных аппаратов, корпусов судов, рабочих колес, направляющих и отводящих каналов турбин. Здесь ни одна из существенных физических и технологических задач не может быть решена в отрыве от разработки формы. Существует специфическая область — картография, где выдача сложной геометрической информации вообще является конечной целью производства.

От формы изделия зависит его эстетическое восприятие, которое может меняться под воздействием моды и других факторов. Прагматическая и эстетическая компоненты входят в геометрию различных изделий в неодинаковых пропорциях. Иногда они достигают полного единства, например в совершенных обводах современного воздушного лайнера, а иногда отдельные детали конструкций могут не обладать эстетическим воздействием, но выполнять важные функции. Эстетические взгляды конструктора имеют особое значение при проектировании архитектурных сооружений, кузовов легковых автомобилей и, разумеется, модных моделей одежды и обуви.

Создание геометрических форм продолжается и на этапе подготовки производства, когда проектируется и изготавливается различная технологическая оснастка, включающая материальные носители геометрической информации: шаблоны, эталоны, ступени — в машиностроении, раскройные лекала, колодки и пресс-формы — в швейной и обувной отраслях.

Потребности современной техники необычайно расширили диапазон используемых геометрических форм. В последнее десятилетие появились задачи, связанные с автоматической реализацией самого процесса формообразования на станках с числовым программным управлением. Наиболее трудоемкой из этих задач является расчет траектории обработки сложной поверхности с учетом требуемой

точности, геометрии режущего инструмента и кинематических особенностей станка.

Однако в целом проблема сложного формообразования не нова. На всех этапах развития технологии люди более или менее успешно справлялись с конкретными геометрическими задачами. С древнейших времен, задолго до рождения геометрии как науки, были выработаны приемы воспроизведения поверхностей, которые мы теперь называем интерполяционными. Так, интерполяционный прием, заключающийся в построении куска поверхности по заданной форме его границ, был освоен с изобретением гончарного круга. Роль образующей здесь играла рука гончара, интерполяция реализовалась механическим перемещением по направляющим окружностям. Сегодня этот тип формообразования развился в так называемый кинематический способ задания поверхностей.

Способ интерполяции поверхности по заданному каркасу задолго до появления цельнометаллических судов и самолетов применялся у северных народов при изготовлении лодок-каяков, представляющих собой обтянутый шкурами каркас из костей крупных животных.

Напомним один старинный способ изготовления лодки из цельного ствола дерева. Прежде обрабатывают наружную поверхность корпуса, что является сравнительно простым делом. Трудность заключается в том, чтобы при удалении внутреннего материала добиться равномерной толщины бортов и дна. С этой целью всю наружную поверхность корпуса покрывают сетью тонких отверстий, засверленных на фиксированную глубину. В отверстия забивают пробки из дерева контрастного цвета. Затем при срезании слоя за слоем материала внутри корпуса следят за появлением цветных пятнышек, что служит сигналом достижения нужной толщины стенки. Подобный способ есть не что иное, как интерполяция точно заданной поверхности.

Ныне отрасль знания, к которой относятся перечисленные выше задачи, претерпела столь существенные качественные и количественные изменения, что появились высказывания о рождении новой научно-технической дисциплины — *инженерной геометрии*. Учитывая историю техники, авторы не разделяют подобной радикальной оценки. Инженерная геометрия как раздел геометрии, дающий необходимые математические средства для описания форм изделий, их изображения и воспроизведения, не нова. С другой стороны, было бы ошибочным отождествлять исторические традиции с современным подходом к решению пусть даже и известных инженерно-геометрических задач.

Своеобразие нынешнего состояния старой отрасли знания — инженерной геометрии — состоит, по нашему мнению, в изменении природы ее методов. Первоначально методика решения инженерно-геометрических задач была неделима от самого производства и имела по необходимости механическую природу. В последующем произошло разделение процесса на проектирование, технологическую подготовку и собственно производство, причем на первых двух этапах в качестве моделирующего инструмента стали использоваться графические приемы: начертательная геометрия и техническое черчение.

На современном этапе с развитием вычислительной техники появилась возможность решать инженерно-геометрические задачи путем численного моделирования. В памяти компьютера создаются на цифровой основе геометрические объекты и геометрические отношения, адекватные реальным, а решение сводится к численному оперированию с информационными, или, как принято говорить,

математическими моделями. В этой связи за рубежом получил распространение термин «вычислительная геометрия» (*computational geometry*), но авторы не находят достаточных оснований для его употребления. Существо дела заключается в дополнении инструментария инженерной геометрии новыми методами численно-информационной природы. Трансформация основ инженерно-геометрической методологии не просто облегчает традиционную работу. Новая методика позволяет ставить задачи, о которых прежде и не помышляли, и дает средства для их решения.

Сплайн-функции — универсальный математический аппарат инженерной геометрии. Для обработки информации на ЭВМ должна быть сформирована математическая модель поверхности изделия. Это сравнительно нетрудно сделать, когда сложная поверхность может быть разбита на элементы, каждый из которых представляет собой простейший геометрический объект: круговой цилиндр или конус, сферу или участок плоскости, ограниченный отрезками прямых или дугами конических сечений. Так обстоит дело в общем машиностроении. В тех случаях, когда поверхности не образуются простейшими элементами, как, например, в авиастроении, приходится искать иное решение.

Впервые идею математического описания поверхностей агрегатов самолета высказал и применил на практике в 30-х годах известный советский авиаконструктор А. Бартини. Специалистам хорошо знакомы кривые Бартини. В последующие десятилетия для этих целей использовались главным образом определенного вида аналитические кривые и поверхности, например кривые и поверхности второго порядка, выбором параметров которых удается описать требуемую форму деталей и агрегатов. Многообразие способов (а их известно несколько десятков) говорит о том, что все они носят узко предметный характер и непригодны в качестве основы даже для отраслевых систем обработки геометрической информации с помощью ЭВМ. Математический аппарат для инженерно-геометрических расчетов самого широкого назначения дала теория сплайн-функций.

Английское слово *spline* означает «упругая рейка». Такую рейку используют в качестве гибкого лекала при вычерчивании плоских кривых по опорным точкам. Форма осевой линии рейки на участке между двумя соседними опорами описывается в прямоугольных координатах при соответствующем выборе единиц измерения уравнением равновесия:

$$y'' [1 + (y')^2]^{-3/2} = M(x). \quad (1)$$

Здесь слева стоит выражение кривизны кривой, а справа — изгибающий момент.

При горизонтальном расположении опор и небольших прогибах, когда квадратом первой производной можно пренебречь по сравнению с единицей, уравнение (1) упрощается до формы $y'' = M(x)$. При условии интегрируемости функции $M(x)$ решение этого уравнения получается двумя квадратурами и имеет вид

$$y = \int_0^x \left(\int_0^t M(s) ds \right) dt + Cx + D. \quad (2)$$

Константы C и D определяются заданием двух граничных условий в концах участка. Поскольку каждая опора является концом предшествующего участка и началом последующего, значение момента в опоре является одновременно граничным условием для обоих участков. Таким образом, на стыках действует условие сопряжения, обеспечивающее правильную стыковку соседних участков. Описания формы кривой на каждом из участков имеют одну и ту же структуру (2) и различаются только значениями числовых констант. В частности, если момент линейно зависит от x , то решением на каждом участке будет полином третьей степени.

Абстрагируясь от задачи о равновесной форме рейки, приходим к понятию функции с кусочной структурой и повторяющимся на каждом звене строением, но с различными значениями параметров. Такие функции и их обобщения на случай нескольких переменных получили название *сплайн-функций* или просто *сплайнов*.

Наиболее употребительными в вычислительных методах являются полиномиальные сплайны, хотя нередко применяются и неполономиальные. По сравнению с другими математическими конструкциями, используемыми для описания сложных геометрических форм, сплайны обладают по меньшей мере тремя важными преимуществами: во-первых, лучшими аппроксимативными свойствами, что при равных информационных затратах дает большую точность или равную точность при менее информативных исходных данных; во-вторых, простотой реализации полученных на их основе алгоритмов на ЭВМ и, наконец, в-третьих, — универсальностью, позволяющей использовать одни и те же аппроксимирующие конструкции для различных геометрических объектов.

Универсальность сплайнов не означает, что их следует применять во всех случаях жизни. Например, в общем машиностроении сплайны используются эпизодически, но существуют области, где они являются незаменимым средством описания объектов и инструментов геометрических расчетов. Прежде всего следует назвать такие отрасли машиностроения, как судостроение и авиастроение, производство автомобилей, турбин [10, 13, 20, 27, 28, 30, 31]. Сплайны нашли применение также в легкой промышленности [2, 15, 30], архитектуре [30], в геологии при построении карт в связи с задачами поиска нефти и газа [5]. Показана возможность применения сплайнов для сжатия информации при передаче фотоснимков по каналам связи и обработке речевых сигналов в реальном масштабе времени [19].

Автоматизированные системы обработки геометрической информации. В нашей стране эти системы впервые появились в начале 60-х годов. Первый период характеризовался созданием систем, автоматизирующих традиционно сложившиеся методы и приемы в проектировании и на производстве. Например, в судостроении и авиастроении в первую очередь автоматизация коснулась весьма трудоемких, но легко формализуемых плазовых работ [13, 20, 31]. Другим направлением стало применение вычислительной техники и устройств машинной графики для изготовления технической доку-

ментации [6, 16, 30]. Из нетрадиционных работ этого периода важное место заняла автоматизация расчета управляющих программ для оборудования с ЧПУ [13, 20, 27, 28, 31]. Подобные системы, с одной стороны, автоматизировали лишь отдельные звенья производственного цикла, а с другой, — как правило, были приспособлены к условиям конкретного производства. Это сужало область применения разработок и ограничивало возможности кооперации в их развитии.

В начале 70-х годов стала очевидной необходимость создания интегрированных систем, обеспечивающих автоматизацию проектирования и технологической подготовки производства в едином цикле. При создании таких систем подразумеваются определенные изменения технологии в направлении стандартизации технической информации и усовершенствования производственных процессов за счет применения ЭВМ и оборудования с ЧПУ.

Совершенствование изделий, связанное с использованием новых научных данных и поиском качественно новых инженерных решений, требует систематического внесения изменений в систему. Поэтому наиболее перспективным является создание адаптивной системы, способной настраиваться на решение задач того или иного производства. В ее основе лежит автоматизация геометрических расчетов. Базовый уровень системы не зависит от конкретной предметной области. Решение всех других задач, и конструкторских и технологических, опирается на эту основу и проводится на единой информационной базе и едиными средствами обработки информации. Система может быть адаптирована к условиям конкретных предметных областей (производство самолетов, турбин, автомобилей) и к изменяющимся условиям внутри области. Адаптация заключается в формулировании специальных задач (конструкторских, физических, технологических) и программной реализации алгоритмов их решения на основе базового уровня. Возможность такого приспособления обеспечивает не только универсальность, но и живучесть системы. В результате адаптации получается проблемно-ориентированный вариант системы, причем это достигается несравненно меньшими затратами, нежели создание новой системы.

Глава I

ЭЛЕМЕНТЫ ТЕОРИИ СПЛАЙН-ФУНКЦИЙ

В этой главе изложены основные понятия теории сплайнов и достаточно простые методы сплайн-аппроксимации. Рассмотрены задачи аппроксимации сплайнами невысоких степеней (первой и третьей), наиболее употребительные на практике, среди них интерполяция, квазиинтерполяция и сглаживание экспериментальных данных. Этому предшествуют сведения об аппроксимации многочленами.

Материалы главы дают теоретическое обоснование рабочего аппарата для моделирования кривых и поверхностей сложной формы, чему посвящены последующие главы книги.

1. АППРОКСИМАЦИЯ ФУНКЦИЙ МНОГОЧЛЕНАМИ

Классы функций. На отрезке $[a, b]$ вещественной оси будем рассматривать множество B , элементы которого суть непрерывные функции. Если $f(t)$ и $g(t)$ — две функции из этого множества, а c — вещественное число, то $cf(t)$ и $f(t) + g(t)$ тоже являются его элементами. Говорят, что B есть *линейное пространство*, а его элементы — *точки линейного пространства*.

Расстояние $\rho(f, g)$ между двумя точками пространства или, как говорят, *метрика*, вводится как вещественная функция пары точек, принимающая неотрицательные значения и удовлетворяющая следующим аксиомам метрики:

- а) $\rho(f, g) = 0$, тогда и только тогда, когда $f = g$.
- б) $\rho(f, g) = \rho(g, f)$ (симметрия);
- в) $\rho(f, r) \leq \rho(f, g) + \rho(g, r)$ (неравенство треугольника).

Линейное пространство с метрикой называется *метрическим пространством*.

В множестве B введем расстояние формулой

$$\rho(f, g) = \|f(t) - g(t)\| = \max_{t \in [a, b]} |f(t) - g(t)|.$$

Это расстояние удовлетворяет всем аксиомам метрики и делает B метрическим пространством.

Каждый элемент $f(t)$ метрического пространства можно характеризовать числом $\|f(t)\|$ (расстоянием до нулевого элемента), называемым *нормой элемента*.

В пространстве B будем рассматривать *классы $C^m[a, b]$ функций*, имеющих на отрезке $[a, b]$ непрерывные производные до m -го порядка включительно. Если непрерывны производные до $(m - 1)$ -го

порядка, а m -е производные кусочно-непрерывны, то такой класс обозначаем $C^{(m)} [a, b]$. Значения функций $f(t)$ и их производных $f^{(m)}(t)$ из этих классов можно представлять в виде разложений по формуле Тейлора:

$$f(t) = f(t_k) + f'(t_k)(t - t_k) + \dots + \frac{f^{(m-1)}(t_k)}{(m-1)!} (t - t_k)^{m-1} + \eta_m(t),$$

$$\text{где } \eta_m(t) = \frac{1}{(m-1)!} \int_{t_k}^t (t - \zeta)^{m-1} f^{(m)}(\zeta) d\zeta.$$

Пространства многочленов. Многочлены

$$P_n(t) = a_0 + a_1 t + \dots + a_n t^n$$

являются элементами пространства B . Множество M_n многочленов степени не выше n само является линейным пространством, или, как говорят, подпространством в B . Функции $1, t, \dots, t^n$, из которых конструируются многочлены степени n , *линейно независимы*, т. е. $P_n(t) \neq 0$, если хотя бы один коэффициент $a_k \neq 0$. Их число $n + 1$ есть *размерность пространства M_n* , а сами они образуют *базис* последнего.

Вообще, *базисом $n + 1$ -мерного функционального пространства* может быть любая совокупность $n + 1$ его линейно независимых функций. Так, в множестве M_n базисом является следующая система функций. Разобьем отрезок $[a, b]$ на n частей точками $t_0 < t_1 < \dots < t_n = b$ и построим многочлены

$$\Phi_k(t) = \prod_{j=0, j \neq k}^n \frac{t - t_j}{t_k - t_j}, \quad k = 0, \dots, n. \quad (3)$$

Они линейно независимы и образуют базис в M_n , т. е. любой многочлен из M_n можно представить в виде $P_n(t) = c_0 \Phi_0(t) + \dots + c_n \Phi_n(t)$.

Так как

$$\Phi_k(t_i) = \begin{cases} 1, & \text{если } k = i; \\ 0, & \text{если } k \neq i, \end{cases}$$

то $c_k = P_n(t_k)$, $k = 0, \dots, n$.

Задачи *аппроксимации функций* многочленами (а затем и сплайнами) будем характеризовать по их содержанию, точности приближения и специфике вычислений.

Типы задач аппроксимации. Простейшим методом аппроксимации является *интерполяция*. В узлах $t_k \in [a, b]$ заданы значения функции $f_k = f(t_k)$, $k = 0, \dots, n$. Требуется найти многочлен $P_n(t)$ из множества M_n , удовлетворяющий условиям $P_n(t_k) = f_k$, $k = 0, \dots, n$. Решение задачи дает *интерполяционный многочлен Лагранжа*

$$L_n(t) = f_0 \Phi_0(t) + \dots + f_n \Phi_n(t).$$

Его можно представить также через *разделенные разности*.

зуют пространство размерности $(n + 1)^2$. В нем можно использовать базисы как из функций $t^k s^l$, $k, l = 0, \dots, n$, так и из функций $\Phi_k(t) \Phi_l(s)$, $k, l = 0, \dots, n$. В последнем случае

$$P_n(t, s) = \sum_{k=0}^n \sum_{l=0}^n c_{kl} \Phi_k(t) \Phi_l(s);$$

где $c_{kl} = P_n(t_k, s_l)$.

Для многочленов $P_n(t, s)$ можно ставить те же задачи, что и в случае одной переменной, но свойства решений одномерных задач сюда не переносятся. Например, решение задачи интерполяции существует не при всяком расположении узлов.

Метод наименьших квадратов здесь состоит в минимизации

$$e_2 = \sum_{v=1}^{N_*} \sum_{\mu=1}^{M_*} q_{v\mu} [P_n(t_v, s_\mu) - f(t_v, s_\mu)]^2.$$

В качестве весов $q_{v\mu}$ можно принять площади, приходящиеся на точки (t_v, s_μ) . Если веса выбираются иначе, то снова может проявиться неустойчивость метода. Следует стремиться к равномерности распределения данных по области Ω и постоянству весов. В общем случае вводятся регуляризирующие функционалы, например, вида

$$\iint_{\Omega} \{ [D^{2,0} P_n(t, s)]^2 + 2 [D^{1,1} P_n(t, s)]^2 + [D^{0,2} P_n(t, s)]^2 \} dt ds.$$

Погрешность методов аппроксимации характеризуется величиной $\|R_n(t, s)\| = \|P_n(t, s) - f(t, s)\|$.

Вычислительный аспект. При решении задач аппроксимации многочленами сталкиваются с несколькими трудностями. Это, во-первых, трудности решения систем линейных уравнений при больших n . Они непомерно возрастают при переходе к двум и более переменным. Во-вторых, при работе с многочленами в процессе вычислений происходит быстрое накопление погрешностей округлений. Они велики уже при $n \approx 20$, но требуется n часто в несколько раз большее. Следует отметить также, что при использовании многочленов высоких степеней их графики, как правило, имеют осцилляции.

От этих недостатков свободны многочлены невысоких степеней и сплайны как кусочно-многочленные функции таких степеней (в данной книге не выше третьей), а также сплайны, конструируемые из других элементарных функций. Поэтому мы рассматриваем задачи аппроксимации многочленами первой, второй и третьей степеней, которые имеют и самостоятельное значение, и входят как элементы в решения задач приближения сплайнами.

2. АППРОКСИМАЦИЯ МНОГОЧЛЕНАМИ ПЕРВОЙ, ВТОРОЙ И ТРЕТЬЕЙ СТЕПЕНИ

Линейная интерполяция. Это простейший метод приближения функций. Он широко применяется на практике. В частности, различные таблицы составляются так, чтобы промежуточные значения

функции с принятой точностью можно было вычислить с помощью линейной интерполяции.

Итак, пусть заданы значения $f_0 = f(t_0)$ и $f_1 = f(t_1)$. Если обозначить $h = t_1 - t_0$ и ввести параметр $u = (t - t_0)/h$, $0 \leq u \leq 1$, то базисные функции (3) будут $\Phi_0(t) = (1 - u)$, $\Phi_1(t) = u$, и интерполяционная формула принимает вид (рис. 1)

$$L_1(t) = (1 - u)f_0 + uf_1. \quad (5)$$

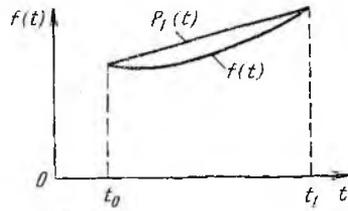


Рис. 1. Линейная интерполяция

На этом примере опишем технику получения оценок погрешности аппроксимации функций многочленами невысоких степеней, когда в аппроксимационной формуле присутствует небольшое число значений функции (в данном случае два).

Процесс получения оценки погрешности разбивается на четыре этапа. Пусть $f(t) \in C_{[a, b]}^{(2)}$.

1. Остаточный член интерполяции

$$R_1(t) = L_1(t) - f(t) = (1 - u)f_0 + uf_1 - f(t).$$

Значения f_0, f_1 заменим их разложениями по формуле Тейлора (с переменной ролей t и t_n) при $m = 2$:

$$f_0 = f(t) - hf'(t) + \int_t^{t_0} (t_0 - \zeta) f''(\zeta) d\zeta;$$

$$f_1 = f(t) + h(1 - u)f'(t) + \int_t^{t_1} (t_1 - \zeta) f''(\zeta) d\zeta.$$

Подставив эти выражения в $R_1(t)$, получаем интегральное представление остаточного члена интерполяции

$$R_1(t) = (1 - u) \int_{t_0}^t (\zeta - t_0) f''(\zeta) d\zeta + u \int_t^{t_1} (t_1 - \zeta) f''(\zeta) d\zeta.$$

В каждом из интегралов сделаем замену переменной интегрирования $\zeta = t_0 + h\tau$, $0 \leq \tau \leq 1$. Тогда

$$R_1(t) = h^2 \left[(1 - u) \int_0^u \tau f''(\zeta) d\tau + u \int_u^1 (1 - \tau) f''(\zeta) d\tau \right].$$

2. Отсюда следует

$$|R_1(t)| \leq h^2 \left[(1 - u) \int_0^u \tau |f''(\zeta)| d\tau + u \int_u^1 (1 - \tau) |f''(\zeta)| d\tau \right].$$

Заменяем $|f''(\zeta)|$ на $\max_{t_0 \leq t \leq t_1} |f''(t)| = \|f''(t)\|_{[t_0, t_1]}$, где символ $\|\cdot\|_{[t_0, t_1]}$ обозначает, что норма вычислена на отрезке $[t_0, t_1]$. Не-

Разделенная разность нулевого порядка — это значение функции в точке $f(t_0)$. Разделенная разность k -го порядка определяется через разделенные разности $k-1$ -го порядка по формуле

$$f[t_0, \dots, t_k] = \frac{f[t_1, \dots, t_k] - f[t_0, \dots, t_{k-1}]}{t_k - t_0}.$$

Если производная $f^{(k)}(t)$ непрерывна, то верна формула

$$f(t_0, \dots, t_k) = \frac{f^{(k)}(\xi)}{k!},$$

где ξ — некоторая точка из отрезка $[t_0, t_k]$.

Теперь интерполяционный многочлен перепишем в форме Ньютона:

$$L_n(t) = f_0 + (t - t_0)f[t_0, t_1] + \dots + (t - t_0) \dots (t - t_{n-1})f[t_0, \dots, t_n]. \quad (4)$$

Важный круг задач составляют задачи о *наилучших приближениях* непрерывных функций: равномерном (чебышевском) и среднем квадратическом. Они весьма сложны по реализации. Один из путей уменьшения трудностей состоит в построении формул аппроксимации многочленами $P_n(t)$, только похожих на наилучшие, близких к ним (пример приведен ниже в п. 2). Другой путь заключается в решении задач приближения не на отрезке $[a, b]$, а на дискретном множестве точек $\{t_\nu, \nu = 1, \dots, N_*\} \in [a, b]$, $N_* > n + 1$. Опишем дискретный аналог среднего квадратического приближения.

Метод наименьших квадратов. Требуется найти многочлен $P_n(t)$, минимизирующий *функционал*

$$e_2 = \sum_{\nu=1}^{N_*} q_\nu [P_n(t_\nu) - f(t_\nu)]^2,$$

где q_ν — веса. Коэффициенты многочлена определяются из равенств, представляющих собой *необходимые условия минимума функционала* $de_j/dc_j = 0$, $j = 0, \dots, n$. Если обозначить

$$s_{jk} = \sum_{\nu=1}^{N_*} q_\nu \Phi_j(t_\nu) \Phi_k(t_\nu); \quad r_j = \sum_{\nu=1}^{N_*} q_\nu \Phi_j(t_\nu) f(t_\nu),$$

то они дают систему уравнений вида

$$s_{j0}c_0 + \dots + s_{jn}c_n = r_j, \quad j = 0, \dots, n.$$

Определитель этой системы называется *определителем Грамма*. Если все исходные данные равноценны, то можно положить

$$q_1 = (t_1 - t_2 - 2a)/2; \quad q_{N_*} = (2b - t_{N_*-1} - t_{N_*})/2; \\ q_\nu = (t_{\nu+1} - t_{\nu-1})/2; \quad \nu = 2, \dots, N_* - 1,$$

и выражение e_2 есть квадратурная формула прямоугольников.

В общем случае $\tilde{f}(t_\nu)$ известны с погрешностями, т. е. известны $\tilde{f}(t_\nu)$ такие, что $|\tilde{f}(t_\nu) - f(t_\nu)| \leq \delta_\nu$. Веса q_ν выбирают с учетом погрешностей. Чем меньше δ_ν , тем больше q_ν , и точка «закрепляется». И наоборот. Таким образом обеспечивается гибкость в применении метода. Однако при этом определитель системы может стать малой

величиной, в результате чего вычисления становятся неустойчивыми [3] по отношению к погрешностям в исходных данных и погрешностям округлений. Такой опасности нет при равномерном распределении точек t_v на отрезке $[a, b]$ и изменении лишь отдельных весов q_v . В противном случае вместо классического варианта метода следует применять его модификации с добавлением к e_2 так называемого *регуляризирующего функционала*. Выбор последнего зависит от специфики конкретной задачи. Например,

$$J(P_n) = e_2 + \int_a^b [P_n''(t)]^2 dt.$$

При такой постановке задачи реализуется некоторый компромисс между стремлением обеспечить близость значений многочлена $P_n(t_v)$ к заданным значениям $\tilde{f}(t_v)$ и минимизировать кривизну его графика.

Погрешность аппроксимации и сходимость. *Погрешностью метода аппроксимации* называется расстояние между приближающим многочленом и аппроксимируемой функцией в заданной метрике. Если обозначить через $R_n(t) = P_n(t) - f(t)$ *остаточный член аппроксимации*, то погрешность будет $\|R_n(t)\|$. Эта величина зависит от степени многочлена и дифференциальных свойств функции $f(t)$. При увеличении степени многочлена получаем последовательность аппроксимаций. Если $\|R_n(t)\| \rightarrow 0$, при $n \rightarrow \infty$, то говорят, что *аппроксимационный процесс сходится*, в противном случае — *расходится*. Например, последовательность наилучших приближений непрерывной функции всегда сходится. В то же время сходимость интерполяционных многочленов Лагранжа гарантируется далеко не во всех случаях.

Случай двух независимых переменных. В области Ω двух переменных t и s рассматривается пространство B вещественных непрерывных функций. Метрика в нем вводится формулой

$$\rho(f, g) = \max_{(t, s) \in \Omega} |f(t, s) - g(t, s)|,$$

а элемент $f(t, s)$ характеризуется нормой $\|f(t, s)\| = \rho(f, 0)$.

Через $C^{m, m}[\Omega]$ обозначим множество непрерывных функций $f(t, s)$, имеющих непрерывные частные и смешанные производные $D^{p, q}f(t, s)$, $p, q \leq m$. Класс $C^{(m, m)}[\Omega]$ состоит из функций с непрерывными производными до порядка $p, q \leq m - 1$ и кусочно-непрерывными производными до порядка $p, q \leq m$.

Многочлены двух переменных

$$P_n(t, s) = \sum_{k+l=0}^n c_{kl} t^k s^l,$$

образуют пространство (подпространство в B) размерности $(n + 1)(n + 2)/2$ с базисными функциями $t^k s^l$, $k + l = 0, \dots, n$.

В теории сплайнов наряду с такими многочленами рассматриваются многочлены степени n по каждой из переменных. Они обра-

равенство от такой замены усиливается. Вынося $\|f''(t)\|_{[t_0, t_1]}$ за знаки интегралов, получаем

$$|R_1(t)| \leq h^2 \left[(1-u) \int_0^u \tau d\tau + u \int_u^1 (1-\tau) d\tau \right] \|f''(t)\|_{[t_0, t_1]}.$$

3. Вычисляя интегралы, находим поточечную оценку погрешности (в зависимости от t) в виде

$$|R_1(t)| \leq 0,5h^2u(1-u) \|f''(t)\|_{[t_0, t_1]}. \quad (6)$$

Правая часть этого неравенства является мажорантой для модуля остаточного члена.

4. Из неравенства (6) выводится оценка максимальной погрешности (погрешности по норме). Для этого достаточно подсчитать максимум функции $u(1-u)$. Он достигается в середине отрезка при $u = 0,5$ и равен $0,25$. Окончательно

$$\|R_1(t)\|_{[t_0, t_1]} \leq 0,125h^2 \|f''(t)\|_{[t_0, t_1]}. \quad (7)$$

Конечно-разностная аппроксимация производной (двухточечная схема). Дифференцируя (5) по t , получаем, что производная функции $f'(t)$ на отрезке $[t_0, t_1]$ приближается разделенной разностью $L'_1(t) = f'[t_0, t_1]$, которая называется *конечно-разностной аппроксимацией производной* по двум точкам.

Остаточный член аппроксимации $R_1(t) = L'_1(t) - f'(t)$. Повторяя рассуждения, по аналогии с (6) находим

$$|R_1(t)| \leq 0,5h[(1-u)^2 + u^2] \|f''(t)\|_{[t_0, t_1]}.$$

Максимальное значение правой части достигается на концах отрезка $[t_0, t_1]$. Получаем

$$\|R_1(t)\|_{[t_0, t_1]} \leq 0,5h \|f''(t)\|_{[t_0, t_1]}. \quad (8)$$

В дальнейшем оценки погрешности аппроксимации приводятся без выводов. При желании читатель может провести их самостоятельно, используя изложенную методику, или обратиться к [8].

З а м е ч а н и е. Все приведенные в этом параграфе оценки погрешности интерполяции многочленами улучшить нельзя. Это значит, что существуют такие функции $f(t)$ в рассматриваемом классе, что в оценках либо имеет место знак равенства, либо равенство достигается, если константу уменьшить на произвольно малую величину [8]. В частности, оценка (7) точна для многочленов второй степени, когда $f''(t) = \text{const}$.

Приближения многочленами первой степени, близкие к наилучшим равномерным. Наряду с интерполяцией по формуле (5), на отрезке $[t_0, t_1]$ рассмотрим аппроксимацию

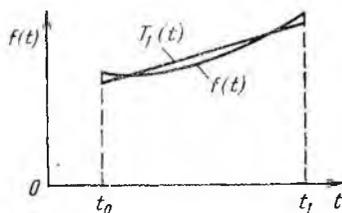
$$P_1(t) = (1-u)[f_0 - h^2 f''_0/16] + u[f_1 - h^2 f''_1/16]. \quad (9)$$

Ее погрешность для функции $f(t) \in C^{(3)}[a, b]$ дается формулой

$$\|R_1(t)\|_{[t_0, t_1]} \leq \frac{h^2}{16} \|f''(t)\|_{[t_0, t_1]} + ch \|f'''(t)\|_{[t_0, t_1]}. \quad (10)$$

Из (10) видно, что коэффициент при $\|f''(t)\|_{[t_0, t_1]}$ вдвое меньше, чем при интерполяции (оценка (7)). Эта аппроксимация для много-

Рис. 2. Приближение многочленом первой степени, аналогичное чебышевскому



членов второй степени ($f''(t) = 0$) является наилучшим равномерным (чебышевским) приближением. Для других функций она дает приближение, качественно похожее на чебышевское (рис. 2). Второй член правой части (10) при $h \rightarrow 0$ есть малая более высокого порядка по сравнению с первым членом. В этом смысле построенное приближение можно назвать *асимптотически наилучшим равномерным приближением*.

Параболическая интерполяция. Наряду с линейной интерполяцией — это часто используемый метод приближения функций, в том числе при работе с таблицами. Пусть заданы значения функции $f_k = f(t_k)$, $k = 0, 1, 2$. Интерполяционный многочлен возьмем в виде (4):

$$L_2(t) = f_0 + (t - t_0) f [t_0, t_1] + (t - t_0)(t - t_1) f [t_0, t_1, t_2].$$

Точность приближения функции $f(t)$ так же, как и ранее, зависит от того, к какому классу она принадлежит, и от расстояния между узлами. Обозначим $h_0 = t_1 - t_0$, $h_1 = t_2 - t_1$, $H = \max(h_0, h_1)$. Остаточный член интерполяции $R_2(t) = L_2(t) - f(t)$. Тогда, если $f(t) \in C^{(2)} [t_0, t_2]$, то

$$\|R_2(t)\|_{[t_0, t_2]} \leq 0,1546H^2 \|f''(t)\|_{[t_0, t_2]}.$$

Сравнивая эту оценку с оценкой линейной интерполяции (7), видим, что при $h_0 = h_1 = h$ они практически совпадают, незначительно различаясь коэффициентами.

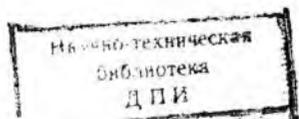
Если интерполируемая функция из более высокого класса гладкости, $f(t) \in C^{(3)} [t_0, t_2]$, то

$$\|R_2(t)\|_{[t_0, t_2]} \leq (\sqrt{3}/27) H^3 \|f'''(t)\|_{[t_0, t_2]}.$$

Как видно из последнего неравенства, здесь порядок приближения относительно H на единицу выше, и при малых H точность повышается. Однако дальнейшее повышение класса гладкости уже не приводит к увеличению порядка приближения относительно H или уменьшению коэффициента. Происходит, как говорят, насыщение интерполяционного процесса. Для многочленов первой степени этому соответствует класс $C^{(2)}$, второй степени — $C^{(3)}$, третьей — $C^{(4)}$.

Конечно-разностная аппроксимация производных (трехточечная схема). Дифференцируя $L_2(t)$, получаем формулу аппроксимации производной

$$L_2'(t) = f [t_0, t_1] + (t - t_0) + (t - t_1) f [t_0, t_1, t_2].$$



В частности, в узлах интерполяции имеем

$$\tilde{f}_0 = (1 + \mu_1) f [t_0, t_1] - \mu_1 f [t_1, t_2]; \quad (11a)$$

$$\tilde{f}_1 = \lambda_1 f [t_0, t_1] + \mu_1 f [t_1, t_2]; \quad (11b)$$

$$\tilde{f}_2 = -\lambda_1 f [t_0, t_1] + (1 + \lambda_1) f [t_1, t_2], \quad (11в)$$

где $\lambda_1 = h_1 (h_0 + h_1)^{-1}$, $\mu_1 = 1 - \lambda_1$.

Это — формулы конечно-разностной аппроксимации производной по трем точкам. Они приближают производную функции $f(t) \in C^{(2)} [t_0, t_2]$ в узлах интерполяции с точностью

$$|R'_2(t_k)| \leq K_{1k} H \|f''(t)\|_{[t_0, t_2]}, \quad (12)$$

где $K_{10} = K_{12} = 2/3$, $K_{11} = 1/2$.

Если $f(t) \in C^{(3)} [t_0, t_2]$, то

$$|R'_2(t_k)| \leq K_{1k} H^2 \|f'''(t)\|_{[t_0, t_2]}, \quad k = 0, 1, 2,$$

где $K_{10} = K_{12} = 1/3$, $K_{11} = 1/6$.

В обоих случаях коэффициент K_{11} , отвечающий $t = t_1$, меньше, чем $K_{10} = K_{12}$. Это значит, что в средних узлах конечно-разностная аппроксимация дает более точное приближение, чем в крайних.

Рассмотрим, наконец, аппроксимацию второй производной функции $f''(t)$. Отметим, что речь может идти только о непрерывной функции. Дифференцируя $L'_2(t)$, находим

$$L'_2(t) = 2f [t_0, t_1, t_2]. \quad (13)$$

Правая часть равна $f''(\xi)$, где $\xi \in [t_0, t_2]$. Поэтому $|L'_2(t)| \leq \|f''(t)\|_{[t_0, t_2]}$, и большего сказать нельзя.

Если же $f(t) \in C^{(3)} [t_0, t_2]$, то имеет место аппроксимация второй производной. В узлах интерполяции оценки будут

$$|R'_2(t_k)| \leq K_{2k} H \|f'''(t)\|_{[t_0, t_2]}, \quad k = 0, 1, 2, \quad (14)$$

где $K_{20} = K_{22} = 1$, $K_{21} = 1/3$.

Кубическая интерполяция. Рассмотрим интерполяционный многочлен Лагранжа

$$L_3(t) = f_0 + (t - t_0) f [t_0, t_1] + (t - t_0)(t - t_1) f [t_0, t_1, t_2] + \\ + (t - t_0)(t - t_1)(t - t_2) f [t_0, t_1, t_2, t_3].$$

Эта формула дает примерно такую же точность приближения, что и формула параболической интерполяции для функций из классов $C^{(2)} [t_0, t_3]$ и $C^{(3)} [t_0, t_3]$. Навысший порядок H^4 достигается, если $f(t) \in C^{(4)} [t_0, t_3]$. Так как многочлен L_3 симметричен относительно индексов 0,1 и 2,3, то оценки погрешности будут одинаковыми для промежутков $[t_0, t_1]$ и $[t_2, t_3]$:

$$\|L_3(t) - f(t)\|_{[t_k, t_{k+1}]} \leq K_{0k} H^4 \|f^{(4)}(t)\|_{[t_0, t_3]}, \quad (15)$$

где $K_{00} = K_{02} = 1/24$; $K_{01} \leq 3/128$; $H = \max(h_0, h_1, h_2)$. Как видно, $K_{01} < K_{00} = K_{02}$. Это означает, что точность приближения на среднем звене отрезка $[t_0, t_3]$ выше, чем на крайних звеньях.

Конечно-разностная аппроксимация производных (четырёхточечная схема). Кубической интерполяции соответствует аппроксимация производных в узлах

$$\begin{aligned} \bar{f}'_0 = & (1 + \mu_1 + \sigma_1) f[t_0, t_1] - \left(\mu_1 + \sigma_1 + \sigma_1 \frac{\mu_2}{\lambda_1} \right) f[t_1, t_2] + \\ & + \sigma_1 \frac{\mu_2}{\lambda_1} f[t_2, t_3]; \end{aligned} \quad (16a)$$

$$\bar{f}'_1 = (\lambda_1 - \lambda_1 \sigma_1) f[t_0, t_1] + (\mu_1 + \lambda_1 \sigma_1 + \mu_2 \sigma_1) f[t_1, t_2] - \mu_2 \sigma_1 f[t_2, t_3]. \quad (16b)$$

Вторые производные аппроксимируются так:

$$\bar{f}''_0 = 2 \{ (1 + 2\sigma_1 + \tau_1) f[t_0, t_1, t_2] - (2\sigma_1 + \tau_1) f[t_1, t_2, t_3] \}, \quad (17a)$$

$$\bar{f}''_1 = 2 \{ (1 - \sigma_1 + \tau_1) f[t_0, t_1, t_2] + (\sigma_1 - \tau_1) f[t_1, t_2, t_3] \}. \quad (17b)$$

Здесь $\sigma_1 = h_0 (h_0 + h_1 + h_2)^{-1}$, $\tau_1 = h_1 (h_0 + h_1 + h_2)^{-1}$.

Формулы (16) и (17) для узлов t_2, t_3 получаются из формул для t_0 и t_1 соответственно заменой $t_0 \rightleftharpoons t_3, t_1 \rightleftharpoons t_2$.

Для функций $f(t) \in C^{(4)} [t_0, t_3]$ оценки погрешности формул (16) будут

$$|f'_k - \bar{f}'_k| \leq K_{1k} H^3 \|f^{IV}(t)\|_{[t_0, t_3]}, \quad k = 0, 1, 2, 3, \quad (18)$$

где $K_{10} = K_{13} = 1/4$, $K_{11} = K_{12} = 1/12$.

Оценки погрешности формул (17) суть

$$|f''_k - \bar{f}''_k| \leq K_{2k} H^2 \|f^{IV}(t)\|_{[t_0, t_3]}, \quad k = 0, 1, 2, 3, \quad (19)$$

где $K_{20} = K_{23} = 11/12$, $K_{21} = K_{22} = 1/6$.

Эрмитова кубическая интерполяция. Так называется интерполяция, когда в узлах t_0, t_1 задаются значения функции f_0, f_1 и ее производной f'_0, f'_1 и строится кубический многочлен $\mathcal{H}_3(t)$ (интерполяционный многочлен Эрмита) по условиям $\mathcal{H}_3(t_k) = f_k, \mathcal{H}'_3(t_k) = f'_k, k = 0, 1$. Он имеет вид (рис. 3)

$$\mathcal{H}_3(t) = \varphi_1(u) f_0 + \varphi_2(u) f_1 + h \varphi_3(u) f'_0 + h \varphi_4(u) f'_1, \quad (20)$$

где $h = t_1 - t_0$;

$$\begin{aligned} \varphi_1(u) &= (1 - u)^2 (1 + 2u); & \varphi_2(u) &= \varphi_1(1 - u) = u^2 (3 - 2u); \\ \varphi_3(u) &= u (1 - u)^2; & \varphi_4(u) &= -\varphi_3(1 - u) = -u^3 (1 - u). \end{aligned} \quad (21)$$

Вывод оценок погрешности эрмитовой интерполяции принципиально не отличается от предыдущих случаев. Дополнение состоит лишь в том, что по формуле Тейлора в узлах приходится представлять не только значения функции f_0 и f_1 , но и значения производной f'_0 и f'_1 , которые входят в выражение интерполяционного

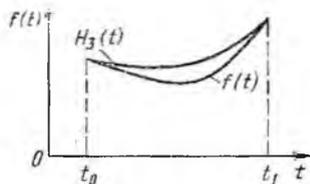


Рис. 3. Эрмитова интерполяция кубическим многочленом

многочлена $\mathcal{H}_3(t)$. При этом, естественно, число членов в представлениях значений производной на единицу меньше.

Для интерполируемой функции $f(t) \in C^{(2)}[t_0, t_1]$ оценки

$$\|R_3^{(p)}(t)\|_{[t_0, t_1]} \leq K_0 h^{2-p} \|f''(t)\|_{[t_0, t_1]}, \quad p = 0, 1, \quad (22)$$

где $K_0 = 1/16$; $K_1 = 0,2515$.

Вторая производная многочлена $\mathcal{H}_3(t)$ не аппроксимирует вторую производную функции $f(t)$. Можно только показать, что

$$\|\mathcal{H}_3''(t)\|_{[t_0, t_1]} \leq (5/3) \|f''(t)\|_{[t_0, t_1]}. \quad (23)$$

Если функция $f(t) \in C^{(4)}[t_0, t_1]$, то аппроксимируются и вторые производные. Оценки имеют вид

$$\|R_3^{(p)}(t)\|_{[t_0, t_1]} \leq K_0 h^{4-p} \|f^{IV}(t)\|_{[t_0, t_1]}, \quad p = 0, 1, 2, \quad (24)$$

где $K_0 = 1/384$; $K_1 = \sqrt{3}/216$; $K_2 = 1/12$.

Привлекают внимание очень малые значения K_0 и K_1 . Сравнивая оценку (24) при $p = 0$ с оценкой интерполяции (15), видим, что в (24) коэффициент K_0 меньше K_{01} в (15) в 9 раз. Оценки для производных также лучше, чем (18) и (19).

3. АППРОКСИМАЦИЯ ФУНКЦИЙ СПЛАЙНАМИ ОДНОЙ И ДВУХ ПЕРЕМЕННЫХ

Самое общее понятие сплайнов уже было дано во введении. Здесь мы уточним его, ограничиваясь конструкциями, которые используются в последующем для решения задач инженерной геометрии.

Определение сплайнов одной переменной. На отрезке задается сетка узлов

$$\Delta: a = t_1 < \dots < t_i < \dots < t_N = b.$$

Функция $S_{n,\nu}(t)$ называется *сплайном степени n класса C^m* ($0 \leq m \leq n$; $\nu = n - m$) с *узлами* на сетке Δ , если

а) на каждом отрезке $[t_i, t_{i+1}]$, $i = 1, \dots, N - 1$, функция $S_{n,\nu}(t)$ является многочленом степени n , т. е.

$$S_{n,\nu}(t) = P_n^i(t) = \sum_{k=0}^n a_k^i (t - t_i)^k \quad \text{для } t \in [t_i, t_{i+1}];$$

б) $S_{n,\nu}(t) \in C^m[a, b]$ (рис. 4).

Производные сплайна $S_{n,\nu}^{(p)}(t)$ порядка выше m и до n включительно, вообще говоря, имеют конечные разрывы в точках t_i , $i = 2, \dots, N - 1$. Индекс $\nu = n - m$ указывает на число таких производных. Число ν часто называют *дефектом сплайна*.

Пространство сплайнов. Множество сплайнов, удовлетворяющих определению, обозначим через $S_{n,\nu}(\Delta)$. Множество $S_{n,\nu}(\Delta)$ есть линейное пространство. С этим связаны термины *линейные сплайны* или *линейная теория сплайнов*.

Каждое звено сплайна зависит от $n + 1$ коэффициентов определяющего его многочлена. Общее число коэффициентов $(n + 1)(N - 1)$. Но в точках $t_i, i = 2, \dots, N - 1$ должны выполняться условия непрерывности первых $n - v$ производных сплайна, которые дают $(n - v + 1) \times (N - 2)$ линейных уравнений для коэффициентов a_p^i . Таким образом,

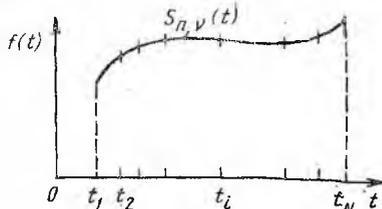


Рис. 4. Сплайн одной переменной

число независимых коэффициентов будет $d_{n,v} = n + 1 + v(N - 2)$. Это есть размерность пространства сплайнов $S_{n,v}(\Delta)$. Ее можно представить как размерность $d_n = n + 1$ множества многочленов на $[t_1, t_2]$ плюс число разрывов производных во «внутренних» узлах сетки Δ .

В пространстве $S_{n,v}(\Delta)$ можно построить *базисы*. Например, аналогом базиса из степеней t^k в пространстве многочленов M_n здесь является система функций

$$(t - t_1)^k, k = 0, \dots, n; (t - t_i)_+^\alpha, \alpha = n - v + 1, \dots, \\ (1 \leq v \leq n), i = 2, \dots, N - 1,$$

где $t_+^\alpha = [\max(t, 0)]^\alpha$ — *усеченные степенные функции*. При этом

$$S_{n,v}(t) = \sum_{k=0}^n a_k^1 (t - t_1)^k + \sum_{i=2}^{N-1} \sum_{\alpha=n-v+1}^n c_\alpha^i (t - t_i)_+^\alpha. \quad (25)$$

Сплайны двух и более переменных. Такие сплайны характеризуются двумя признаками: формой области с ее разбиением на подобласти и определением множества сплайнов. Мы ограничимся случаем двух переменных, когда заданная область — прямоугольник $\Omega = [a, b] \times [c, d]$. В области Ω вводится сетка линий $\Delta = \Delta_t \times \Delta_s$, причем $\Delta_t: a = t_1 < \dots < t_i < \dots < t_N = b$, $\Delta_s: c = s_1 < \dots < s_j < \dots < s_M = d$, делящая Ω на прямоугольники $\Omega_{ij} = [t_i, t_{i+1}] \times [s_j, s_{j+1}]$.

Отметим сначала функции $\bar{S}_{n,v}(t; s)$, которые являются сплайнами степени n класса C^m только по переменной t на сетке Δ_t , а s входит как параметр:

$$\bar{S}_{n,v}(t, s) = \sum_{k=0}^n \bar{a}_k^i(s) (t - t_i)^k, (t, s) \in [t_i, t_{i+1}] \times [c, d], \quad (26a)$$

где $\bar{a}_k^i(s)$ — функции переменной s . Аналогично, если $\tilde{S}_{n,v}(t, s)$ есть сплайн по переменной s на сетке Δ_s , то

$$\tilde{S}_{n,v}(t, s) = \sum_{l=0}^n a_l^j(t) (s - s_j)^l, (t, s) \in [a, b] \times [s_j, s_{j+1}]. \quad (26b)$$

Такие функции называются *частичными сплайнами*. Их обычно используют на промежуточном этапе при аппроксимации сплайнами двух переменных.

Функция $S_{n,v}(t, s)$ называется *сплайном двух переменных степени n класса $C^{m,m}$ с линиями склейки на сетке Δ* , если:

а) в каждой клетке Ω_{ij} функция $S_{n,v}(t, s)$ является многочленом степени n по каждой из переменных t, s , т. е.

$$S_{n,v}(t, s) = \sum_{k=0}^n \sum_{l=0}^n a_{kl}^{ij} (t - t_i)^k (s - s_j)^l, \quad (t, s) \in \Omega_{ij};$$

б) $S_{n,v}(t, s) \in C^{m,m}[\Omega]$.

Множество этих сплайнов образует линейное пространство $S_{n,v}(\Delta)$. Всякий сплайн двух переменных является сплайном по каждой из них и может быть записан формулой (26а). При фиксированном s такие функции образуют пространство сплайнов $S_{n,v}(\Delta_s)$ размерности $d_{n,v} = n + 1 + v(N - 2)$. Функции $\bar{a}_k^i(s)$ являются сплайнами по переменной s и представляются формулами типа (26б) с постоянными коэффициентами a_i^j . Число независимых коэффициентов a_i^j равно $\bar{d}_{n,v} = n + 1 + v(M - 2)$. А так как число независимых функций $\bar{a}_k^i(s)$ есть $d_{n,v}$, то общее число независимых коэффициентов a_{kl}^{ij} равно $d_{n,v}\bar{d}_{n,v}$. Это произведение и есть размерность пространства сплайнов $S_{n,v}(\Delta)$.

Задачи аппроксимации сплайнами. Это, с одной стороны, те же задачи, что и для многочленов, а с другой — специфические проблемы, рассматриваемые только в теории сплайнов.

К первым относится задача интерполяции. Для сплайнов нечетных степеней, например первой и третьей, характерна схема, при которой узлы интерполяции и узлы сплайна совпадают (см. п. 4—6). В случае четных степеней, например параболических сплайнов [21], узлы не совпадают, а чередуются между собой. Для сплайнов рассматриваются задачи о наилучших приближениях, а также метод наименьших квадратов и асимптотически наилучшие приближения.

Новой задачей является задача сглаживания экспериментальных данных как задача минимизации некоторого специального функционала при ограничении на уклонение сплайна от заданных значений (п. 7). В п. 9 рассмотрены методы локальной аппроксимации, в которых каждое звено сплайна полностью определяется исходными данными на его отрезке — носителе или на нескольких соседних с ним отрезках. В каком пространстве сплайнов $S_{n,v}(\Delta)$ искать решение той или иной задачи, зависит от требований, предъявляемых к гладкости сплайна.

Погрешность аппроксимации и сходимость. Погрешность аппроксимации сплайнами будем характеризовать так же, как и в случае многочленов, нормой остаточного члена $\|R_{n,v}(t)\| = \|S_{n,v}(t) - f(t)\|$. Она снова зависит от того, к какому классу: $C^m[a, b]$, или $C^{(m)}[a, b]$, принадлежит функция $f(t)$, и от шагов h_i сетки Δ .

Изменяя координаты узлов сетки Δ и их число, получим последовательность сеток и соответствующую им последовательность аппроксимационных сплайнов. Если $\|R_{n,v}(t)\| \rightarrow 0$ при $N \rightarrow \infty$ и $H = \max h_i \rightarrow 0$, то говорят, что аппроксимационный процесс

сходится. Если $\|R_{n,\nu}(t)\| = O(H^m)$, $m > 0$, то говорят, что имеет место *сходимость с порядком m* . В дальнейшем мы будем рассматривать только сходящиеся процессы.

Уменьшение шага ведет к увеличению числа узлов. В случае многочленов это связано с возрастанием их степени, что является источником различных трудностей. Степень же сплайна не изменяется, и это принципиальный момент их теории.

Погрешность аппроксимации в случае двух переменных характеризуется нормой остаточного члена $\|R_n(t, s)\| = \|S_{n,\nu}(t, s) - f(t, s)\|$ в области Ω . Она зависит, во-первых, от того, к какому классу: $C^{m,m}[\Omega]$ или $C^{(m,m)}[\Omega]$, принадлежит функция $f(t, s)$, и, во-вторых, от шагов по обоим переменным t и s . Если обозначить $l_j = s_{j+1} - s_j$, $L = \max l_j$, то типичными будут формулы вида $\|S_{n,\nu}(t, s) - f(t, s)\| = O(H^m + L^m)$, обеспечивающие сходимость аппроксимационного процесса.

Вычислительный аспект. При переходе от аппроксимации многочленами к аппроксимации сплайнами преследуют две цели. Первая, это улучшение качества приближения: при одинаковых вычислительных затратах абсолютные погрешности аппроксимации сплайнами меньше, чем погрешности аппроксимации многочленами, а при одинаковых погрешностях уменьшается объем вычислений. Сплайны позволяют избежать осцилляций. Для сходимости аппроксимации к аппроксимируемой функции предъявляются более слабые требования, чем в случае многочленов. Например, интерполяция сплайнами невысоких степеней сходится даже для непрерывных функций.

Вторая цель — резкое уменьшение вычислительных трудностей как при построении алгоритмов решения задач, так и при дальнейшей работе с аппроксимантами, которые на каждом звене представляют собой многочлены невысоких степеней или иные элементарные функции. С этой точки зрения наиболее простыми оказываются задачи локальной аппроксимации. В таких случаях решения выписываются обычно в явном виде. Нелокальные задачи сводятся к решению систем линейных уравнений размерности порядка N с разреженными (*ленточными*) матрицами (см. п. 6—8), для которых существуют эффективные алгоритмы. В этом заключается одно из преимуществ сплайнов перед многочленами высоких степеней, где матрицы заполненные. В случае аппроксимации сплайнами двух переменных проблема часто сводится к решению серий одномерных задач, что также представляет весьма существенное преимущество.

При работе со сплайнами можно использовать либо кусочно-многочленное представление, либо представление через базисные функции. В первом случае достигается наибольшая экономия в числе арифметических операций, но зато приходится хранить большой объем информации о многочленах — $(n+1)(N-1)$ чисел. Во втором случае достаточно хранить $d_{n,\nu} = n+1 + \nu(N-2)$ коэффициентов сплайна, но при этом увеличивается число арифметических операций. Следует предупредить, что не все базисы (а их в теоретических исследованиях используется несколько) пригодны для

организации вычислений. Большой частью их применение приводит к быстрому накоплению погрешностей округления. В частности, при использовании усеченных степенных функций затруднения возникают при вычислении второй суммы в (25) при больших N . От этого недостатка свободен лишь базис, состоящий из функций с конечными носителями (отличных от нуля на конечном отрезке), так называемых B -сплайнов, рассматриваемых ниже в п. 8.

4. СПЛАЙНЫ ПЕРВОЙ СТЕПЕНИ

Рассмотрение конкретных сплайнов начнем с хорошо известного примера, за внешней простотой которого скрываются многие специфические особенности сплайнов.

Интерполяция. Сплайны первой степени $S_1(t)$ класса C на сетке Δ — это кусочно-линейные функции. (Здесь и далее второй индекс v сплайна, если он равен единице, опускается.)

Пусть в узлах сетки Δ заданы значения функции $f_i = f(t_i)$. Интерполяционный сплайн $S_1(t)$ определяется условиями $S_1(t_i) = f_i$, $i = 1, \dots, N$. Геометрически он представляет собой ломаную, проходящую через точки (t_i, f_i) (рис. 5). Между соседними узлами реализуется линейная интерполяция. Согласно (5) при $t \in [t_i, t_{i+1}]$ сплайн будет иметь вид

$$S_1(t) = (1 - u)f_i + uf_{i+1}; \quad u = (t - t_i)/h_i. \quad (27)$$

Если в исходные данные f_i внести возмущения $\delta f_i \leq \delta$, то сплайн изменится в пределах δ . Возмущение его производной оценивается значением δ/h_i и растет с уменьшением h_i .

Погрешность интерполяции. Пусть $f(t) \in C^{(2)}[a, b]$. Тогда согласно (7) и (8)

$$\|S_1^{(p)}(t) - f^{(p)}(t)\|_{[t_i, t_{i+1}]} \leq K_p h_i^{2-p} \|f''(t)\|_{[t_i, t_{i+1}]}, \quad p = 0, 1, \quad (28)$$

где $K_0 = 1/8$, $K_1 = 1/2$. Чтобы записать оценки для всего отрезка $[a, b]$, достаточно заменить h_i на H , а $\|\cdot\|_{[t_i, t_{i+1}]}$ на $\|\cdot\|$:

$$\|S_1^{(p)}(t) - f^{(p)}(t)\| \leq K_p H^{2-p} \|f''(t)\|, \quad p = 0, 1. \quad (29)$$

Интерполяция с заданной точностью. На практике обычно требуется осуществить интерполяцию с некоторой заданной точностью. Эта задача решается путем выбора сетки на отрезке $[a, b]$ с учетом свойств интерполируемой функции. Согласно (28) точность интерполяции определяется гладкостью функции внутри отрезков $[t_i, t_{i+1}]$. Поэтому точки, в которых функция имеет особенности, например разрывы производных, следует включить в число узлов интерполяции.

Оценка (29) при $p = 0$ дает возможность определить максимальный шаг сетки для обеспечения заданной точности $\|S_1(t) - f(t)\| \leq \varepsilon$. Для этого

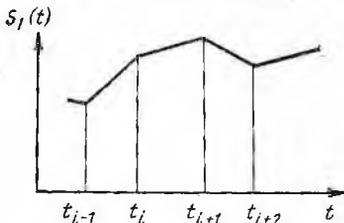


Рис. 5. Сплайн первой степени

достаточно выбрать шаг h так, чтобы $h^2 \|f''(t)\| \leq \delta \varepsilon$, т. е. $h = \sqrt{\delta \varepsilon / \|f''(t)\|}$, и требуемая точность обеспечивается, например, на равномерной сетке с шагом h . При этом число узлов, как правило, будет очень завышенным.

С помощью оценки (28) при $p = 0$ можно строить более экономичные по числу шагов сетки, если учитывать значения $\|f''(t)\|_{[t_i, t_{i+1}]}$ на каждом отрезке. Так, если узлы интерполяции выбирать из условий

$$h_i^2 \|f''(t)\|_{[t_i, t_{i+1}]} = \delta \varepsilon, \quad (30)$$

то задача будет решена с числом узлов, вообще говоря, меньшим, чем при равномерном разбиении.

Пусть, например, функция $|f''(t)|$ монотонно убывает на $[a, b]$. Если узел t_i найден, то в силу монотонности $\|f''(t)\|_{[t_i, t_{i+1}]} = |f''(t_i)|$, и можно вычислить h_i . Если $|f''(t_i)| \neq 0$, то полагаем $t_{i+1} = t_i + h_i$. Если $|f''(t_i)| = 0$, то $t_{i+1} = b$. Начиная этот процесс от левого конца отрезка $[a, b]$, последовательно находим все узлы сетки. Если $|f''(t)|$ монотонно возрастает, то построение следует начинать с правого конца отрезка $[a, b]$. Алгоритм можно применять, если функция $|f''(t)|$ кусочно-монотонна. Алгоритм выбора узлов сплайна, при котором не требуется выделять участки монотонности функции $|f''(t)|$, будет изложен в п. 5.

Аппроксимация сплайнами первой степени. Аналогом асимптотически наилучшего равномерного приближения многочленом первой степени (9) является следующая аппроксимация сплайном:

$$S_1(t) = (1-u) \left[f_i - \frac{1}{16} \max(h_{i-1}^2, h_i^2) f_i'' \right] + u \left[f_{i+1} - \frac{1}{16} \max(h_i^2, h_{i+1}^2) f_{i+1}'' \right], \quad t \in [t_i, t_{i+1}], \quad h_{-1} = h_N = 0. \quad (31)$$

Погрешность аппроксимации согласно (10)

$$\|S_1(t) - f(t)\|_{[t_i, t_{i+1}]} \leq \frac{1}{16} H_i^2 [\|f''(t)\|_{[t_i, t_{i+1}]} + H_i \|f'''(t)\|_{[t_i, t_{i+1}]}],$$

где $H_i = \max(h_{i-1}, h_i, h_{i+1})$.

Производные в (31) можно заменить конечно-разностными аппроксимациями (11). При этом в силу (14) в (31) будет внесена погрешность порядка $O(H^3)$. Отсюда ясно, что главный (первый) член погрешности аппроксимации не изменится.

5. ЭРМИТОВЫ КУБИЧЕСКИЕ СПЛАЙНЫ

Интерполяция. В узлах сетки Δ заданы значения функции и ее производной $f_i = f(t_i)$, $f_i' = f'(t_i)$, $i = 1, \dots, N$. Требуется найти кубический сплайн $S_{3,2}(t)$ такой, что

$$S_{3,2}(t_i) = f_i; \quad S_{3,2}'(t_i) = f_i', \quad i = 1, \dots, N.$$

Это — эрмитов кубический сплайн. Его первая производная непрерывна, а вторая и третья в общем случае разрывны в узлах сетки. График представляет собой гладкую линию (см. рис. 4).

На отрезке $[t_i, t_{i+1}]$ реализуется эрмитова интерполяция кубическим многочленом (20). Звено сплайна имеет вид

$$S_{3,2}(t) = f_i \varphi_1(u) + f_{i+1} \varphi_2(u) + h_i [f_i' \varphi_3(u) + f_{i+1}' \varphi_4(u)], \quad (32)$$

где $\varphi_k(u)$ — функции (21). Это локальная интерполяция.

Возмущения исходных данных $\delta f_i \leq \delta$, $\delta f_i' \leq \delta'$ вызывают у сплайна и его производных возмущения порядка $O(\delta h_i^p + \delta' h_i^{p+1})$.

Погрешность интерполяции. Для функций $f(t) \in C^{(2)}[a, b]$ имеем оценки (22), которые для сплайнов дают

$$|S_{3,2}^{(p)}(t) - f^{(p)}(t)| \leq K_p h_i^{2-p} \|f''(t)\|_{[t_i, t_{i+1}]}, \quad p = 0, 1, \quad (33)$$

где $K_0 = 1/16$; $K_1 = 0,2515$.

Для функции $f(t) \in C^{(4)}[a, b]$ из (24) получаем

$$|S_{3,2}^{(p)}(t) - f^{(p)}(t)| \leq K_p h_i^{4-p} \|f^{IV}(t)\|_{[t_i, t_{i+1}]}, \quad p = 0, 1, 2, \quad (34)$$

где $K_0 = 1/384$; $K_1 = \sqrt{3}/216$; $K_2 = 1/12$.

Эти оценки остаются справедливыми и для функций с кусочно-непрерывными вторыми и третьими производными, если точки их разрыва включаются в число узлов интерполяции. В инженерной практике к такому классу относятся часто используемые гладкие кусочно-аналитические функции.

Из формул (33), (34) получаются оценки, справедливые на всем отрезке $[a, b]$, заменой h_i на H и $\|\cdot\|_{[t_i, t_{i+1}]}$ на $\|\cdot\|$.

Интерполяция с заданной точностью. Задача интерполяции с точностью

$$|R(t)| = |S_{3,2}(t) - f(t)| \leq \varepsilon, \quad t \in [a, b] \quad (35)$$

решается двояко. Шаги равномерной сетки для функций $f(t) \in C^{(2)}[a, b]$ позволяет определить оценка (33):

$$h \leq 4 \sqrt{\varepsilon / \|f''(t)\|}, \quad (36)$$

а шаги для функций $f(t) \in C^{(4)}[a, b]$ — оценка (34):

$$h \leq \sqrt[4]{384\varepsilon / \|f^{IV}(t)\|}. \quad (37)$$

Для построения неравномерных сеток при интерполяции функций $f(t) \in C^{(2)}[a, b]$ можно использовать алгоритм, приведенный в п. 4. Для функций $f(t) \in C^{(4)}[a, b]$ в этом алгоритме нужно всюду заменить $f''(t)$ на $f^{IV}(t)$.

Ниже рассмотрен другой алгоритм построения неравномерной сетки, не требующий вычисления производных $f''(t)$ или $f^{IV}(t)$. Он достаточно трудоемок, но зато позволяет получать сетки практически с минимальным числом узлов.

На отрезке $[a, b]$ возьмем множество точек (контрольных) $\tau_j = a + j\delta t$, $j = 0, \dots$, и предположим, что остаточный член интерполяции в них удовлетворяет условию

$$|R(\tau_j)| = |S_{3,2}(\tau_j) - f(\tau_j)| \leq \varepsilon_1 < \varepsilon.$$

Выясним, когда на основании этого можно утверждать, что заданная точность обеспечивается на всем отрезке $[a, b]$.

Построим сплайн первой степени $S_1(t)$, интерполирующий функцию $R(t)$ по значениям $R(\tau_j)$. Поскольку $|R(\tau_j)| \leq \varepsilon_1$, то $|S_1(t)| \leq \varepsilon_1$ на всем отрезке $[a, b]$. Отсюда ясно, что достаточно добиться выполнения неравенства

$$|\delta R(t)| = |S_1(t) - R(t)| \leq \varepsilon - \varepsilon_1, \quad (38)$$

и тогда условие (35) будет выполнено.

Согласно (28)

$$|\delta R(t)| \leq 0,125(\delta t)^2 \|R''(t)\|. \quad (39)$$

Если функция $f(t) \in C^2[a, b]$, то учитываем, что $\|R''(t)\| \leq \|S_{3,2}''(t)\| + \|f''(t)\|$, и на основании (23) получаем

$$|\delta R(t)| \leq 3^{-1}(\delta t)^2 \|f''(t)\|.$$

На шаги δt налагаем ограничение

$$\delta t \leq \sqrt{3(\varepsilon - \varepsilon_1) / \|f''(t)\|}. \quad (40)$$

При этом ограничении неравенства (38) и (35) будут выполнены.

Рассмотрим узловые моменты алгоритма. Ввиду локальности эрмитова сплайна задача состоит в последовательном определении по возможности максимальных размеров каждого шага сетки Δ . Предполагается, что известна оценка $\|f''(t)\| \leq A$. Зададим значение ε_1 и по формулам (40) и (36) определим максимально допустимый шаг δt и шаг h равномерной сетки, обеспечивающей точность ε интерполяции: $\delta t = \sqrt{3(\varepsilon - \varepsilon_1)/A}$; $h = 4\sqrt{\varepsilon/A}$.

Пусть найден узел t_{i-1} . Узел t_i находим путем последовательных приближений $t_i^{(k)}$ $k = 0, 1, \dots$. В качестве нулевого приближения принимаем $t_i^{(0)} = t_{i-1} + h_{i-2}$. Если оказалось, что $t_i^{(0)} > b$, полагаем $t_i^{(0)} = b$. Каждое последующее приближение $t_i^{(k)}$ выбирается из некоторого допустимого отрезка $[\alpha^{(k)}, \beta^{(k)}]$. Для $t_i^{(1)}$ устанавливаем $\alpha^{(1)} = t_{i-1} + h$, $\beta^{(1)} = b$.

Приближение $t_i^{(k)}$ определяется следующим образом. По значениям $f(t_{i-1})$, $f'(t_{i-1})$, $f(t_i^{(k-1)})$, $f'(t_i^{(k-1)})$ строится звено кубического сплайна. Подсчитываются его уклонения e'_j от $f(t)$ в точках $\tau_j = t_{i-1} + j\delta t$, $j = 1, \dots, m_h$ и находится $e^* = \max e'_j$. Возможны три случая:

1) $e^* < \varepsilon_1$. Экстраполируем звено сплайна вправо и вычисляем его уклонения от $f(t)$ в точках $\tau_l = t_i^{(k-1)} + l\delta t$, $l = 1, 2, \dots, l_h$, до тех пор, пока $|R(\tau_l)| \leq \varepsilon_1 - e^*$, а $\tau_l \in [\alpha^{(k)}, \beta^{(k)}]$. В качестве k -го приближения принимаем точку $t_i^{(k)} = t_i^{(k-1)} + l_h\delta t$. Устанавливаем границы допустимого отрезка для последующего приближения равными $\alpha^{(k+1)} = t_i^{(k-1)}$, $\beta^{(k+1)} = \beta^{(k)}$;

2) $e^* = \varepsilon_1$. Принимаем $t_i^{(k)} = t_i^{(k-1)} + \delta t$. Если оказалось, что $t_i^{(k)} \geq \beta^{(k)}$, полагаем $t_i^{(k)} = \beta^{(k)}$. Корректируем границы допустимого отрезка $\alpha^{(k+1)} = t_i^{(k-1)}$, $\beta^{(k+1)} = \beta^{(k)}$;

3) $\varepsilon^* > \varepsilon_1$. Принимаем $t_i^{(k)} = \alpha^{(k)} + (t_i^{(k-1)} - \alpha^{(k)}) \sqrt{\varepsilon_1/\varepsilon^*}$, если $\sqrt{\varepsilon_1/\varepsilon^*} \leq 1/2$. В противном случае $t_i^{(k)} = \alpha^{(k)} + (t_i^{(k-1)} - \alpha^{(k)})/2$. Изменяем границы допустимого промежутка $\alpha^{(k+1)} = \alpha^{(k)}$, $\beta^{(k+1)} = t_i^{(k-1)}$.

Нетрудно убедиться, что процесс последовательного приближения сходится, так как допустимый отрезок $[\alpha^{(k)}, \beta^{(k)}]$ на каждом шаге уменьшается, и как только он становится меньше δt , можно считать, что узел t_i определен.

Основные вычислительные затраты этого алгоритма приходятся на определение значений $f(\tau_j)$, $S_{3,2}(\tau_j)$ и $|S_{3,2}(\tau_j) - f(\tau_j)|$. Их объем зависит от поведения второй производной $f''(t)$, от того, насколько точно взято значение A и как выбрано ε_1 . Из (40) очевидно, что большему шагу δt соответствует меньшее значение ε_1 , т. е. уменьшение числа контрольных точек возможно только за счет уменьшения допустимого отклонения $S_{3,2}(t)$ от $f(t)$ в этих точках. На практике рекомендуется принимать значения $\varepsilon_1 = 0,9\varepsilon$. В этом случае $\delta t = \sqrt{0,3\varepsilon/A}$ и общее число контрольных точек $m = (b - a)/\delta t$. Отметим, что конечный результат практически не зависит от того, насколько точно устанавливается оценка $\|f''(t)\| \leq A$.

В случае интерполяции функции $f(t) \in C^{(4)}[a, b]$ вместо формулы (36) используют формулу (37). Из (39), (37) и оценки (34) при $p = 0$ следует

$$\delta R(t) \leq 96^{-1} (\delta t)^2 h^2 \|f^{IV}(t)\| \leq (\delta t)^2 \sqrt{\varepsilon \|f^{IV}(t)\|/24}.$$

Чтобы выполнялось условие (38), следует положить

$$\delta t \leq \sqrt[4]{24(\varepsilon - \varepsilon_1)^2/\varepsilon \|f^{IV}(t)\|}.$$

Для применения алгоритма нужно знать оценку $\|f^{IV}(t)\| \leq A$. Эрмитовы кубические сплайны двух переменных. Интерполяция. В области $\Omega = [a, b] \times [c, d]$ в узлах сетки $\Delta = \Delta_t \times \Delta_s$ заданы значения некоторой функции $f(t, s)$ и ее производных $D^{p,q}f(t, s)$, $p, q = 0, 1$, т. е. заданы $f_{ij}^{(p,q)} = D^{p,q}f(t_i, s_j)$, $i = 1, \dots, N$; $j = 1, \dots, M$. Задача эрмитовой интерполяции кубическим сплайном двух переменных $S_{3,2}(t, s)$ состоит в построении сплайна, удовлетворяющего условиям

$$D^{p,q}S_{3,2}(t_i, s_j) = f_{ij}^{(p,q)}, \quad p, q = 0, 1, \quad i = 1, \dots, N; \quad j = 1, \dots, M.$$

Нахождение сплайна $S_{3,2}(t, s)$ состоит в последовательном построении двух интерполяционных частичных сплайнов. Вначале по формуле (32) получаем

$$\begin{aligned} \bar{S}_{3,2}(t; s) = & \varphi_1(u) f(t_i, s) + \varphi_2(u) f(t_{i+1}, s) + h_i \varphi_3(u) D^{1,0}f(t_i, s) + \\ & + h_i \varphi_4(u) D^{1,0}f(t_{i+1}, s), \quad t \in [t_i, t_{i+1}]. \end{aligned}$$

Находим производную $D^{0,1}\bar{S}_{3,2}(t; s)$ по переменной s при $s = s_j, s_{j+1}$. Операции интерполирования по t и дифференцирования по s перестановочны. Интерполируя функцию $\bar{S}_{3,2}(t; s)$ частичным

сплайном по s , получаем эрмитов кубический сплайн двух переменных.

Определим вектор-строку $\varphi^T(u)$, вектор-столбец $\varphi(v)$, $v = (s - s_j)/l_j$, и матрицу F формулами

$$\varphi^T(u) = [\varphi_1(u), \varphi_2(u), h_i\varphi_3(u), h_i\varphi_4(u)];$$

$$F = \begin{bmatrix} f_{ij} & f_{i,j+1} & f_{ij}^{(0,1)} & f_{i,j+1}^{(0,1)} \\ f_{i+1,j} & f_{i+1,j+1} & f_{i+1,j}^{(0,1)} & f_{i+1,j+1}^{(0,1)} \\ f_{ij}^{(1,0)} & f_{i,j+1}^{(1,0)} & f_{ij}^{(1,1)} & f_{i,j+1}^{(1,1)} \\ f_{i+1,j}^{(1,0)} & f_{i+1,j+1}^{(1,0)} & f_{i+1,j}^{(1,1)} & f_{i+1,j+1}^{(1,1)} \end{bmatrix}; \quad \varphi(v) = \begin{bmatrix} \varphi_1(v) \\ \varphi_2(v) \\ L_j\varphi_3(v) \\ L_j\varphi_4(v) \end{bmatrix}.$$

Тогда сплайн может быть записан в виде

$$S_{3,2}(t, s) = \varphi^T(u) F \varphi(v). \quad (41)$$

Погрешность интерполяции. Ее получаем как сумму погрешностей, образующихся на каждом из двух шагов решения задачи.

Если $f(t, s) \in C^{(2,2)}[\Omega]$, то

$$\|D^{p,q}[S_{3,2}(t, s) - f(t, s)]\| \leq K_p H^{2-p} \|D^{2,0} f(t, s)\| +$$

$$+ K_q L^{2-q} \|D^{0,2} f(t, s)\| + K_p K_q H^{2-p} L^{2-q} \|D^{2,2} f(t, s)\|, \quad p, q = 0, 1, \quad (42a)$$

где $K_0 = 1/16$; $K_1 = 0,2515$.

Если $f(t, s) \in C^{(4,4)}[\Omega]$, то

$$\|D^{p,q}[S_{3,2}(t, s) - f(t, s)]\| \leq K_p H^{4-p} \|D^{4,0} f(t, s)\| +$$

$$+ K_q L^{4-q} \|D^{0,4} f(t, s)\| + K_p K_q H^{4-p} L^{4-q} \|D^{4,4} f(t, s)\|, \quad p, q = 0, 1, 2, \quad (42b)$$

где $K_0 = 1/384$; $K_1 = \sqrt{3}/216$; $K_2 = 1/12$.

Интерполяция с заданной точностью. При достаточно малых H и L величинами $O(H^2 L^2)$ в (42a) и $O(H^4 L^4)$ в (42b) можно пренебречь. Тогда равномерная сетка в случае $f(t, s) \in C^{(2,2)}[\Omega]$ строится исходя из требования, чтобы каждое из оставшихся слагаемых в правой части (42a) в отдельности не превышало заданного значения, т. е.

$$\frac{1}{16} h^2 \|D^{2,0} f(t, s)\| \leq \varepsilon_t; \quad \frac{1}{16} l^2 \|D^{0,2} f(t, s)\| \leq \varepsilon_s. \quad (43a)$$

Здесь $\varepsilon_t + \varepsilon_s = \varepsilon$ — заданная точность приближения. Таким образом, задача об определении двумерной сетки $\Delta = \Delta_t \times \Delta_s$ сводится к независимому определению одномерных сеток Δ_t и Δ_s . Если $\|D^{2,0} f(t, s)\| \leq A$; $\|D^{0,2} f(t, s)\| \leq B$, то величины ε_t и ε_s можно задать формулами $\varepsilon_t = \varepsilon A (A + B)^{-1}$, $\varepsilon_s = \varepsilon B (A + B)^{-1}$.

В случае $f(t, s) \in C^{(4,4)}[a, b]$ следует взять соотношения

$$\frac{1}{384} h^4 \|D^{4,0} f(t, s)\| \leq \varepsilon_t; \quad \frac{1}{384} l^4 \|D^{0,4} f(t, s)\| \leq \varepsilon_s. \quad (43b)$$

Построение экономичных неравномерных сеток осложняется тем, что оптимальные шаги, например, по переменной t будут разными при разных s . Но сетка должна быть прямоугольной.

Выделим в области Ω точки (\bar{t}_i, \bar{s}_j) , в которых функция $f(t, s)$ имеет характерные особенности, и введем линии $t = \bar{t}_i$, $s = \bar{s}_j$. Построим сетку Δ_t , выполняя условия типа (35) лишь на линиях $s = \bar{s}_1, \dots, \bar{s}_m$. Первый узел сетки Δ_t $t_1 = a$. Затем для каждого \bar{s}_j одним из изложенных методов определяем узлы $t_2(\bar{s}_j)$. В качестве второго узла сетки Δ_t принимаем $t_2 = \min \{t_2(\bar{s}_j)\}$. Таким же обра-

зом находим t_3, t_4 и т. д. Аналогичным образом строится сетка Δ_s .

При удачном выборе значений $\{\bar{t}_i\}$, $\{\bar{s}_j\}$, получив сетку $\Delta = \Delta_t \times \Delta_s$, можно считать задачу решенной, т. е. требование $|S_{3,2}(t, s) - f(t, s)| \leq \varepsilon$ будет выполняться во всей области Ω . Если полной уверенности в этом нет, то следует проверить неравенство в некоторых дополнительных точках, например в центрах клеток Ω_{ij} .

6. КУБИЧЕСКИЕ СПЛАЙНЫ

Этим термином обозначаются *кубические сплайны класса C^2* с двумя непрерывными производными в отличие от эрмитовых кубических сплайнов класса C^1 .

Задача интерполяции. Пусть на отрезке $[a, b]$ в узлах сетки Δ заданы значения некоторой функции $f_i = f(t_i)$, $i = 1, \dots, N$. Найти кубический сплайн $S_3(f; t)$ такой, что

$$S_3(f; t_i) = f_i, \quad i = 1, \dots, N. \quad (44)$$

Сплайн $S_3(f; t)$ на каждом из отрезков $[t_i, t_{i+1}]$ определяется четырьмя коэффициентами, а на всем отрезке $[a, b]$ — $4(N-1)$ коэффициентами. Поскольку он непрерывен и непрерывны его первая и вторая производные во внутренних узлах t_i , $i = 2, \dots, N-1$, сетки Δ , это дает $3(N-2)$ равенств для коэффициентов. Вместе с равенствами (44) получаются $4N-6$ соотношений. Два дополнительных соотношения задаются в виде *краевых условий*, т. е. ограничений на значения сплайна и его производных на концах отрезка $[a, b]$ (или вблизи концов). Из них наиболее употребительными являются следующие типы:

$$\text{I. } S_3'(f; a) = f'(a); \quad S_3'(f; b) = f'(b);$$

$$\text{II. } S_3''(f; a) = f''(a); \quad S_3''(f; b) = f''(b);$$

$$\text{III. } S_3^{(p)}(f; a) = S_3^{(p)}(f; b), \quad p = 1, 2;$$

$$\text{IV. } S_3'''(f; t_r - 0) = S_3'''(f; t_r + 0), \quad r = 2, N-1.$$

Условия типа III носят название периодических. Их применяют, если интерполируемая функция $f(t)$ — периодическая с периодом $b - a$. Условия типа IV означают непрерывность третьих производных в точках t_2 и t_{N-1} , т. е. сплайн выражается одним кубическим многочленом на отрезке $[t_1, t_3]$ и $[t_{N-2}, t_N]$, а не двумя. Возможны также случаи, когда на одном конце отрезка задано условие одного из типов I, II, IV, а на противоположном конце — другое,

Кубический сплайн $S_3(f; t)$ на отрезке $[t_i, t_{i+1}]$ представляется формулой, аналогичной формуле (32) для эрмитова кубического сплайна,

$$S_3(t) = f_i \varphi_1(t) + f_{i+1} \varphi_2(t) + h_i [m_i \varphi_3(t) + m_{i+1} \varphi_4(t)], \quad (45)$$

но здесь его производные m_i в узлах сетки не известны.

Определенный таким образом кубический сплайн непрерывен вместе с первой производной. Величины m_i выбираются так, чтобы была непрерывна и вторая производная. Эти условия $S_3'(t_i - 0) = S_3'(t_i + 0)$ дают $N - 2$ равенства:

$$\lambda_i m_{i-1} + 2m_i + \mu_i m_{i+1} = c_i = 3\tilde{f}'_i, \quad i = 2, \dots, N - 1, \quad (46)$$

где \tilde{f}'_i определяются формулами (11б); $\lambda_i = h_i (h_i + h_{i-1})^{-1}$; $\mu_i = 1 - \lambda_i$.

Граничные уравнения, соответствующие крайним условиям:

$$I. 2m_1 = c_1^* = 2f'_1; \quad 2m_N = c_N^* = 2f'_N;$$

$$II. 2m_1 + m_2 = c_1^* = 3f'[t_1, t_2] - 0,5h_1 f''_1;$$

$$m_{N-1} + 2m_N = c_N^* = 3f'[t_{N-1}, t_N] + 0,5h_{N-1} f''_N.$$

В периодическом случае $f_1 = f_N$; $f_2 = f_{N+1}$; $m_1 = m_N$; $m_2 = m_{N+1}$; $h_N = h_1$ верно уравнение (46) при $i = 1$, и система имеет вид

$$2m_1 + \mu_1 m_2 + \lambda_1 m_{N-1} = c_1 = 3\tilde{f}'_1;$$

$$\lambda_i m_{i-1} + 2m_i + \mu_i m_{i+1} = c_i = 3\tilde{f}'_i, \quad i = 2, \dots, N - 2;$$

$$\mu_{N-1} m_1 + \lambda_{N-1} m_{N-2} + 2m_{N-1} = c_{N-1} = 3\tilde{f}'_{N-1}.$$

В случае граничных условий типа IV система будет

$$(1 + \gamma_1) m_2 + \gamma_1 m_3 = \tilde{f}'_2 + 2\gamma_1 f'[t_2, t_3];$$

$$\lambda_i m_{i-1} + 2m_i + \mu_i m_{i+1} = 3\tilde{f}'_i, \quad i = 3, \dots, N - 3;$$

$$\gamma_N m_{N-2} + (1 + \gamma_N) m_{N-1} = \tilde{f}'_{N-1} + 2\gamma_N f'[t_{N-2}, t_{N-1}],$$

где $\gamma_1 = h_1/h_2$; $\gamma_N = h_{N-1}/h_{N-2}$. Эта система определяет m_2, \dots, m_{N-1} , а m_1 и m_N находят из соотношений

$$m_1 + (1 - \gamma_1^2) m_2 - \gamma_1^2 m_3 = 2(f[t_1, t_2] - \gamma_1^2 f[t_2, t_3]);$$

$$-\gamma_N^2 m_{N-2} + (1 - \gamma_N^2) m_{N-1} + m_N = 2(f[t_{N-1}, t_N] - \gamma_N^2 f[t_{N-2}, t_{N-1}]).$$

Итак, построение интерполяционного кубического сплайна по формуле (45) сводится к отысканию значений m_i путем решения одной из полученных систем.

Общий вид этих систем $Az = g$. Их матрицы во всех четырех случаях ленточные с диагональным преобладанием, т. е., если a_{ij} — элемент i -й строки и j -го столбца, то $r_i = |a_{ii}| - \sum_{j \neq i} |a_{ij}| > 0$ для всех i . Определители таких матриц отличны от нуля, и системы имеют решения, притом единственные. Это значит, что интерполяционный кубический сплайн, удовлетворяющий условиям (44) и одному из типов I—IV крайних условий, существует и единствен.

Выбор граничных условий. Через граничные условия и конструкцию сплайна включаются параметры, выбирая которые, можно управлять его поведением, особенно возле концов отрезка $[a, b]$. Если $f(t)$ — периодическая функция, то используют крайние условия типа III. Если известны $f'(t)$ или $f''(t)$ в точках a и b , то естественно воспользоваться крайними условиями типа I или II. При этом, если существует выбор, то следует предпочесть первые. Если производные неизвестны, то в большинстве случаев наилучшим решением будет применение крайних условий типа IV.

Отметим также, что вместо значений производных можно использовать их разностные аналоги. При этом точность интерполяции на интервалах вблизи концов падает. Иногда предлагается принимать $S_3^*(f; a) = S_3^*(f; b) = 0$. В этом случае вблизи концов точность интерполяции функции и ее первой производной уменьшается и становится соизмеримой с точностью интерполяции сплайном первой степени, что резко ухудшает всю картину.

Алгоритм. Решения интерполяционных систем находятся методом прогонки (п. 10). Алгоритм прогонки устойчив по отношению к погрешностям округления. Поэтому, если исходные данные f_i заданы с погрешностью δ и, значит, погрешность прахих частей систем уравнений $\delta' = 3\delta (\min_i h_i)^{-1}$, то значения m_i будут определены с этой точностью. Вычисления сплайна и его производных осуществляются так же, как вычисления эрмитовых кубических сплайнов.

В рассмотренной задаче каждая из величин m_i , $i = 1, \dots, N$, а значит, и сплайн $S_3(t)$ зависят от всей совокупности исходных данных о функции и ее производных. Это пример нелокальной аппроксимации. Тем не менее сплайн имеет достаточно четко выраженные локальные свойства.

Локальные свойства. Значение функции $f_i = f(t_i)$ существенно влияет на поведение сплайна только на ближайших к i отрезках и затухает в удаленных промежутках в геометрической прогрессии со знаменателем, меньшим $1/2$. Иначе, если при построении сплайна на отрезке $[t_i, t_{i+1}]$ пренебречь информацией в узлах t_j таких, что $i - j > k$, $j - i > k + 1$, то допускаемая при этом погрешность составит примерно $(1/2)^k S_3(t)$. Граничные условия также существенно влияют на поведение сплайна только вблизи концов отрезка $[a, b]$ и слабо влияют при удалении от них на некоторое число шагов сетки Δ .

Если шаг $h_k = t_{k+1} - t_k$ значительно меньше других шагов, то вид сплайна на $[a, t_k]$ существенно зависит только от значений функции в узлах t_1, \dots, t_k , граничного условия в t_1 и значения f_k . Аналогично, вид сплайна на $[t_{k+1}, b]$ не зависит от значений функции на $[a, t_k]$. В пределе, когда получается двойной узел $t_k = t_{k+1}$, имеем два полностью независимых сплайна слева и справа от этого узла. Если узлы сетки размещены попарно близко, то в пределе, когда они сливаются, кубический сплайн с граничными условиями типа I стремится к эрмитову кубическому сплайну.

Указанным свойством кубического сплайна можно воспользо-
ваться для устранения отрицательного влияния (в виде осцилляций)
разрыва второй производной функции $f(t)$. Для этого точку раз-
рыва следует поместить между двумя «близкими» узлами.

Погрешность интерполяции. Вследствие нелокальности интер-
полляции оценки погрешности могут быть даны только через харак-
теристики интерполируемой функции на всем отрезке $[a, b]$.

Сначала получим оценки погрешности производных в узлах
сетки Δ . Рассмотрим, к примеру, систему в периодическом случае.
Преобразуем ее уравнения к виду

$$\lambda_i (m_{i-1} - f_{i-1}) + 2(m_i - f_i) + \mu_i (m_{i+1} - f_{i+1}) = \bar{c}_i, \quad i = 1, \dots, N-1,$$

где $\bar{c}_i = 3f'_i - \lambda_i f'_{i-1} - 2f'_i - \mu_i f'_{i+1}$.

Матрица системы относительно величин $m_i - f_i$ не изменилась,
т. е. осталась с диагональным преобладанием. Для решения систем
 $Az = g$ с такими матрицами справедливо неравенство

$$\max_i |z_i| \leq (\min_i r_i)^{-1} \max_i |g_i|.$$

В нашем случае $\min_i r_i \geq 1$, а $\max_i |\bar{c}_i|$ находится стандартным
способом. Для функций $f(t) \in C^{(2)}[a, b]$ получается неравенство

$$\max_i |m_i - f_i| \leq \frac{5}{6} H \|f''(t)\|, \quad (47)$$

а для функций $f(t) \in C^{(4)}[a, b]$

$$\max_i |m_i - f_i| \leq \frac{1}{24} H^3 \|f^{IV}(t)\|. \quad (48)$$

Далее рассмотрим неравенство

$$|S_3^{(p)}(f) - f^{(p)}(t)| \leq |S_{3,2}^{(p)}(t) - f^{(p)}(t)| + |S_3^{(p)}(t) - S_{3,2}^{(p)}(t)|. \quad (49)$$

Оценки первого слагаемого в правой части имеются — это (33),
(34). Для второго слагаемого на интервале $[t_i, t_{i+1}]$ на основании
формулы для сплайнов (45) и (32) получаем при $t \in [t_i, t_{i+1}]$

$$|S_3^{(p)}(t) - S_{3,2}^{(p)}(t)| \leq c_p h_i^{1-p} \max\{|m_i - f_i|, |m_{i+1} - f_{i+1}|\}. \quad (50)$$

Тогда из (49), учитывая (33), (47) и (50), для $f(t) \in C^{(2)}[a, b]$
находим

$$\|S_3^{(p)}(t) - f^{(p)}(t)\| \leq K_p H^{2-p} \|f''(t)\|, \quad p = 0, 1, \quad (51)$$

где $K_0 = 13/48$; $K_1 = 0,8623$.

Если $f(t) \in C^{(4)}[a, b]$, аналогично получаем

$$\|S_3^{(p)}(t) - f^{(p)}(t)\| \leq K_p H^{4-p} \|f^{IV}(t)\|, \quad p = 0, 1, \quad (52)$$

где $K_0 = 5/384$; $K_1 = 1/24$.

Оценки (52) остаются справедливыми и для функций с разрывами
третьих производных в узлах сетки Δ .

Приведенные результаты справедливы для задач с граничными
условиями типов I—III, а для граничных условий типа IV они верны

при $\gamma_1 \leq 0,36$, $\gamma_N \leq 0,36$. Впрочем, влияние γ_1 и γ_N сказывается только вблизи концов отрезка $[a, b]$.

Если $f(t) \in C^4(a, b)$, то для условий типов I и III при равномерной сетке оценки совпадают (с точностью до малых более высокого порядка) с оценками (34) для эрмитовых сплайнов.

Интерполяция с заданной точностью. На равномерной сетке она осуществляется точно так же, как для эрмитовых кубических сплайнов, с помощью оценок (51) или (52) при $p = 0$.

Если $f(t) \in C^4[a, b]$, то для построения сеток, близких к равномерным в том смысле, что

$$h_i = h_{i-1} + O(h_{i-1}^2), \quad (53)$$

на основании сделанного выше замечания можно использовать методику для эрмитовых кубических сплайнов.

Рассмотрим случай $f(t) \in C^{(2)}[a, b]$. Для построения сетки Δ , обеспечивающей интерполяцию с точностью ε , будем использовать формулу (49). Потребуем, чтобы слагаемые в ее правой части по отдельности не превышали значений

$$\|S_{3,2}(t) - f(t)\| \leq \varepsilon_2; \quad \|S_3(t) - S_{3,2}(t)\| \leq \varepsilon_3,$$

где $\varepsilon_2 + \varepsilon_3 = \varepsilon$ — заданная точность приближения.

Оценка приближения эрмитовым сплайном $S_{3,2}(t)$ имеется — это формула (33) с коэффициентом $K_0 = 1/16$. Обозначим

$$\delta_i S_3^*(t) = \max\{|m_i - f_i|, |m_{i+1} - f_{i+1}|\}.$$

Из (50), учитывая (47), получаем

$$\|S_3(t) - S_{3,2}(t)\| \leq \frac{1}{4} h_i \delta_i S_3^*(t) \leq \frac{5}{24} h_i H \|f''(t)\| \leq \varepsilon_3.$$

Значения ε_2 и ε_3 можно выбрать пропорциональными числам $1/16$ и $5/24$, т. е. $\varepsilon_2 = 3\varepsilon/13$, $\varepsilon_3 = 10\varepsilon/13$.

Искомую сетку Δ будем определять последовательными приближениями, начиная с сетки $\Delta^{(0)}$, на которой эрмитов сплайн $S_{3,2}(t)$ интерполирует $f(t)$ с погрешностью $3\varepsilon/13$. Ее можно получить с помощью алгоритма, описанного в п. 5 данной главы. Итерация с номером l включает следующие действия. На сетке $\Delta^{(l-1)}$ строится кубический сплайн $S_3(t)$, и на каждом отрезке $[t_i, t_{i+1}]$ проверяется условие

$$0,25h_i |\delta_i S_3^*(t)| \leq 10\varepsilon/13. \quad (54)$$

Если оно не выполняется, то точка $t_i^* = (t_i + t_{i+1})/2$ — середина отрезка $[t_i, t_{i+1}]$ — включается в число узлов. Полученные таким образом узлы $\{t_i^*\}$ вместе с разбиением $\Delta^{(l-1)}$ образуют новую сетку $\Delta^{(l)}$. Процесс заканчивается, как только $\Delta^{(l)}$ совпадает с $\Delta^{(l-1)}$.

Основные затраты при использовании этого алгоритма приходятся на определение начального приближения $\Delta^{(0)}$.

Кубические сплайны двух переменных. Интерполяция. Рассматривается множество кубических сплайнов $S_3(t, s)$ на сетке $\Delta = \Delta_t \times \Delta_s$ в области $\Omega = [a, b] \times [c, d]$. Задача интерполяции

заключается в построении сплайна, принимающего в узлах сетки заданные значения

$$S_3(t_i, s_j) = f_{ij}, \quad i = 1, \dots, N; \quad j = 1, \dots, M, \quad (55)$$

и удовлетворяющего граничным условиям. Они, как и в случае одной переменной, четырех типов:

$$I. D^{1,0} S_3(t_i, s_j) = f_{ij}^{(1,0)}, \quad i = 1, N; \quad j = 1, \dots, M;$$

$$D^{0,1} S_3(t_i, s_j) = f_{ij}^{(0,1)}, \quad i = 1, \dots, N; \quad j = 1, M;$$

$$D^{1,1} S_3(t_i, s_j) = f_{ij}^{(1,1)}, \quad i = 1, N; \quad j = 1, M;$$

$$II. D^{2,0} S_3(t_i, s_j) = f_{ij}^{(2,0)}, \quad i = 1, N; \quad j = 1, \dots, M;$$

$$D^{0,2} S_3(t_i, s_j) = f_{ij}^{(0,2)}, \quad i = 1, \dots, N; \quad j = 1, M;$$

$$D^{2,2} S_3(t_i, s_j) = f_{ij}^{(2,2)}, \quad i = 1, N; \quad j = 1, M.$$

III. Условия периодичности сплайна и его производных до порядка $D^{p,q} S_3(t, s)$, $p, q = 0, 1, 2$ по переменной t с периодом $b - a$ и по переменной s с периодом $d - c$.

IV. Условия непрерывности старших производных, а именно: на линиях $t = t_2$ и $t = t_{N-1}$ непрерывны все производные до порядка $D^{3,2} S_3(t, s)$, а на линиях $s = s_2$ и $s = s_{M-1}$ — до порядка $D^{2,3} S_3(t, s)$. Кроме того, в областях $[t_r, t_{r+2}] \times [s_k, s_{k+2}]$, $r = 1, N - 2$, $k = 1, M - 2$, прилегающих к вершинам прямоугольника Ω , непрерывны все производные, включая $D^{3,3} S_3(t, s)$. Кроме перечисленных, возможны и смешанные краевые условия, т. е. условия разных типов по разным переменным.

Кубический сплайн $S_3(t, s)$ может быть также представлен в виде (41), но с матрицей F_m , отличающейся от F тем, что вместо $f_{ij}^{(p,q)}$ здесь стоят величины $m_{ij}^{(p,q)}$, $p, q = 0, 1$:

$$S_3(t, s) = \varphi^T(u) F_m \varphi(v). \quad (56)$$

Задача состоит в том, чтобы из уравнений (55) и граничных условий найти неизвестные величины $m_{ij}^{(1,0)}$, $m_{ij}^{(0,1)}$ и $m_{ij}^{(1,1)}$.

Алгоритм. Исходные данные представляются в виде табл. 1. Во внутренней части таблицы размещены значения функции. Окаймляющие строки и столбцы заполняются только в случае краевых условий типов I и II. Если краевые условия смешанные, например по переменной t типа I или II, а по s — типа III или IV, то свободными остаются граничные строки.

Алгоритм основан на том, что на линиях сетки, например $s = s_j$, сплайн $S_3(t, s)$ и его частные производные по s являются кубическими сплайнами от переменной t .

Шаг 1. Строятся кубические сплайны от переменной t , $S_3(t, s_j)$, $j = 0, \dots, M$, по строкам (см. табл. 1), включая граничные (если они имеются) с краевыми условиями из граничных столбцов. Задача сводится к решению систем уравнений, число которых зависит от вида краевых условий и равно $M + 2$ (типы I, II), M (тип IV)

ТАБЛИЦА 1. Исходные данные задачи интерполирования кубическим сплайном двух переменных

$f_{1M}^{(p, q)}$	$f_{1M}^{(0, q)}$	$f_{2M}^{(0, q)}$...	$f_{NM}^{(0, q)}$	$f_{NM}^{(p, q)}$
$f_{1M}^{(p, 0)}$	f_{1M}	f_{2M}	...	f_{NM}	$f_{NM}^{(p, 0)}$
.
.
$f_{12}^{(p, 0)}$	f_{12}	f_{22}	...	f_{N2}	$f_{N2}^{(p, 0)}$
$f_{11}^{(p, 0)}$	f_{11}	f_{21}	...	f_{N1}	$f_{N1}^{(p, 0)}$
$f_{11}^{(p, q)}$	$f_{11}^{(0, q)}$	$f_{21}^{(0, q)}$...	$f_{N1}^{(0, q)}$	$f_{N1}^{(p, q)}$

и $M - 1$ (тип III). В результате находим $m_{ij}^{(1, 0)} = D^{1, 0} S_3(t_i, s_j)$, $(t_i, s_j) \in \Delta$. Затем эти значения располагаем снова в виде таблицы, аналогичной табл. 1 без граничных столбцов. В граничных строках при условиях типа I и II помещаются значения $m_{ij}^{1, q} = D^{1, q} S_3(t_i, s_j)$, $q = 1$, или 2; $i = 1, \dots, N$; $j = 1, M$.

Шаг 2. По данным табл. 1 строятся сплайны $S_3(t_i, s)$, $i = 1, \dots, N$, в результате находятся значения $m_{ij}^{(0, 1)} = D^{0, 1} S_3(t_i, s_j)$. Причем на этом шаге решается N одномерных задач в непериодическом случае и $N - 1$ задач — в периодическом.

Шаг 3. Строятся кубические сплайны по переменной s , $\tilde{S}_3(t_i, s)$, $i = 1, \dots, N$, по столбцам второй таблицы. Это будут, очевидно, частные производные $D^{1, 0} S_3(t_i, s)$ по t искомого сплайна на линиях $t = t_i$. Значения производных сплайнов $D_{s_i}^{0, 1} \tilde{S}_3(t_i, s)$ в узлах сетки Δ представляют собой смешанные производные искомого сплайна на сетке, т. е. $m_{ij}^{(1, 1)} = D^{1, 1} S_3(t_i, s_j)$. Число решаемых задач такое же, как на втором шаге.

Построение двумерного сплайна свелось к решению $2N + M + 2$ (не менее $2N + M - 3$) одномерных задач. Можно пользоваться и другим вариантом описанного алгоритма, поменяв ролями переменные t и s . В этом случае придется решить $N + 2M + 2$ (не менее $N + 2M - 3$) одномерных задач.

Одномерные задачи можно решать методом дробонки. В этом случае при решении задач интерполяции по одной из переменных коэффициенты левых частей уравнений в системах будут одинаковыми, что избавит нас от повторения некоторой части вычислений. Вычисление сплайна проводится по формуле (56).

Погрешность интерполяции. Для граничных условий типов I, II, III справедливы оценки (42а), (42б), но со значениями констант K_0, K_1 , как в (51) и (52). Для $f(t, s) \in C^{4, 4}[\Omega]$ при равномерных сетках Δ_t и Δ_s коэффициенты в (42б) сохраняются.

Интерполяция с заданной точностью. Для сплайна $S_3(t, s)$ интерполяция осуществляется, как и для эрмитового сплайна $S_{3,2}(t, s)$ (см. п. 5), путем независимого построения сеток Δ_t и Δ_s , обеспечивающих точность приближения ε_t и ε_s . В случае, если интерполируемая функция $f(t, s) \in C^{4,4}[\Omega]$, то достаточно взять сетку Δ эрмитова сплайна $S_{3,2}(t, s)$, контролируя ограничения типа (53) по шагам h_i и l_j . Если $f(t, s) \in C^{(2,2)}[\Omega]$, то сетка Δ_t для эрмитова сплайна должна обеспечивать аппроксимацию с погрешностью $3\varepsilon_t/13$, а погрешность, равная $10\varepsilon_t/13$, ограничивает отклонение сплайна $S_3(t, s)$ от эрмитова в соответствии с формулой (54). Аналогичное обстоятельство учитывается и при построении сетки Δ_s . Хотя описанные процедуры принципиально не сложны, тем не менее нахождение двумерных сеток требует известного опыта исследователя.

Другая схема решения задачи интерполяции. В ряде случаев более удобным является иное представление интерполяционного кубического сплайна, в котором вместо величин m присутствуют $M_i = S_3'(t_i)$, $i = 1, \dots, N$, а именно: при $t \in [t_i, t_{i+1}]$

$$S_3(t) = \psi_1(u) f_i + \psi_2(u) f_{i+1} + (1/6) h_i^2 [\psi_3(u) M_i + \psi_4(u) M_{i+1}], \quad (57)$$

где $\psi_1(u) = 1 - u$; $\psi_2(u) = u$; $\psi_3(u) = -u(1-u)(2-u)$;
 $\psi_4(u) = -u(1-u)^2$.

Формула (57) обеспечивает непрерывность функции $S_3(t)$ и ее второй производной в узлах сетки Δ . Величины M_i находятся из условия непрерывности в узлах первой производной, т. е. $S_3'(t_i-0) = S_3'(t_i+0)$. Эти равенства вместе с крайними условиями одного из типов I—IV образуют систему относительно неизвестных M_i . Для условий типов I и II она имеет вид

$$\begin{aligned} 2M_1 + \lambda_1^* M_2 &= d_1^*; \\ \mu_i M_{i-1} + 2M_i + \lambda_i M_{i+1} &= d_i = 6f[t_{i-1}, t_i, t_{i+1}], \quad i = 2, \dots, N-1; \\ \mu_N^* M_{N-1} + 2M_N &= d_N^*. \end{aligned} \quad (58)$$

В случае граничных условий типа I $\lambda_1^* = \mu_N^* = 1$,

$$d_1^* = \frac{6}{h_1} \left(\frac{f_2 - f_1}{h_1} - f_1' \right); \quad d_N^* = \frac{6}{h_{N-1}} \left(f_N' - \frac{f_N - f_{N-1}}{h_{N-1}} \right),$$

а для условий типа II $\lambda_1^* = \mu_N^* = 0$, $d_1^* = 2f_1''$, $d_N^* = 2f_N''$.

В периодическом случае система имеет вид

$$\begin{aligned} 2M_1 + \lambda_1 M_2 + \mu_1 M_{N-1} &= d_1; \\ \mu_i M_{i-1} + 2M_i + \lambda_i M_{i+1} &= d_i, \quad i = 2, \dots, N-2; \\ \lambda_{N-1} M_1 + \mu_{N-1} M_{N-2} + 2M_{N-1} &= d_{N-1}; \quad M_1 = M_N. \end{aligned} \quad (58a)$$

Можно выписать уравнения и при граничных условиях типа IV. Матрицы систем во всех четырех случаях — по-прежнему матрицы с диагональным преобладанием. Системы решаются методом прогонки.

О выпуклости кубического сплайна. Для исследования некоторых свойств решения систем удобнее использовать не уравнения (58), а преобразованные уравнения. Умножим i -е уравнение в (58) на 2 и вычтем из него $(i - 1)$ -е, умноженное на μ_i , и $(i + 1)$ -е, умноженное на λ_i . Таким образом получаем уравнения

$$\begin{aligned} -\mu_{i-1}\mu_i M_{i-2} + (3 + \mu_{i-1}\mu_i + \lambda_i\lambda_{i+1}) M_i - \lambda_i\lambda_{i+1} M_{i+2} = \\ = 2d_i - \mu_i d_{i-1} - \lambda_i d_{i+1}, \quad i = 3, \dots, N - 2, \end{aligned} \quad (59)$$

плюс четыре уравнения, зависящие от граничных условий.

При интерполяции функции часто требуется сохранить некоторые ее качественные характеристики. Например, если $f(t)$ — выпуклая вниз функция ($f''(t) > 0$), то естественно потребовать, чтобы сплайн был тоже выпуклым. Из условия выпуклости функции $f(t)$ следует, что $d_i > 0$, $i = 1, \dots, N$. Так как в силу (57) на отрезке $[t_i, t_{i+1}]$ производная $S_3''(t) = \int_{t_i}^t (1-u) M_i + u M_{i+1}$, то $S_3''(t) \geq 0$, если все $M_i \geq 0$.

Матрица системы (59) — с диагональным преобладанием, причем диагональные элементы положительны, а недиагональные — отрицательны. Она принадлежит к числу матриц монотонного вида [8]. Так называются матрицы A систем $Az = g$, если из условий $g_i \geq 0$ вытекает $z_i \geq 0$, $i = 1, \dots, N$. Следовательно, если $2d_i - \mu_i d_{i-1} - \lambda_i d_{i+1} \geq 0$, $i = 2, \dots, N - 1$, и $2d_1 - \lambda_1 d_2 \geq 0$, $2d_N - \mu_N d_{N-1} \geq 0$, то $M_i \geq 0$, $i = 1, \dots, N$, и $S_3''(t) \geq 0$ на $[a, b]$. Получены достаточные условия выпуклости. Если они нарушены, то функция $S_3''(t)$ может быть знакопеременной, что проявляется в виде осцилляций графика сплайна $S_3(t)$.

Аппроксимация вторых производных. Из систем уравнений (59) с помощью стандартных приемов, как и при выводе оценки (47), находим для функций $f(t) \in C^{(2)}[a, b]$

$$|M_i| \leq (11/5) \|f''(t)\|.$$

Для $f(t) \in C^4[a, b]$, аналогично (48), можно получить

$$|M_i - f''_i| \leq H^2 \|f^{IV}(t)\|/6. \quad (60)$$

Достаточно кропотливый анализ показывает, что в этом случае наибольшие отклонения вторых производных сплайна и функции $f(t)$ достигаются в узлах сетки. Следовательно,

$$\|S_3''(t) - f''(t)\| \leq H^2 \|f^{IV}(t)\|/6. \quad (61)$$

Эта оценка дополняет формулу (52).

Обобщенные кубические сплайны. Так называются функции $S_G(t) \in C^2[a, b]$, которые на каждом отрезке $[t_i, t_{i+1}]$ имеют вид

$$\begin{aligned} S_G(t) = A_i(1-u) + B_i u + C_i(1-u) \left[\frac{(1-u)^2}{1+\psi_i(u)} - 1 \right] + \\ + D_i u \left[\frac{u^2}{1+\psi_i(u)} - 1 \right], \end{aligned}$$

где $\varphi_i(u)$ и $\psi_i(u)$ — заданные дважды непрерывно дифференцируемые функции при $u \in [0, 1]$. Предполагаются выполненными условия $\varphi_i(u) > -1$; $\psi_i(u) > -1$; $\varphi_i(0) = 0$; $\psi_i(1) = 0$.

При интерполяции сплайном $S_G(t)$ имеем $A_i = f_i$; $B_i = f_{i+1}$. Коэффициенты C_i и D_i определяются так, чтобы $S_G(t) \in C^2[a, b]$. Если обозначить $M_i = S_G''(t_i)$, то $C_i = h_i^2 \Phi_i M_i$; $D_i = h_i^2 \Psi_i M_{i+1}$, где

$$\Phi_i = [6 - \varphi_i''(0) + 2(3 + \varphi_i'(0))\varphi_i'(0)]^{-1};$$

$$\Psi_i = [6 - \psi_i''(1) - 2(3 - \psi_i'(1))\psi_i'(1)]^{-1}$$

при ограничениях $0 < \Phi_i < \infty$, $0 < \Psi_i < \infty$.

Из условия непрерывности $S_G(t)$ в узлах t_i получаем

$$\begin{aligned} \mu_i \Phi_{i-1} M_{i-1} + [\mu_i (2 - \psi_i'(1)) \Psi_{i-1} + \lambda_i (2 + \varphi_i'(0)) \Phi_i] M_i + \lambda_i \Psi_i M_{i+1} = \\ = f[t_{i-1}, t_i, t_{i+1}], \quad i = 2, \dots, N-1. \end{aligned}$$

Для однозначности определения сплайна $S_G(t)$ требуются два крайних условия. Их можно брать такими же, как и в случае кубического сплайна.

Ограничения на функции $\psi_i(t)$ и $\varphi_i(t)$, достаточные для существования обобщенного сплайна, весьма слабые, и это позволяет строить сплайны с самыми разнообразными свойствами.

Например, положим

$$\varphi_i(u) = p_i [1 - (1 - u)^\alpha]; \quad \psi_i(u) = q_i [1 - u^\alpha],$$

где $\alpha > 0$ — целое; $p_i \geq 0$, $q_i \geq 0$ — постоянные. Это рациональные сплайны.

Справедливы следующие утверждения. Если значения f_i таковы, что вторые разделенные разности по точкам (t_{i-1}, t_i, t_{i+1}) и (t_i, t_{i+1}, t_{i+2}) положительны, то при всех достаточно больших p_{i-1} , p_i , q_i , q_{i+1} выполняется неравенство $S_G''(t) > 0$, $t \in [t_i, t_{i+1}]$, т. е. сплайн $S_G(t)$ на отрезке $[t_i, t_{i+1}]$ будет выпуклой функцией.

Если значения f_i таковы, что для некоторого i выполнено условие $f_{i-1} < f_i < f_{i+1} < f_{i+2}$, то при всех достаточно больших p_{i-1} , p_i , q_i , q_{i+1} выполняется неравенство $S_G''(t) > 0$, $t \in [t_i, t_{i+1}]$, т. е. сплайн $S_G(t)$ на отрезке $[t_i, t_{i+1}]$ будет монотонно возрастающей функцией.

Очевидно, при $p_i = q_i = 0$, $i = 1, \dots, N-1$, рациональный сплайн превращается в кубический. С другой стороны, справедливо отождествление

$$\lim_{p_i, q_i \rightarrow \infty} S_G(t) = (1 - u)f_i + uf_{i+1}, \quad t \in [t_i, t_{i+1}],$$

и сплайн первой степени тоже можно считать частным случаем рационального сплайна.

Перечисленные свойства позволяют эффективно использовать обобщенные сплайны для приближения с сохранением монотонности и выпуклости (вогнутости) функций. Не вызывает затруднений и

наличие точек излома или участков с большими градиентами. В этом большое преимущество обобщенных сплайнов перед обычными кубическими.

7. ЗАДАЧА СГЛАЖИВАНИЯ КУБИЧЕСКИМИ СПЛАЙНАМИ¹

Экстремальное свойство интерполяционных кубических сплайнов. Рассмотрим задачу о минимизации функционала

$$J(f) = \int_a^b |f''(t)|^2 dt.$$

В качестве множества допустимых функций возьмем класс $C^{(2)}[a, b]$. Будем рассматривать также подклассы: а) $\bar{C}^{(2)}[a, b]$ периодических функций с периодом $b - a$ и б) $\bar{C}^{(2)}[a, b]$ функций, удовлетворяющих краевым условиям $f'(a) = z'_1$ и $f'(b) = z'_N$ (z'_1 и z'_N — заданные числа).

Пусть с сеткой Δ связано множество кубических сплайнов $S_{3,1}(\Delta)$. Экстремальное свойство состоит в том, что среди всех функций $f(t) \in C^{(2)}[a, b]$, принимающих заданные значения $f(t_i) = z_i^0$, $i = 1, \dots, N$ на сетке Δ , кубический сплайн $S_3(t)$ с краевыми условиями

$$S_3^-(a) = S_3^+(b) = 0, \quad (62)$$

минимизирует функционал $J(f)$. Если функции $f(t)$ принадлежат одному из классов $\bar{C}^{(2)}[a, b]$, $\bar{C}^{(2)}[a, b]$, то минимум функционала доставляет кубический сплайн из того же множества. Во всех случаях такой сплайн единствен.

Сглаживающие кубические сплайны. Если исходные данные содержат погрешность, т. е. имеются z_i^0 такие, что

$$|f_i - z_i^0| \leq \delta_i, \quad \delta_i \geq 0, \quad i = 1, \dots, N, \quad (63)$$

то строить интерполяционный сплайн не имеет смысла. Попробуем получить более «гладкий» сплайн (рис. 6).

Будем минимизировать функционал $J(f)$ на множестве функций $C^{(2)}[a, b]$ при условиях (63) (задача на условный экстремум). Оказывается, что среди функций $f(t) \in C^{(2)}[a, b]$, с ограничениями (63) кубический сплайн $S_3(t)$, удовлетворяющий условиям (62), минимизирует функционал $J(f)$. Такой сплайн называется *сглаживающим*. Если функции $f(t)$ принадлежат одному из классов $\bar{C}^{(2)}[a, b]$ или $\bar{C}^{(2)}[a, b]$ с теми же ограничениями (63), то минимум функционалу доставляет кубический сплайн из того же множества.

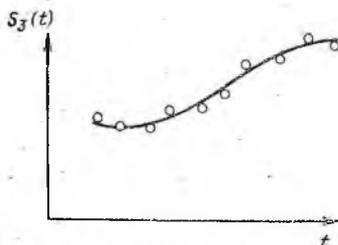


Рис. 6. Сглаживающий кубический сплайн

¹ В подготовке материала этого параграфа принял участие Н. Н. Павлов.

Действительно, если для каждой функции $f(t)$ построить интерполяционный кубический сплайн, то, как следует из экстремального свойства, $J[S_3(t)] \leq J(f)$. Поэтому функция, минимизирующая $J(f)$, если она существует, должна быть кубическим сплайном.

Введем обозначение для разрыва третьей производной сплайна в узлах сетки Δ :

$$D_i = \begin{cases} S_3'''(t_i + 0); \\ S_3'''(t_i + 0) - S_3'''(t_i - 0), & i = 2, \dots, N-1; \\ -S_3'''(t_N - 0). \end{cases} \quad (64)$$

Для сглаживающего сплайна выполняются следующие соотношения:

$$\begin{aligned} D_i &\geq 0, & \text{при } S_3(t_i) = z_i^0 - \delta_i; \\ D_i &\leq 0, & \text{при } S_3(t_i) = z_i^0 + \delta_i; \\ D_i &= 0, & \text{при } z_i^0 - \delta_i < S_3(t_i) < z_i^0 + \delta_i. \end{aligned} \quad (65)$$

Обратно, если для сплайна выполняются условия (65), то он является сглаживающим. Формулы (65), устанавливающие свойства сглаживающих сплайнов, называются *условиями характеристики*.

Существование и единственность решения задачи сглаживания. Покажем теперь, что *сглаживающий сплайн существует*. Заметим, что если все $\delta_i = 0$, то задача вырождается в задачу интерполирования, которая имеет решение и притом единственное.

Рассмотрим функционал $J(S_3)$. Он может быть преобразован к виду

$$\int_a^b [S_3''(t)]^2 dt = \sum_{i=1}^N D_i S_3(t_i) + S_3'(t) S_3''(t) \Big|_a^b. \quad (66)$$

Согласно представлению сплайна (57)

$$D_i = (M_2 - M_1)/h_1; \quad (67a)$$

$$D_i = (M_{i+1} - M_i)/h_i - (M_i - M_{i-1})/h_{i-1}, \quad i = 2, \dots, N-1; \quad (67b)$$

$$D_N = (M_N - M_{N-1})/h_{N-1}. \quad (67b)$$

В периодическом случае $h_1 = h_N$, $M_1 = M_N$, $M_2 = M_{N+1}$, и все величины D_i определяются формулами (67b), где $i = 1, \dots, N-1$.

С другой стороны, величины M_i и значения сплайна $z_i = S_3(t_i)$ связаны линейными системами (58) или (58a) (через величины d_i). Следовательно, D_i — линейные комбинации значений z_i , $i = 1, \dots, N$, и сумма справа в (66) есть квадратичная форма этих значений. В случае, если $S_3(t) \in C^{(2)}[a, b]$ или $\bar{C}^{(2)}[a, b]$, второе слагаемое в правой части (66) равно нулю, а если $S_3(t) \in \bar{C}^{(2)}[a, b]$, оно равно $z_N M_N - z_1 M_1$, т. е. является линейной формой от значений сплайна z_i .

Следовательно, функционал $F(z) = J(S_3)[z = (z_1, \dots, z_N) - \text{вектор}]$ есть неотрицательная квадратичная форма относительно величин z_i . Его минимум отыскивается в параллелепипеде

$$z_i^0 - \delta_i \leq z_i \leq z_i^0 + \delta_i, \quad i = 1, \dots, N. \quad (68)$$

Неотрицательная квадратичная форма в выпуклой замкнутой области, какой является параллелепипед, достигает наименьшего значения (подобно многочлену второй степени на отрезке). Итак, задача сглаживания имеет решение.

Вопрос единственности сглаживающего сплайна не является таким простым, как для интерполяционного сплайна, и на него не может быть дан однозначный ответ. Однако в каждом классе можно указать достаточные условия его единственности.

Если среди сплайнов из $\tilde{C}^{(2)} [a, b]$, удовлетворяющих ограничениям (63), не содержится ни одного многочлена нулевой степени, то сглаживающий сплайн из этого класса единствен. Условием единственности в классе $C^{(2)} [a, b]$ будет отсутствие многочленов нулевой и первой степени, а для класса $\tilde{C}^{(2)} [a, b]$ — многочленов второй степени и ниже.

Таким образом, вопрос об единственности сглаживающего сплайна существенно зависит от выбора ограничений (63). Если величины δ_i малы, что означает близость сглаживающего сплайна к интерполяционному, то естественно ожидать его единственность. При больших δ_i единственности может и не быть.

Алгоритм. Рассмотрим задачу построения сглаживающего сплайна, которая, как было отмечено выше, сводится к минимизации квадратичного функционала $F(z)$ в параллелепипеде (68). Решение будем искать *методом штрафных функций* [7].

Нетрудно видеть, что наша задача эквивалентна задаче отыскания минимума функционала

$$I = F(z) + \sum_{i=1}^N \chi_i(z_i),$$

где функции $\chi_i(z_i)$ определены следующим образом:

$$\chi_i(z_i) = \begin{cases} 0, & \text{если } z_i^0 - \delta_i \leq z_i \leq z_i^0 + \delta_i; \\ +\infty, & \text{если } z_i < z_i^0 - \delta_i, z_i > z_i^0 + \delta_i. \end{cases}$$

Рассмотрим последовательность функционалов

$$I_L(z) = F(z) + \sum_{i=1}^N V_i^L(z_i), \quad (69)$$

где L пробегает значения 1, 2, ..., а $V_i^L(z_i)$ — выпуклые функции, причем $\lim_{L \rightarrow \infty} V_i^L(z_i) = \chi_i(z_i)$ для любых z_i .

Пусть z^L доставляет минимум функционалу $I_L(z)$. Тогда последовательность $\{z^L\}$ сходится к решению исходной задачи. Положим

$$V_i^L(z_i) = r_L [(z_i - z_i^0 - \delta_i)_+^3 + (z_i^0 - z_i - \delta_i)_+^3] + (z_i - z_i^0)^2 / r_L,$$

а последовательность $\{r_L\}$ выбирается такой, что $r_L > 0$ и $\lim_{L \rightarrow \infty} r_L = +\infty$.

где $u_i = 2(h_{i-1} + h_i)$;

$$H = \begin{bmatrix} 0 & 0 & 0 \\ h_1^{-1} & -(h_1^{-1} + h_2^{-1}) & h_2^{-1} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ h_{N-2}^{-1} & -(h_{N-2}^{-1} + h_{N-1}^{-1}) & h_{N-1}^{-1} \\ 0 & 0 & 0 \end{bmatrix};$$

$$R^{(k)} = \begin{bmatrix} \rho_1^{(k)} \\ \cdot \\ \cdot \\ \cdot \\ \rho_N^{(k)} \end{bmatrix}$$

Тогда система (70) в матричной форме примет вид

$$z + R^{(k)} H^T M = z^{(k)} - 0,5q_i^{(k)}. \quad (71)$$

Величины z_i и M_i связаны соотношениями $AM = 6Hz$.

Подставляя z из (71) для случая краевых условий (62), получаем

$$(A + 6HR^{(k)}H^T)M = 6H[z^{(k)} - 0,5q_i^{(k)}]. \quad (72)$$

Матрица $A + 6HR^{(k)}H^T$ системы (72) положительно определенная [8]. Следовательно, она невырождена, из чего следуют существование и единственность решений. Последнее утверждение справедливо для всех трех случаев: $C^{(2)}[a, b]$; $\tilde{C}^{(2)}[a, b]$; $\bar{C}^{(2)}[a, b]$.

Исключая периодический случай, систему уравнений (72), имеющую пятидиагональную структуру, можно записать в виде

$$\begin{aligned} a_1 M_1 + b_1 M_2 + c_1 M_3 &= g_1; \\ b_1 M_1 + a_2 M_2 + b_2 M_3 + c_2 M_4 &= g_2; \\ c_{i-2} M_{i-2} + b_{i-1} M_{i-1} + a_i M_i + b_i M_{i+1} + c_i M_{i+2} &= g_i, \quad i = 3, \dots, N-2; \\ c_{N-3} M_{N-3} + b_{N-2} M_{N-2} + a_{N-1} M_{N-1} + b_{N-1} M_N &= g_{N-1}; \\ c_{N-2} M_{N-2} + b_{N-1} M_{N-1} + a_N M_N &= g_N. \end{aligned} \quad (73)$$

Коэффициенты этой системы определяются формулами

$$\begin{aligned} a_i &= \frac{1}{3}(h_{i-1} + h_i) + \frac{1}{h_{i-1}^2} \rho_{i-1}^{(k)} + \left(\frac{1}{h_{i-1}} + \frac{1}{h_i} \right)^2 \rho_i^{(k)} + \frac{1}{h_i^2} \rho_{i+1}^{(k)}, \\ & \quad i = 2, \dots, N-1; \\ b_i &= \frac{1}{6} h_i - \frac{1}{h_i} \left[\left(\frac{1}{h_{i-1}} + \frac{1}{h_i} \right) \rho_i^{(k)} + \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right) \rho_{i+1}^{(k)} \right], \\ & \quad i = 2, \dots, N-2; \\ c_i &= \frac{1}{h_i h_{i+1}} \rho_{i+1}^{(k)}, \quad i = 2, \dots, N-3; \end{aligned} \quad (74)$$

$$g_i = \frac{y_{i+1}^{(k)} - y_i^{(k)}}{h_i} - \frac{y_i^{(k)} - y_{i-1}^{(k)}}{h_{i-1}}, \quad i = 2, \dots, N-1,$$

где $y_i^{(k)} = z_i^{(k)} - 0,5q_i^{(k)}$.

Если сглаживающий сплайн $S_3(t)$ удовлетворяет условиям $S_3^*(a) = S_3^*(b) = 0$, то $a_1 = a_N = 1$; $b_1 = c_1 = c_{N-2} = b_{N-1} = g_1 = g_N = 0$.

Если $S_3(t)$ удовлетворяет условиям $S_3'(a) = z'_1$; $S_3'(b) = z'_N$, то

$$\begin{aligned} a_1 &= \frac{h_1}{3} + \frac{1}{h_1^2} (\rho_1^{(k)} + \rho_2^{(k)}); & g_1 &= \frac{y_2^{(k)} - y_1^{(k)}}{h_1} - z'_1; \\ b_1 &= \frac{h_1}{6} - \frac{1}{h_1} \left(\frac{1}{h_1} + \frac{1}{h_2} \right) \rho_2^{(k)} - \frac{1}{h_1^2} \rho_1^{(k)}; & c_1 &= \frac{1}{h_1 h_2} \rho_2^{(k)}; \\ a_N &= \frac{h_{N-1}}{3} + \frac{1}{h_{N-1}^2} (\rho_{N-1}^{(k)} + \rho_N^{(k)}); & g_N &= z'_N - \frac{y_{N-1}^{(k)} - y_N^{(k)}}{h_{N-1}}; \\ b_{N-1} &= \frac{h_{N-1}}{6} - \frac{1}{h_{N-1}} \left(\frac{1}{h_{N-1}} + \frac{1}{h_{N-2}} \right) \rho_{N-1}^{(k)} - \frac{1}{h_{N-1}^2} \rho_N^{(k)}; \\ c_{N-2} &= \frac{\rho_{N-1}^{(k)}}{h_{N-1} h_{N-2}}. \end{aligned}$$

В периодическом случае система состоит из уравнений

$$c_{i-2} M_{i-2} + b_{i-1} M_{i-1} + a_i M_i + b_i M_{i+1} + c_i M_{i+2} = g_i, \quad i = 1, \dots, N-1, \quad (75)$$

где коэффициенты определяются формулами (74). При этом коэффициенты с индексами $N+k-1$ и k считаются равными: $M_1 = M_N$, $h_1 = h_N$, $a_1 = a_N$, Для решения систем (73) и (75) могут быть использованы методы обыкновенной (см. п. 10) и немонотонной прогонки [8].

После вычисления M_i из системы (73) или (75) величины z_i определяются соотношениями (70).

Выделим основные этапы решения исходной задачи. Фиксируем некоторое $r_1 > 0$ и будем минимизировать функционал $I_1(z)$. Итерационный процесс отыскания его минимума строится следующим образом. Выбирается начальное приближение $\rho_i^{(0)}$, вычисляются q_i^0 и решается система (73) (в периодическом случае — (75)). В результате получаем $z^{(1)}$. Затем, вычислив $\rho_i^{(1)}$, находим $z^{(2)}$, и т. д. Процесс прекращается, как только $\max_i |z_i^{(k+1)} - z_i^{(k)}|$ становится

меньше некоторого $\varepsilon > 0$. Тогда $z^{(k+1)}$ принимается за точку минимума функционала $I_1(z)$. После этого выбираем $r_2 > r_1$ и переходим к следующему шагу. Здесь уже за начальное приближение принимается решение, полученное на предыдущем шаге.

Продолжаем увеличивать r_L до тех пор, пока для всех i не станет выполняться

$$|z_i^L - z_i^0| \leq \delta_i + \varepsilon', \quad (76)$$

где $\varepsilon' > 0$ — требуемая точность выполнения ограничений (68). Члены последовательности $\{r_L\}$ будем вычислять по формуле

$$r_L = 2 \max_i |D_i^{L-1}| / 3\varepsilon',$$

что обеспечит в конечном счете выполнение неравенства (76).

Отметим также, что при больших δ_i , когда сглаживающий сплайн лишь в немногих точках выходит на грани параллелепипеда ограничений, сама задача может стать плохо обусловленной. При этом использование описанного алгоритма не приведет к желаемому результату.

Остановимся на вопросе сходимости метода Ньютона при минимизации функционала $I_L(z)$. Хотя описанный выше алгоритм во всех примерах обеспечил сходимость, в общем случае она не гарантируется. Поэтому рекомендуется использовать так называемый *метод Ньютона с регулировкой шага*, при котором кроме описанной выше процедуры получения решения на каждом шаге требуется проверка выполнения условия

$$I_L(z^{(k+1)}) - I_L(z^{(k)}) \leq \theta \alpha \sum_{i=1}^N \frac{\partial}{\partial z_i} I_L(z^{(k)}) (z_i^{(k+1)} - z_i^{(k)}),$$

где $0 < \theta < 1/2$; $\alpha \leq 1$. Если оно выполняется при $\alpha = 1$, то в качестве $k+1$ -го приближения решения берут точку $z^{(k+1)}$, в противном случае берут точку $z^{(k)} + \alpha(z^{(k+1)} - z^{(k)})$, где величина α удовлетворяет этому же условию.

В примере, приведенном в п. 9 данной главы, упомянутые выше величины имеют значения: $\delta_i = \delta = 0.1$; $\varepsilon' = 0.01\delta$; $\varepsilon = 0.1\varepsilon'$. Точность ε' достигается при $r_L = 4200$.

8. В-СПЛАЙНЫ

Сплайны класса C^{n-1} с конечными носителями минимальной длины. Рассмотрим пространство сплайнов $S_{n,1}(\Delta)$: Выделим на $[a, b]$ отрезок $[t_i, t_{i+n+1}]$, ограниченный узлами сетки Δ , и поставим задачу: построить сплайн $\bar{B}_n^i(t)$ такой, что $\bar{B}_n^i(t) \equiv 0$ вне интервала (t_i, t_{i+n+1}) , но на интервале (t, t_{i+n+1}) $\bar{B}_n^i(t) \not\equiv 0$. Такого рода функции называются функциями с *конечным носителем* или *финитными функциями*.

Рассмотрим представление сплайна через усеченные степенные функции (25). В сумме по α в данном случае будет только член с $\alpha = n$. Поскольку $\bar{B}_n^i(t) \equiv 0$ при $t \leq t_i$ и $t \geq t_{i+n+1}$, то $a_k^i = 0$, $k = 0, \dots, n$; $c_r^i = 0$, $r = 2, \dots, i-1, i+n+2, \dots, N-1$, и

$$\bar{B}_n^i(t) = \sum_{r=i}^{i+n+1} c_n^r (t - t_r)_+^n. \quad (77)$$

Так как сплайн $\tilde{B}_n^i(t) = 0$ при $t \geq t_{i+n+1}$, то его производные до порядка n равны нулю в точке t_{i+n+1} . При этом получаем

$$\sum_{r=i}^{i+n+1} c_n^r (t_{i+n+1} - t_r)^l = 0, \quad l = n, \dots, 0, \quad (78)$$

т. е. однородную систему $n + 1$ линейных уравнений с $n + 2$ неизвестными c_n^r . Каждый из определителей $(n + 1)$ -го порядка матрицы системы есть *определитель Вандермонда* и отличен от нуля. Поэтому неизвестные c_n^r можно выразить через одно из них, например c_n^i :

$$c_n^r = \frac{\omega_{n+1, i}^r(t_i)}{\omega_{n+1, i}^r(t_r)} c_n^i, \quad r = i, \dots, i + n + 1,$$

$$\text{где } \omega_{n+1, i}^r(t_p) = \prod_{i=1, j \neq p}^{i+n+1} (t_p - t_j).$$

Из (77) находим

$$\tilde{B}_n^i(t) = a_n^i \sum_{r=i}^{i+n+1} \frac{(t - t_r)_+^n}{\omega_{n+1, i}^r(t_r)}. \quad (79)$$

Здесь обозначено $a_n^i = \omega_{n+1, i}^i(t_i) c_n^i$, причем c_n^i — произвольное число. Таким образом, условия (78) задачи определяют искомым сплайн $\tilde{B}_n^i(t)$ с точностью до постоянного множителя.

Выясним, что будет, если носитель сплайна сузить, например, до интервала (t_i, t_{i+n}) . Условия, что $S_n(t) \equiv 0$ вне (t_i, t_{i+n}) , дают систему уравнений вида (78), в которой слева суммирование по r будет от i до $i + n$. Это система $n + 1$ уравнений с $n + 1$ неизвестными, определитель которой отличен от нуля. Такая система имеет только тривиальное решение $c_n^r = 0, r = i, \dots, i + n$, и, значит, $S_n(t) \equiv 0$. Итак, минимальный интервал, на котором функция $\tilde{B}_n^i(t)$ может отличаться от нуля, а вне его быть равной нулю, содержит $n + 2$ узла сетки. Такие сплайны получили название *B-сплайнов*.

Нормализованные B-сплайны и их свойства. До сих пор мы не фиксировали точно значения коэффициентов a_n^i в формулах B-сплайнов (79). Между тем специальным выбором этих коэффициентов удастся обеспечить полезные свойства B-сплайнов.

Положим $a_n^i = (-1)^{n+1} (t_{i+n+1} - t_i)$. Сплайны с такими коэффициентами обозначим

$$B_n^i(t) = (-1)^{n+1} (t_{i+n+1} - t_i) \sum_{r=i}^{i+n+1} \frac{(t - t_r)_+^n}{\omega_{n+1, i}^r(t_r)}. \quad (80)$$

Эти сплайны называются *нормализованными B-сплайнами*. Перечислим некоторые их свойства.

Свойство 1. *Справедливо тождество*

$$B_n^i(t) = \frac{t - t_i}{t_{i+n} - t_i} B_{n-1}^i(t) + \frac{t_{i+n+1} - t}{t_{i+n+1} - t_{i+1}} B_{n-1}^{i+1}(t). \quad (81)$$

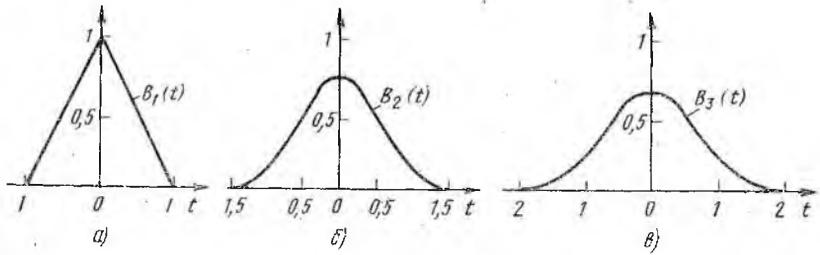


Рис. 7. В-сплайны степеней:
а — первой; б — второй; в — третьей

Его проверка осуществляется непосредственной подстановкой в правую часть тождества выражений $B_{n-1}^i(t)$ и $B_{n-1}^{i+1}(t)$ по формулам типа (80) и приведением подобных членов относительно $(t - t_r)_+^{n+1}$, $r = i, \dots, i + n + 1$.

Тождество (81) представляет собой рекуррентную формулу для вычисления В-сплайнов n -й степени через В-сплайны $(n - 1)$ -й степени. При этом принимается, что $B_0^i(t) \equiv 1$ на $[t_i, t_{i+1})$ и $B_0^i(t) \equiv 0$ вне этого интервала.

Свойство 2. Нормализованные В-сплайны положительны на своих интервалах-носителях.

Справедливость этого утверждения следует из того, что коэффициенты при В-сплайнах в правой части (81) положительны и $B_0^i(t) = 1$ на $[t_i, t_{i+1})$.

Расширим сетку узлов множества сплайнов $S_{n,1}(\Delta)$ точками $t_{-n+1} < \dots < t_0 < t_1$ и $t_N < t_{N+1} < \dots < t_{N+n}$ и построим сплайны $B_n^i(t)$, $i = -n + 1, \dots, N - 1$.

Свойство 3. Справедливо тождество

$$\sum_{i=-n+1}^{N-1} B_n^i(t) \equiv 1, \quad t \in [t_1, t_N]. \quad (82)$$

Оно доказывается путем суммирования обеих частей тождества (81) по индексу i .

Графики нормализованных В-сплайнов для младших степеней $n = 1, 2, 3$ на равномерной сетке узлов Δ : $t_1 + ih$, $i = 1, \dots, N - 1$, приведены на рис. 7.

Функции $B_n^i(t)$, $i = -n + 1, \dots, N - 1$ являются линейными комбинациями усеченных степенных функций. Используя этот факт, можно показать, что они линейно независимы и образуют базис в пространстве сплайнов $S_{n,1}(\Delta)$, т. е. всякий сплайн из этого множества представляется в виде

$$S_n(t) = \sum_{i=-n+1}^{N-1} b_i B_n^i(t),$$

где b_i — постоянные коэффициенты.

Случай многих переменных. Если расширить сетку Δ_s так же, как Δ_t , то для пространства $S_{n,1}(\Delta_s)$ тоже можно построить базис

из нормализованных B -сплайнов: $\bar{B}_n^j(s)$, $j = -n + 1, \dots, M - 1$. Тогда всякий сплайн двух переменных представляется как

$$S_n(t, s) = \sum_{i=-n+1}^{N-1} \sum_{j=-n+1}^{M-1} b_{ij} B_n^i(t) \bar{B}_n^j(s).$$

Попарные произведения $B_n^i(t) \bar{B}_n^j(s)$ образуют базис в пространстве $S_{n,1}(\Delta)$ сплайнов двух переменных.

Аппарат B -сплайнов позволяет легко записать общий вид сплайна от любого числа независимых переменных. Например, в случае трех переменных

$$S_n(t, s, z) = \sum_{i=-n+1}^{N-1} \sum_{j=-n+1}^{M-1} \sum_{k=-n+1}^{L-1} b_{ijk} B_n^i(t) \bar{B}_n^j(s) \hat{B}_n^k(z).$$

В п. 6 мы рассмотрели задачу интерполяции при кусочно-многочленном представлении кубических сплайнов. Здесь познакомимся с техникой применения для этих целей B -сплайнов.

Представление кубических сплайнов через B -сплайны. Множество кубических B -сплайнов $B_3^i(t)$, $i = -2, \dots, N - 1$, определено на расширенной сетке узлов t_{-2}, \dots, t_{N+3} . Дополнительные узлы вводятся произвольным образом, например, $t_{-r+1} = t_1 - rh_1$, $t_{N+r} = t_N + rh_{N-1}$, $r = 1, 2, 3$.

В приложениях кубические B -сплайны удобно нумеровать не по левому узлу интервала-носителя (t_i, t_{i+4}) , а по его среднему узлу t_{i+2} . Обозначим $B_i(t) = B_3^{i-2}(t)$, тогда

$$S_3(t) = \sum_{i=0}^{N+1} b_i B_i(t). \quad (83)$$

Значения сплайна $B_i(t)$ и его двух первых производных в узлах t_{i-2}, t_{i+2} равны нулю, а в узлах t_{i-1}, t_i, t_{i+1} их можно получить с помощью рекуррентной формулы (81). Они сведены в табл. 2. Значения в произвольной точке интервала-носителя проще всего находить по общим формулам кубических сплайнов (45) или (57).

Интерполяция. Если $S_3(t)$ — интерполяционный сплайн, то коэффициенты определяются из условий интерполяции (44). Подставляя в них (83), находим

$$b_{i-1} B_{i-1}(t_i) + b_i B_i(t_i) + b_{i+1} B_{i+1}(t_i) = f_i, \quad i = 1, \dots, N. \quad (84)$$

Сюда добавляются граничные условия, например, типа I:

$$\begin{aligned} b_0 B_0'(t_1) + b_1 B_1'(t_1) + b_2 B_2'(t_1) &= f'_1; \\ b_{N-1} B_{N-1}(t_N) + b_N B_N(t_N) + b_{N+1} B_{N+1}(t_N) &= f'_N. \end{aligned} \quad (85)$$

Уравнения (84), (85) служат для определения коэффициентов b_i .

Разрешимость системы следует из единственности решения задачи интерполяции.

Для получения системы при граничных условиях типа II достаточно в (85) заменить значения первых производных функций $B_i(t)$

ТАБЛИЦА 2. Значения кубического В-сплайна и его производных в узлах сетки Δ

t	t_{i-2}	t_{i-1}	t_i	t_{i+1}	t_{i+2}
$B_i(t)$	0	$\frac{h_{i-2}^2}{(t_i - t_{i-2})(t_{i+1} - t_{i-2})}$	$1 - \frac{1}{t_{i+1} - t_{i-1}} \left(\frac{h_i^2}{t_{i+1} - t_{i-2}} + \frac{h_{i-1}^2}{t_{i+2} - t_{i-1}} \right)$	$\frac{h_{i+1}^2}{(t_{i+2} - t_i)(t_{i+2} - t_{i-1})}$	0
$B_i'(t)$	0	$\frac{3h_{i-2}}{(t_i - t_{i-2})(t_{i+1} - t_{i-2})}$	$\frac{3}{t_{i+1} - t_{i-1}} \left(\frac{h_i}{t_{i+1} - t_{i-2}} - \frac{h_{i-1}}{t_{i+2} - t_{i-1}} \right)$	$-\frac{3h_{i+1}}{(t_{i+2} - t_i)(t_{i+2} - t_{i-1})}$	0
$B_i''(t)$	0	$\frac{6}{(t_i - t_{i-2})(t_{i+1} - t_{i-2})}$	$-\frac{6}{t_{i+1} - t_{i-1}} \left(\frac{1}{t_{i+1} - t_{i-2}} + \frac{1}{t_{i+2} - t_{i-1}} \right)$	$\frac{6}{(t_{i+2} - t_i)(t_{i+2} - t_{i-1})}$	0

и $f(t)$ на значения их вторых производных. В случае условий периодичности имеем $N - 1$ уравнение вида (84), где точки t_1 и t_N отождествляются. Использование представления (83) при граничных условиях типа IV неудобно вследствие того, что граничные уравнения в этом случае содержат по пять неизвестных.

Сравним данный путь решения задачи интерполяции с рассмотренным в п. 6. Технически удобно в случае граничных условий типов I и II предварительно исключить из системы величины b_0 и b_{N+1} . Тогда матрица становится трехдиагональной. Объем вычислений при формировании матриц и векторов правых частей здесь и в (46) совершенно одинаков. Матрица системы (46) имеет диагональное преобладание, а рассматриваемая матрица в общем случае этим свойством не обладает. Установлено, что диагональное преобладание сохраняется при условии

$$\rho = \max_{|i-j|=1} \frac{h_i}{h_j} < \frac{1 + \sqrt{13}}{2} \approx 2,305,$$

которое представляет собой ограничение на отношение соседних шагов сетки. Это ограничение не кажется слишком жестким, но его выполнение позволяет решать задачу методом прогонки, не прибегая к более сложным вычислительным процедурам.

Применение В-сплайнов для вычисления значений сплайнов приводит к увеличению числа операций по сравнению с алгоритмом, указанным в п. 5. Но зато количество хранимой информации здесь меньше, а именно: вместо $3N$ чисел требуется запоминать $2(N + 2)$ числа.

О решении нестандартных задач. Представление сплайнов через В-сплайны позволяет без особого труда выписывать интерполяционные уравнения в нестандартных ситуациях. Например, пусть в точке t^* кроме значения функций $f(t^*)$ задано значение произ-

водной $f'(t^*)$. Требуется построить кубический сплайн, удовлетворяющий условиям

$$S_3(t^*) = f(t^*); S_3'(t^*) = f'(t^*). \quad (86)$$

Возможна следующая схема решения. Точка t^* помещается между двумя узлами сетки Δ , скажем, t_k и t_{k+1} , в которых интерполяционные условия не задаются. Вместо них имеем два соотношения (86), содержащие коэффициенты b_{k-1} , b_k , b_{k+1} , b_{k+2} . Эти соотношения можно преобразовать так, что первое будет содержать неизвестные b_{k-1} , b_k , b_{k+1} , а второе — b_k , b_{k+1} , b_{k+2} .

Подставленные в систему (84) на соответствующие места с номерами k и $k+1$ уравнения (86) не нарушают трехдиагональную структуру матрицы системы. Как и в уравнениях (84), в этих уравнениях в общем случае диагонального преобладания нет. Для равномерной сетки узлов оно имеет место.

Дополнительное условие $S_3'(t^*) = f'(t^*)$ уместно вводить в задачу интерполяции в зоне большого градиента функции. Оно способствует стабилизации решения, предупреждая или хотя бы, ослабляя естественные в этой ситуации осцилляции.

Случай двух переменных. Кубический сплайн двух переменных возьмем в виде

$$S_3(t, s) = \sum_{i=0}^{N+1} \sum_{j=0}^{M+1} b_{ij} B_i(t) \bar{B}_j(s). \quad (87)$$

Введем систему сплайн-функций

$$v_i(s) = \sum_{j=0}^{M+1} b_{ij} \bar{B}_j(s), \quad i = 0, \dots, N+1. \quad (87a)$$

Тогда сплайн $S_3(t, s)$ может быть выражен формулой

$$S_3(t, s) = \sum_{i=0}^{N+1} v_i(s) B_i(t). \quad (88)$$

Рассмотрим задачу интерполяции с условиями (55) и краевыми условиями, например, типа I. Ее решение заключается в нахождении коэффициентов b_{ij} в (87). Дадим описание алгоритма, основанного на представлении сплайна в форме (88).

Исходную информацию задачи представим в виде табл. 1. Значения функций $v_i(s)$ и их производных в узлах сетки Δ обозначим через $v_{ij}^{(q)} = v_i^{(q)}(s_j)$, $q = 0, 1; i = 0, \dots, N+1, j = 0, \dots, M+1$.

Шаг 1. Используя формулу (88), построим интерполяционные кубические сплайны по всем строкам табл. 1, а именно: $S_3(t, s_j)$, $j = 1, \dots, M$, $D^{0,1} S_3(t, s_j)$, $j = 0, M+1$. В результате получим значения v_{ij} , $i = 0, \dots, N+1, j = 1, \dots, M$, и v_{ij} , $i = 0, \dots, N+1; j = 0, M+1$, которые расположим в виде новой таблицы такого же вида.

Шаг 2. Используя формулу (87a), построим интерполяционные кубические сплайны по столбцам второй таблицы: $v_i(s)$, $i = 0, \dots,$

..., $N + 1$. Эта процедура дает значения b_{ij} , $i = 0, \dots, N + 1$, $j = 0, \dots, M + 1$, и задача решена.

Таким образом, по сравнению с алгоритмом, описанным в п. 6, здесь потребовалось решить не три серии одномерных задач, а только две — общим числом $N + M + 4$. (В периодическом случае таких задач $N + M - 2$). Можно применять и другой вариант алгоритма, поменяв ролями переменные t и s .

Сплайн $S_3(t, s)$ и его производные можно вычислять непосредственно по формуле (87). При этом по сравнению с алгоритмами вычисления сплайнов из п. 6 объем вычислений увеличивается, но зато практически в 4 раза сокращается объем хранимой информации. Здесь кроме узлов сеток Δ_t, Δ_s приходится запоминать только $(N + 2)(M + 2)$ коэффициента b_{ij} , а в алгоритме из п. 6—4 NM величин f_{ij} , $m_{ij}^{(1,0)}$, $m_{ij}^{(0,1)}$, $m_{ij}^{(1,1)}$ или f_{ij} , $M_{ij}^{(2,0)}$, $M_{ij}^{(0,2)}$, $M_{ij}^{(2,2)}$.

В целом представление (87) дает существенные преимущества по сравнению с кусочно-многочленным представлением. Эти преимущества становятся решающими при переходе к сплайнам трех и более переменных.

9. ЛОКАЛЬНАЯ АППРОКСИМАЦИЯ КУБИЧЕСКИМИ СПЛАЙНАМИ

Переход от одного представления сплайнов к другому. Это — важный технический прием при практическом использовании сплайнов. Он позволяет для решения прикладных задач применять различные представления, но при проведении числовых расчетов использовать программное обеспечение, ориентированное на какое либо одно представление.

Переход от представления (83) к (45) или (57) не вызывает затруднения. Действительно, из (83) следует при $t = t_i$

$$S_3(t_i) = b_{i-1}B_{i-1}(t_i) + b_iB_i(t_i) + b_{i+1}B_{i+1}(t_i);$$

$$m_i = b_{i-1}B'_{i-1}(t_i) + b_iB'_i(t_i) + b_{i+1}B'_{i+1}(t_i);$$

$$M_i = b_{i-1}B''_{i-1}(t_i) + b_iB''_i(t_i) + b_{i+1}B''_{i+1}(t_i), \quad i = 1, \dots, N.$$

Отсюда после подстановки данных из табл. 2 находим $S_3(t_i)$, m_i , M_i для представления (45) или (57).

Обратный переход чуть сложнее. Из последних равенств определяем b_{i-1} , b_i , b_{i+1} через $S_3(t_i)$, m_i , M_{i+1} :

$$b_{i-1} = S_3(t_i) - \frac{h_{i-2} + 2h_{i-1}}{3} m_i + \frac{h_{i-1}(h_{i-2} + h_{i-1})}{6} M_i; \quad (89a)$$

$$b_i = S_3(t_i) + \frac{h_i - h_{i-1}}{3} m_i - \frac{h_{i-1}h_i}{6} M_i; \quad (89b)$$

$$b_{i+1} = S_3(t_i) + \frac{2h_{i-1} + h_i}{3} m_i + \frac{h_i(h_i + h_{i+1})}{6} M_i, \quad i = 0, \dots, N + 1. \quad (89в)$$

Для вычисления коэффициентов b_i , $i = 2, \dots, N - 1$ можно пользоваться любой из трех формул (89). Коэффициент b_1 может

быть найден по одной из двух формул: (89а) и (89б), а b_N — по формулам (89б) и (89в). Наконец, для определения b_0 имеем одну формулу (89а), а для b_{N+1} — (89в). В дальнейшем будем всегда использовать формулу (89б) и только для b_0 и b_{N+1} — формулы (89а) и (89в) соответственно.

Локальная аппроксимация функций кубическими сплайнами. Задача заключается в определении коэффициентов b_i в зависимости от значений функции $f(t)$ и ее производных в некоторой окрестности точки t_i . При интерполяции значения $S_3(t_i) = f_i$, $i = 1, \dots, N$ задаются, а значения m_i либо M_i находят из решения соответствующей системы уравнений (см. п. 6), после чего коэффициенты b_i определяют по формулам (89). Возникает вопрос: нельзя ли задавать $b_i = \bar{b}_i$, не вычисляя m_i и M_i , так, чтобы сплайн $\hat{S}_3(t) = \sum_{i=0}^{N+1} \bar{b}_i B_i(t)$ был близок к интерполяционному? Оказывается можно.

Самое простое заключается в том, чтобы взять $\bar{b}_i = S_3(t_i) = f_i$. Тогда отклонение сплайна $\hat{S}_3(t)$ от интерполяционного $S_3(t)$ будет

$$\begin{aligned} \|\hat{S}_3(t) - S_3(t)\| &\leq \sum_{i=0}^{N+1} |\bar{b}_i - b_i| B_i(t) \leq \max_i |\bar{b}_i - b_i| \sum_{i=0}^{N+1} B_i(t) = \\ &= \max_i |\bar{b}_i - b_i|. \end{aligned}$$

В силу формул (89) величина $\max_i |\bar{b}_i - b_i| = O(H)$ есть малая первого порядка относительно шага H . Но это не единственный и далеко не лучший путь. Рассмотрим несколько примеров.

Формула, точная на многочленах первой степени. Функция $f(t) \in C^{(2)}[a, b]$. Интерполяционный кубический сплайн $S_3(t)$ приближает такую функцию с точностью $O(H^2)$ (оценка (51)). Поэтому в формулах (89) пренебрежем членами с M_i , т. е. малыми порядка $O(H^2)$, а величины m_i заменим значениями производных $f'_i = f'(t_i)$. Согласно (47) такая замена вносит в (89) погрешность порядка $O(H^2)$.

Сетка узлов Δ задана. Для шагов расширенной сетки $\bar{\Delta}$ положим $h_{-1} = h_0 = h_1$, $h_{N-1} = h_N = h_{N+1}$. Тогда из (89) для коэффициентов сплайна $\hat{S}_3(t)$ имеем

$$\bar{b}_0 = f_1 - h_1 f'_1; \quad \bar{b}_1 = f_1; \quad \bar{b}_i = f_i + (h_i - h_{i-1}) f'_i / 3, \quad i = 2, \dots, N-1, \quad (90)$$

$$\bar{b}_N = f_N; \quad \bar{b}_{N+1} = f_N + h_N f'_N.$$

Сплайн $\hat{S}_3(t)$ будет в этом случае приближать функцию $f(t)$ с той же точностью $O(H^2)$, что и интерполяционный сплайн $S_3(t)$. Абсолютная величина погрешности складывается из погрешности интерполяции (51) и отклонения аппроксимационного сплайна от интерполяционного. В результате

$$\|\hat{S}_3(t) - f(t)\| \leq (51/80) H^2 \|f''(t)\|, \quad t \in [t_2, t_{N-1}].$$

На интервалах $[t_1, t_2]$ и $[t_{N-1}, t_N]$ константа в общем случае несколько больше. Она остается той же, если h_1 и h_{N-1} вдвое меньше H .

Если аппроксимируемая функция $f(t)$ есть многочлен первой степени ($f''(t) = 0$), то сплайн $\hat{S}_3(t)$, как и $S_3(t)$, восстанавливает $f(t)$ точно. Говорят, что формула $\hat{S}_3(t)$ с коэффициентами (90) *точна на многочленах первой степени*.

Пусть на сетке Δ заданы только значения функции f_i , а значения производной f'_i неизвестны. В этом случае значения f'_i в (90) заменяем их конечно-разностными аппроксимациями \tilde{f}'_i (11):

$$\tilde{b}_1 = f_1, \quad \tilde{b}_i = f_i + (h_i - h_{i-1}) \tilde{f}'_i / 3, \quad i = 2, \dots, N-1, \quad \tilde{b}_N = f_N. \quad (91)$$

Коэффициенты \tilde{b}_0 и \tilde{b}_{N+1} проще всего находить из условий интерполяции в узлах t_1 и t_N : $\tilde{S}_3(t_1) = f_1$; $\tilde{S}_3(t_N) = f_N$. Тем более, что в практических задачах эти значения часто бывают заданы точно. Тогда

$$\tilde{b}_0 = f_1 - \frac{3}{2 + \gamma_1} \left(f_2 - f_1 + \frac{\gamma_1 - 1}{3} h_1 \tilde{f}'_2 \right); \quad (91a)$$

$$\tilde{b}_{N+1} = f_N + \frac{3}{2 + \gamma_N} \left(f_N - f_{N-1} + \frac{\gamma_N - 1}{3} h_{N-1} \tilde{f}'_{N-1} \right), \quad (91b)$$

где $\gamma_1 = h_2/h_1$; $\gamma_N = h_{N-2}/h_{N-1}$.

Сплайн $\tilde{S}_3(t) = \sum_{i=0}^{N+1} \tilde{b}_i B_i(t)$ приближает функцию $f(t) \in C^{(2)}[a, b]$ с той же точностью, что и в предыдущем случае.

Формулы, точные на многочленах третьей степени. Квазиинтерполяция. В формулах (89) будем сохранять все члены справа, заменяя m_i и M_i , $i = 1, \dots, N$ значениями f'_i и f''_i или их аппроксимациями. Все такие формулы локальной аппроксимации $\hat{S}_3(t)$ будут наиболее близки к интерполяционным сплайнам и потому называются квазиинтерполяционными.

1. Прежде всего рассмотрим схему с коэффициентами

$$\tilde{b}_0 = f_1 - h_1 f'_1 + (h_1^2/3) f''_1;$$

$$\tilde{b}_i = f_i + \frac{h_i - h_{i-1}}{3} \tilde{f}'_i - \frac{h_{i-1} h_i}{6} \tilde{f}''_i, \quad i = 1, \dots, N;$$

$$\tilde{b}_{N+1} = f_N + h_{N-1} \tilde{f}'_N + (h_{N-1}^2/3) \tilde{f}''_N.$$

Если функция $f(t) \in C^{(4)}[a, b]$, то замена величин m_i и M_i производными в силу (48) и (60) дает погрешность порядка $O(H^4)$. Оценка погрешности аппроксимации есть

$$\|\hat{S}_3(t) - f(t)\| \leq 0,0408 H^4 \|f^{(4)}(t)\|.$$

2. Пусть заданы только значения функции f_i , $i = 1, \dots, N$, а значения f'_i и f''_i неизвестны. Будем в (89) заменять их конечно-разностными аппроксимациями (11б) и (13). Тогда

$$\tilde{b}_i = f_i + \frac{1}{3} \left[h_i \lambda_i \frac{f_i - f_{i-1}}{h_{i-1}} - h_{i-1} \mu_i \frac{f_{i+1} - f_i}{h_i} \right], \quad i = 2, \dots, N-1. \quad (92)$$

Эти формулы дают приближение порядка $O(H^4)$. Такую же точность при $i = 0, 1, N, N + 1$ можно получить двумя путями.

а) Полагаем $\bar{b}_i = \bar{b}_i$, $i = 0, 1, N, N + 1$, если известны f_1^* , f_1^* , f_N^* , f_N^* . Если последние неизвестны, то следует использовать более точные конечно-разностные аппроксимации (16) и (17), построенные по узлам $t_1 - t_4$ и $t_{N-3} - t_N$. Погрешность аппроксимации будет

$$\|\tilde{S}_3(t) - f(t)\| \leq (7/128) H^4 \|f^{IV}(t)\|, \quad t \in [t_3, t_{N-2}].$$

На отрезках $[t_2, t_3]$, $[t_{N-2}, t_{N-1}]$ и $[t_1, t_2]$, $[t_{N-1}, t_N]$ коэффициенты больше: $25/128$ и $77/128$ соответственно.

б) Величины \bar{b}_i , $i = 0, 1, N, N + 1$ вычисляются из условий интерполяции сплайнами $\tilde{S}_3(t)$ значений f_i в точках t_1, t_2, t_{N-1}, t_N :

$$\bar{b}_1 = B_1^{-1}(t_2) [f_2 - \bar{b}_2 B_2(t_2) - \bar{b}_3 B_3(t_2)];$$

$$\bar{b}_0 = 6 [f_1 - \bar{b}_1 B_1(t_1) - \bar{b}_2 B_2(t_1)];$$

$$\bar{b}_N = B_N^{-1}(t_{N-1}) [f_{N-1} - \bar{b}_{N-1} B_{N-1}(t_{N-1}) - \bar{b}_{N-2} B_{N-2}(t_{N-1})];$$

$$\bar{b}_{N+1} = 6 [f_N - \bar{b}_N B_N(t_N) - \bar{b}_{N-1} B_{N-1}(t_N)].$$

Погрешность аппроксимации на отрезке $[t_3, t_{N-2}]$ не увеличивается, а на отрезках $[t_1, t_2]$ и $[t_{N-2}, t_N]$ несколько уменьшается.

Случай двух переменных. Переход от одного представления сплайна к другому совершается простыми вычислениями, если нужно из формы

$$S_3(t, s) = \sum_{i=0}^{N+1} v_i(s) B_i(t); \quad v_i(s) = \sum_{j=0}^{M+1} b_{ij} \bar{B}_j(s) \quad (93)$$

получить форму (56). Действительно, согласно (93)

$$D^{p,q} S_3(t_i, s_j) = \sum_{k=i-1}^{i+1} \sum_{l=j-1}^{j+1} b_{kl} B_k^{(p)}(t_i) \bar{B}_l^{(q)}(s_j), \quad p, q = 0, 1, 2.$$

Подставляя сюда $B_i^{(p)}(t_i)$ из табл. 2 и аналогичные значения $\bar{B}_j^{(q)}(s_j)$, получаем требуемый результат.

Обратный переход совершается в два шага.

Шаг 1. По известным значениям $S_{ij}^{(p,q)} = D^{p,q} S_3(t_i, s_j)$, $p, q = 0, 1, 2$, с помощью формул типа (89) находим значения сплайнов $v_i(s)$ в формулах (93) и их производных $v_i^{(q)}(s_j)$, $q = 0, 1, 2$:

$$v_0^{(q)}(s_j) = S_{0j}^{(0,q)} - \frac{h_{-1} + 2h_0}{3} S_{0j}^{(1,q)} + \frac{h_0(h_{-1} + h_0)}{6} S_{0j}^{(2,q)}, \quad (94a)$$

$$v_i^{(q)}(s_j) = S_{ij}^{(0,q)} + \frac{h_i - h_{i-1}}{3} S_{ij}^{(1,q)} - \frac{h_{i-1} h_i}{6} S_{ij}^{(2,q)}, \quad i = 1, \dots, N; \quad (94б)$$

$$v_{N+1}^{(q)}(s_j) = S_{Nj}^{(0,q)} + \frac{2h_{N-1} + h_N}{3} S_{Nj}^{(1,q)} + \frac{h_N(h_N + h_{N-1})}{6} S_{Nj}^{(2,q)},$$

$$j = 1, \dots, M. \quad (94в)$$

Шаг 2. По полученным значениям $v_i^{(q)}(s_j)$ с помощью аналогов формулы (89б) по переменной s вычисляем b_{ij} в формулах (93):

$$b_{i,0} = v_i(s_0) - \frac{l_{-1} + 2l_0}{3} v_i'(s_0) + \frac{l_0(l_{-1} + l_0)}{6} v_i''(s_0); \quad (95a)$$

$$b_{ij} = v_i(s_j) + \frac{l_j - l_{j-1}}{3} v_i'(s_j) - \frac{l_{j-1}l_j}{6} v_i''(s_j), \quad j = 1, \dots, M; \quad (95б)$$

$$b_{i,M+1} = v_i(s_M) + \frac{2l_{j-1} + l_j}{3} v_i'(s_M) + \frac{l_j(l_j + l_{j+1})}{6} v_i''(s_M),$$

$$i = 0, \dots, N + 1. \quad (95в)$$

Локальная аппроксимация сплайнами двух переменных. Воспользуемся представлением сплайнов в виде (93). Как и в случае интерполяции, локальные процедуры состоят из двух шагов. Укажем их для двумерных аналогов локальных аппроксимаций.

1. Если в узлах сетки Δ заданы значения функции $f(t, s)$ и всех ее производных $D^{p,q}f(t_i, s_j)$, $p, q = 0, 1, 2$, то можно использовать схемы локальной аппроксимации с производными. Для этого расширим сетку Δ до $\bar{\Delta}$, полагая $h_{-1} = h_0 = h_1$; $h_{N+1} = h_N = h_{N-1}$; $l_{-1} = l_0 = l_1$, $l_{M+1} = l_M = l_{M-1}$. На первом шаге по формулам (94) нужно найти значения сплайнов $\vartheta_i(s_j)$ и их производных $\vartheta_i'(s_j)$, $\vartheta_i''(s_j)$, в узлах сетки Δ , используя заданные значения производных $D^{p,q}f(t_i, s_j)$ вместо $S_{ij}^{(p,q)}$. На втором шаге находят коэффициенты \bar{b}_{ij} по формулам (95). Таким образом получают формулы локальной аппроксимации (93) точности $O(H^4 + L^4)$ (квазиинтерполяция).

Если в формулах (94) и (95) пренебречь членами порядка $O(H^2)$ и $O(L^2)$ соответственно, то из первых находятся $\vartheta_i(s_j)$ и $\vartheta_i'(s_j)$, а из вторых, по-прежнему, — коэффициенты \bar{b}_{ij} . В этом случае имеем формулы локальной аппроксимации точности $O(H^2 + L^2)$.

2. Если использовать конечно-разностные аппроксимации частных производных, то по формулам (94) достаточно найти только $\bar{v}_i(s_j)$, заменяя величины $S_{ij}^{(p,0)}$ конечно-разностными аппроксимациями производных $D^{p,0}f(t_i, s_j)$. Коэффициенты \bar{b}_{ij} находят по формулам (95), где производные $v_i^{(q)}(s_j)$ по переменной s заменяют их конечно-разностными аппроксимациями. Здесь также возможны два случая. Первый — это двумерный аналог локальной аппроксимации, когда в формулах (94) и (95) пренебрегают членами порядка $O(H^2)$ и $O(L^2)$ соответственно. Второй случай — двумерная квазиинтерполяция.

В обоих случаях по исходным данным f_{ij} , $i = 1, \dots, N$; $j = 1, \dots, M$, на первом шаге строят кубические сплайны $\bar{S}_3(t, s_j)$ на линиях сетки $s = s_j$, $j = 1, \dots, M$. При этом используют формулы типа (91), но находят не величины \bar{b}_i , а $\bar{v}_i(s_j)$, $i = 0, \dots, N + 1$. По найденным значениям строят кубические сплайны $\bar{v}_i(s)$ с использованием аналогов формул (91) по переменной s . В результате будут найдены \bar{b}_{ij} , $i = 0, \dots, N + 1$, $j = 0, \dots, M + 1$ и получен сплайн $\bar{S}_3(t, s)$.

Квазиинтерполяционный сплайн строится также с использованием формул (92), их аналогов по переменной s и правил вычисления величин $\tilde{v}_i(s_j)$ и \tilde{b}_{ij} в граничных узлах сетки.

Локальное сглаживание. Формулы локальной аппроксимации можно использовать для сглаживания исходных данных, если последние известны с погрешностями.

Рассмотрим формулу $\tilde{S}_3(t)$ в применении к исходным данным f_i , $i = 1, \dots, N$. Значения сплайна в узлах t_i определяются формулами

$$\tilde{S}_3(t_i) = \tilde{b}_{i-1}B_{i-1}(t_i) + \tilde{b}_iB_i(t_i) + \tilde{b}_{i+1}B_{i+1}(t_i), \quad i = 1, \dots, N. \quad (96)$$

Если коэффициенты \tilde{b}_i взять по формулам (91), то каждый из них будет комбинацией трех значений f_{i-1} , f_i , f_{i+1} функции f , и формулы (96) дадут как бы усредненные по пяти точкам t_{i-2} , ..., ..., t_{i+2} значения. В этом смысле их называют *формулами локального сглаживания*.

В случае равноотстоящих узлов формулы (96) упрощаются, так как $\tilde{b}_i = f_i$, $i = 1, \dots, N$, и принимают вид

$$\tilde{S}_3(t_i) = \frac{1}{6}f_{i-1} + \frac{2}{3}f_i + \frac{1}{6}f_{i+1}, \quad i = 2, \dots, N-1. \quad (97)$$

Коэффициенты \tilde{b}_0 и \tilde{b}_{N+1} (91а) и (91б) получены в предположении, что значения f_1 и f_N не варьируются. В случае необходимости это условие можно снять, если в (90) воспользоваться конечно-разностной аппроксимацией производных f'_1 и f'_N и положить в (97) при $i = 1, N$ $f_0 = \tilde{b}_0$ и $f_{N+1} = \tilde{b}_{N+1}$.

Если исходные значения представить как $z_i^0 = f_i + \delta_i$, $i = 1, \dots, \dots, N$, где f_i — точные значения, а δ_i — погрешности, то сглаженные значения (97) можно записать в виде

$$f_i^{(1)} = f_i + \frac{1}{3}h^2f[t_{i-1}, t_i, t_{i+1}] + \delta_i^{(1)},$$

где $\delta_i^{(1)} = \frac{1}{6}\delta_{i-1} + \frac{2}{3}\delta_i + \frac{1}{6}\delta_{i+1}$.

При малых h величина $(1/3)h^2f[t_{i-1}, t_i, t_{i+1}]$ мала, и, значит, погрешность после сглаживания станет равной $\delta_i^{(1)}$. Обозначим $\bar{\delta} = \max_i |\delta_i|$. Тогда, если $\delta_i = \bar{\delta}$ (или $\delta_i = -\bar{\delta}$), то после сглаживания может оказаться, что при $\delta_{i-1} = \delta_{i+1} = \delta_i$ погрешность $\delta_i^{(1)} = \bar{\delta}$. Это худший случай. В самой же благоприятной ситуации ($\delta_{i-1} = \delta_{i+1} = -\bar{\delta}$) погрешность $\delta_i^{(1)} = \bar{\delta}/3$. Таким образом, максимальная погрешность $\bar{\delta}$ при использовании формулы (97) может только уменьшаться, но не аннулируется полностью, поэтому сглаживание следует проводить несколько раз. Однако при большом числе итераций k , когда погрешность уменьшается настолько, что значения $(1/3)h^2f^{(k-1)}[t_{i-1}, t_i, t_{i+1}]$ и $\delta_i^{(k)}$ станут сравнимыми, начнется срезание полезной информации. В пределе процесс будет давать прямую линию. Если сетка неравномерная (см. формулы (96)), то картина усложняется, но качественный характер ее остается прежним.

Так как сказать заранее, сколько следует делать итераций, нельзя, то должен осуществляться контроль за ходом процесса.

1. Если имеются оценки погрешностей, т. е. $|f_i^{(k)} - z_i^0| \leq |\delta_i|$, то средством контроля является проверка этих условий. Если сглаженное значение f_i выходит за пределы «коридора», то его принимают равным значению на границе «коридора». Процесс сглаживания следует останавливать, когда очередной шаг не дает существенных изменений.

2. Если значения погрешности в исходных данных неизвестны, то результат сглаживания можно контролировать с помощью вторых разделенных разностей $f^{(k)} [t_{i-1}, t_i, t_{i+1}]$ от сглаженных на k -м шаге значений. При этом можно добиться либо их ограниченности заданным ε , либо знакопостоянства, характеризующего выпуклость (вогнутость) функции, либо того и другого одновременно. Если требуемое условие не выполняется лишь в отдельных точках, то повторное сглаживание достаточно проводить в этих и соседних с ними точках.

Аналогичным образом используют для сглаживания формулу $S_3(t)$ с коэффициентами (92), точную для многочленов третьей степени. В этом случае при равномерной сетке на $k+1$ -й итерации имеем

$$f_i^{(k+1)} = f_i - \frac{2}{3} h^4 f^{(k)} [t_{i-2}, t_{i-1}, t_i, t_{i+1}, t_{i+2}] + \delta_i^{(k+1)},$$

причем $\delta_i^{(k+1)} = -\frac{1}{36} \delta_{i-2}^{(k)} + \frac{1}{9} \delta_{i-1}^{(k)} + \frac{5}{6} \delta_i^{(k)} + \frac{1}{9} \delta_{i+1}^{(k)} - \frac{1}{36} \delta_{i+2}^{(k)}$.

Здесь величина $(2/3) h^4 f^{(k)} [t_{i-2}, \dots, t_{i+2}]$ на два порядка меньше, чем аналогичная величина в предыдущем случае. Поэтому процедуру сглаживания можно применять при соответственно меньшей погрешности исходных данных, разумеется, если функция достаточно гладкая. Интенсивность сглаживания на каждой итерации тоже будет значительно меньше, чем в предыдущем случае.

В пределе при многократном повторении процедуры получают многочлен третьей степени. Контроль за ходом процесса осуществляют теми же способами, что и выше.

Процесс сглаживания можно организовать и так: сначала вести его по первой схеме, а когда погрешности достаточно уменьшатся, перейти на вторую схему.

Локальное сглаживание сплайнами двух переменных. Снова воспользуемся формулами (93), где коэффициенты b_{hl} определяются последовательно формулами типа (94) и (95). При этом в (94) значения $S_{ij}^{(p,q)}$, $p, q = 0, 1, 2$ заменяются конечно-разностными аппроксимациями производных $D^{p,q} f(t_i, s_j)$ той или иной точности. В любом случае сглаженное значение зависит от исходных данных на 25-точечном прямоугольнике из точек сетки Δ , а именно: f_{hl} , $k = i-2, \dots, i+2$; $l = j-2, \dots, j+2$.

Простейшая схема получается при равномерной сетке, если в (94) и (95) пренебрегать малыми порядками $O(H^2)$ и $O(L^2)$. Тогда

$$\tilde{S}_3(t_i, s_j) = \sum_{k=i-1}^{i+1} \sum_{l=j-1}^{j+1} f_{kl} B_k(t_i) \bar{B}_l(s_j), \quad i = 2, \dots, N-1; \\ j = 2, \dots, M-1. \quad (98)$$

ТАБЛИЦА 3. Числовые результаты сглаживания

t_i	z_i^0	$S(t_i)$	$\tilde{S}(t_i)$
-55,245	-135,257	-135,249	-135,357
-50,634	-123,455	-123,448	-123,555
-46,110	-111,621	-111,619	-111,713
-41,671	-99,758	-99,754	-99,817
-37,319	-87,862	-87,859	-87,878
-33,053	-75,937	-75,928	-75,904
-28,875	-63,983	-63,969	-63,910
-24,783	-52,000	-51,976	-51,900
-20,778	-39,989	-39,961	-39,889
-16,856	-27,952	-27,932	-27,881
-12,991	-15,899	-15,844	-15,825
-9,177	-3,831	-3,731	-3,730
-5,413	8,252	8,352	8,353
-1,668	20,369	20,269	20,268
2,370	32,454	32,354	32,353
6,783	44,406	44,396	44,418
11,567	56,211	56,280	56,296
16,721	67,857	67,957	67,957
22,239	79,331	79,430	79,431
28,118	90,616	90,690	90,716
34,353	101,702	101,737	101,802
40,937	112,573	112,570	112,673
47,865	123,218	123,210	123,318

B -сплайны принимают значения $2/3$ при совпадении индексов k и i (l и j) и $1/6$, если они разные. Это девятиточечная схема. Данные на границе области проще всего сгладить отдельно и затем фиксировать при сглаживании данных во внутренних узлах.

Схемой (98) можно пользоваться, если максимальная погрешность $\delta = \max_{i,j} |\delta_{ij}|$ достаточно велика по сравнению с величинами порядка $O(H^2 + L^2)$. Сглаживание и здесь проводится многократно. В пределе получается поверхность вида $f(t, s) = c_0 + c_1 t + c_2 s + c_3 ts$. Контроль за ходом процесса на k -м шаге осуществляется по ограничениям $|f_{ij}^{(k)} - f_{ij}| \leq \delta_{ij}$, если они заданы, или по поведению вто-

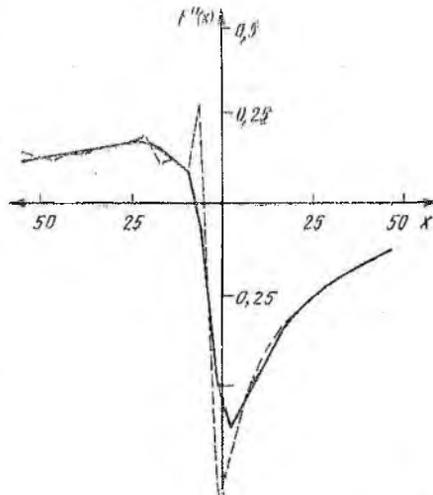


Рис. 8. Графики вторых производных сплайнов: — интерполяционного; - - - сглаживающего

рых разделенных разностей $f_i [t_{i-1}, t_i, t_{i+1}]$ и $f_i [s_{j-1}, s_j, s_{j+1}]$ — в противном случае.

Схема, в которой учитываются более тонкие эффекты, получается, если в (94) и (95) сохранять все члены. Ею можно пользоваться при малой погрешности, превышающей величины $O(H^4 + L^4)$, т. е. на два порядка меньшей, чем в простейшей схеме.

Числовой пример сглаживания. На примере сглаживания сечения профиля лопатки газотурбинного двигателя демонстрируются два способа решения задачи: локальное сглаживание по формуле (96) с коэффициентами (92) и глобальное сглаживание, основанное на экстремальном свойстве сплайнов (см. п. 7). Задана оценка погрешностей исходных данных $|f_i - z_i^0| \leq 0,1, i = 1, \dots, N$. Локальное сглаживание осуществляется по схеме 1 (с. 58). В результате имеем значения $S(t_i)$. Значения сплайна глобального сглаживания обозначены $\bar{S}(t_i)$. Результаты сглаживания сведены в табл. 3. На рис. 8 изображены графики вторых производных интерполяционного и локально сглаживающего сплайнов; график для глобально сглаживающего сплайна мало отличается от последнего и потому не приводится.

10. МЕТОД ПРОГОНКИ

Рассмотрим алгоритмы решения систем уравнений с трех и пятидиагональными матрицами. Такие системы были получены в задачах интерполяции и сглаживания кубическими сплайнами класса C^2 . Алгоритмы даются без обоснования их корректности (возможно выполнение всех необходимых арифметических операций) и устойчивости (не происходит прогрессивного накопления погрешностей округления). За подробностями мы отсылаем читателя к специальным руководствам, например [3, 8].

Система уравнений с трехдиагональной матрицей с диагональным преобладанием. Пусть имеется система

$$\begin{bmatrix} a_1 & b_1 & 0 & \dots & \dots & \dots & 0 & c_1 \\ c_2 & a_2 & b_2 & \dots & \dots & \dots & 0 & 0 \\ \dots & \dots \\ 0 & 0 & 0 & \dots & c_{N-1} & a_{N-1} & b_{N-1} \\ b_N & 0 & 0 & \dots & \dots & 0 & c_N & a_N \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_{N-1} \\ z_N \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ \dots \\ g_{N-1} \\ g_N \end{bmatrix}$$

Алгоритм состоит из двух шагов. Пусть $c_1 = b_N = 0$. Первый шаг (прямая прогонка) — исключение неизвестных z_1, \dots, z_{N-1} из второго и последующих уравнений. В результате чего приходим к системе

$$z_i = v_i z_{i+1} + u_i, \quad i = 1, \dots, N-1; \quad z_N = u_N, \quad (99)$$

где $v_0 = u_0 = 0$;

$$v_i = -\frac{b_i}{a_i + c_i v_{i-1}}; \quad u_i = \frac{g_i - c_i u_{i-1}}{a_i + c_i v_{i-1}}, \quad i = 1, \dots, N. \quad (100)$$

Второй шаг — обратная прогонка. Здесь $z_N = u_N$. Остальные неизвестные находятся по формулам (99).

Для реализации алгоритма требуется $8N$ арифметических операций. На практике приходится решать системы с одной и той же матрицей, но разными правыми частями, например в задачах двумерной интерполяции. В этом случае в формулах (99) следует запоминать величины v_i и $a_i + c_i v_{i-1}$. Тогда для повторного решения системы потребуется только $5N$ арифметических операций.

Пусть теперь c_1 и b_N отличны от нуля. В результате исключения неизвестных (прямая прогонка) получаем

$$z_i = v_i z_{i+1} + u_i + p_i z_N, \quad i = 1, \dots, N-1, \quad (101)$$

где u_i, v_i определяются формулами (100) и

$$p_0 = 1, p_i = -\frac{c_i p_{i-1}}{a_i + c_i v_{i-1}}, \quad i = 1, \dots, N-1.$$

Чтобы найти z_N , выразим все z_i через z_N :

$$z_i = t_i + s_i z_N, \quad i = 1, \dots, N-1. \quad (102)$$

Подставляя эти выражения в (101), приходим к формулам

$$s_N = 1, t_N = 0; \\ s_i = v_i s_{i+1} + p_i, t_i = v_i t_{i+1} + u_i, \quad i = N-1, \dots, 1.$$

Подставляя z_1 и z_{N-1} в последнее уравнение системы, получаем

$$z_N = \frac{u_N + v_N t_1}{1 - p_N - v_N s_1},$$

после чего решение находится по формулам (102).

Всего необходимо выполнить $14N$ арифметических операций. При решении систем с одинаковыми матрицами следует запомнить величины $v_i, a_i + c_i v_{i-1}, s_i, 1 - p_N - v_N s_1$. Тогда для решения систем понадобится $7N$ арифметических операций.

Система уравнений с пятидиагональной матрицей с диагональным преобладанием. Рассмотрим систему $Az = g$ с матрицей

$$A = \begin{bmatrix} a_1 & b_1 & c_1 & 0 & 0 & \dots & \dots & \dots & 0 & e_1 & d_1 \\ d_2 & a_2 & b_2 & c_2 & 0 & \dots & \dots & \dots & 0 & 0 & e_2 \\ e_3 & d_3 & a_3 & b_3 & c_3 & \dots & \dots & \dots & 0 & 0 & 0 \\ \dots & \dots \\ 0 & 0 & 0 & \dots & e_{N-2} & d_{N-2} & a_{N-2} & b_{N-2} & c_{N-2} & & \\ c_{N-1} & 0 & 0 & \dots & 0 & e_{N-1} & d_{N-1} & a_{N-1} & b_{N-1} & & \\ b_N & c_N & 0 & \dots & 0 & 0 & e_N & d_N & a_N & & \end{bmatrix}.$$

Для решения можно использовать метод прогонки. Алгоритм строится аналогично случаю системы с трехдиагональной матрицей. Приведем расчетные формулы.

Пусть вначале $d_1 = e_1 = e_2 = c_{N-1} = b_N = c_N = 0$. Решение системы отыскивается в виде

$$z_i = v_i z_{i+1} + w_i z_{i+2} + u_i, \quad i = 1, \dots, N-1; \quad z_N = u_N. \quad (103)$$

Если обозначить $D_i = d_i + e_i v_{i+1}$; $A_i = a_i + D_i v_{i-1} + e_i w_{i-2}$, то прогоночные коэффициенты будут

$$v_{-1} = v_0 = w_{-1} = w_0 = u_{-1} = u_0 = 0; \\ u_i = \frac{g_i - D_i u_{i-1} - e_i w_{i-2}}{A_i}, \quad i = 1, \dots, N; \quad (104) \\ v_i = -\frac{b_i + D_i w_{i-1}}{A_i}, \quad w_i = -\frac{c_i}{A_i}.$$

Для реализации алгоритма в этом случае требуется $19N$ арифметических операций. При решении последовательности систем с одинаковыми матрицами следует запоминать v_i, w_i, D_i и A_i . Тогда для повторного решения системы понадобится только $9N$ арифметических операций.

Пусть теперь $d_1, e_1, e_2, b_N, c_{N-1}, c_N$ отличны от нуля. В результате прямой прогонки получаем

$$z_i = v_i z_{i+1} + w_i z_{i+2} + u_i + p_i z_N + g_i z_{N-1}, \quad i = 1, \dots, N-2, \quad (105)$$

где u_i, v_i, w_i определяются формулами (104) и

$$p_{-1} = q_0 = 0; \quad p_0 = q_{-1} = 1;$$

$$p_i = -\frac{D_i p_{i-1} + e_i p_{i-2}}{A_i}; \quad q_i = -\frac{D_i q_{i-1} + e_i q_{i-2}}{A_i}, \quad i = 1, \dots, N.$$

Обратная прогонка приводит к формулам

$$z_i = t_i + s_i z_N + r_i z_{N-1}, \quad i = 1, \dots, N-2, \quad (106)$$

где

$$t_{N-1} = s_{N-1} = t_N = r_N = 0; \quad r_{N-1} = s_N = 1;$$

$$t_i = v_i t_{i+1} + w_i t_{i+2} + u_i; \quad s_i = v_i s_{i+1} + w_i s_{i+2} + p_i;$$

$$r_i = v_i r_{i+1} + w_i r_{i+2} + q_i, \quad i = N-2, \dots, 1.$$

Значения z_{N-1} и z_N находят по формулам

$$z_{N-1} = \frac{\beta_1 \alpha_{22} - \beta_2 \alpha_{12}}{\alpha_{11} \alpha_{22} - \alpha_{12} \alpha_{21}}; \quad z_N = \frac{-\beta_1 \alpha_{21} + \beta_2 \alpha_{11}}{\alpha_{11} \alpha_{22} - \alpha_{12} \alpha_{21}},$$

где

$$\alpha_{11} = 1 - q_{N-1} - w_{N-1} r_1; \quad \alpha_{12} = -(p_{N-1} + v_{N-1} + w_{N-1} s_1);$$

$$\alpha_{21} = -(v_N r_1 + w_N r_2 + q_N); \quad \alpha_{22} = 1 - p_N - v_N s_1 - w_N s_2;$$

$$\beta_1 = w_{N-1} t_1 + u_{N-1}; \quad \beta_2 = v_N t_1 + w_N t_2 + u_N.$$

Остальные компоненты решения определяются формулами (106).

Для реализации алгоритма требуется выполнить $39N$ арифметических операций. При решении систем с одинаковыми матрицами следует запоминать $v_i, w_i, D_i, A_i, s_i, r_i, \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}$, и тогда потребуется выполнить $13N$ арифметических операций.

Системы с ленточными матрицами без диагонального преобладания. Диагональное преобладание является достаточным условием применимости метода прогонки, но не необходимым. Поэтому его нарушение обычно не приводит к затруднениям, но всегда гарантировать это нельзя. Для таких случаев разработан метод так называемой *нмонотонной прогонки*.

Метод прогонки представляет собой алгоритм исключения Гаусса [3], приспособленный для решения систем с ленточными матрицами. При этом диагональное преобладание обеспечивает автоматический выбор главного (наибольшего по модулю в строке матрицы) элемента. В алгоритмах нмонотонной прогонки такой элемент предварительно отыскивается. Их изложение для систем с трех- и пяти-диагональными матрицами можно найти в работе [8].

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ КРИВЫХ И ПОВЕРХНОСТЕЙ

В настоящей главе очерчен класс кривых и поверхностей, имеющих применение в задачах инженерной геометрии. Основной темой, данной и следующей глав, является методика аппроксимации этих геометрических объектов. В качестве инструмента аппроксимации выступают сплайны. Изложение основано на результатах теории приближения полиномиальными сплайнами, данной в гл. I.

1. КРИВЫЕ И ПОВЕРХНОСТИ В ИНЖЕНЕРНО-ГЕОМЕТРИЧЕСКИХ ЗАДАЧАХ

Специфика инженерного подхода к геометрии. Понятия кривой и поверхности являются математическими абстракциями, однако далеко не все множество кривых и поверхностей, интересующее математиков, нужно инженерам.

Инженерная геометрия оперирует только с такими объектами, которые размещаются в ограниченной области на *евклидовой плоскости* или *в пространстве*. Когда речь идет, например, о кривой с бесконечными ветвями (парабола, гипербола и др.), фактически подразумевается некоторый ее кусок в конечной области. Далее, в инженерной геометрии рассматриваются только кривые, имеющие конечную длину. Такие кривые называются *спрямляемыми*. Ясно, что кривая, не укладывающаяся ни в какую ограниченную область, не имеет конечной длины. Однако это условие является лишь необходимым, но не достаточным.

Примером неспрямляемой кривой служит кривая, задаваемая в полярных координатах уравнением $r = (\varphi + 5)(\varphi + 1)^{-1}$, где $\varphi \geq 0$. Эта кривая, хотя и расположена в круге $r = 5$, не имеет конечной длины, так как представляет собой спираль, навивающуюся на окружность $r = 1$ (рис. 9). Эта спираль задана на полупрямой $[0, \infty)$ параметра φ . Заменив параметр φ на t , можно задать спираль двумя функциями: $\varphi = t(1 - t)^{-1}$, $r = 5 - 4t$, где $t \in [0, 1)$. Любой кусок спирали, отвечающий отрезку $[a, b] \subset (0, 1)$, имеет конечную длину.

Однако существуют кривые, заданные на замкнутом отрезке, но не являющиеся спрямляемыми. Примером такой кривой на

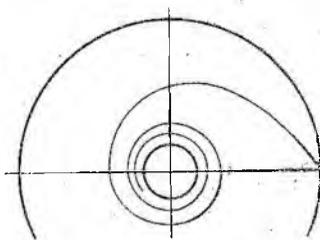


Рис. 9. Спираль, навивающаяся на окружность

плоскости может служить график функции $y(x) = x \sin x^{-1}$ на отрезке $[-1, 1]$. Здесь в качестве параметра выступает декартова координата x . Эта кривая непрерывна, но не спрямляема [25]. Более того, ни один ее кусок, содержащий начало координат, не является спрямляемым. Это обусловлено тем, что в окрестности точки $(x = 0, y = 0)$ кривая «неограниченно колеблется», бесконечное множество раз пересекая ось x . В пространстве аналогичная кривая задается уравнениями

$$x = x; \quad y = x \sin(1/x); \quad z = x \cos(1/x).$$

Однако не всегда такое поведение кривой является признаком отсутствия у нее конечной длины. Так, график функции $y(x) = x^2 \sin x^{-1}$ на отрезке $[-1, 1]$ есть спрямляемая кривая. Достаточным условием спрямляемости является непрерывная дифференцируемость координат текущей точки кривой как функций параметра, меняющегося на замкнутом отрезке.

В инженерной геометрии не используются «слишком колеблющиеся» кривые, хотя бы и спрямляемые. Чтобы исключить кривые, имеющие бесконечное множество волн убывающей амплитуды (или витков убывающего радиуса в пространстве), примем ограничение, основанное на понятии выпуклой дуги. Плоская дуга называется *выпуклой*, если вместе со стягивающей ее хордой она ограничивает выпуклую область, т. е. область, любые две точки которой можно соединить прямолинейным отрезком, лежащим в ней.

Всякая рассматриваемая в инженерной геометрии кривая обладает следующим свойством. *Проекция кривой на произвольную плоскость может быть составлена из конечного числа выпуклых дуг.* Здесь и ниже, когда условия проецирования не оговариваются, подразумевается ортогональная проекция параллельными прямыми. *Любые две кривые с таким свойством могут пересекаться между собой только по конечному числу отрезков (в частности, точек).* Это свойство имеют, например, кривые, задаваемые полиномиальными уравнениями (*алгебраические кривые*).

Параметризация кривых. Говорят, что данная кривая (поверхность) удовлетворяет некоторому условию кусочно, если эту кривую (поверхность) можно разбить на конечное число кусков, для каждого из которых условие выполнено. В инженерной геометрии рассматриваются кривые, удовлетворяющие следующему условию: *множество точек кривой является образом непрерывного кусочно-обратимого (взаимно однозначного) отображения замкнутого отрезка числовой прямой.* Отображаемый отрезок называют *областью изменения параметра*, а само отображение — *параметризацией*. Примеры непрерывных параметризаций были приведены выше.

Одно и то же *точечное множество* может быть множеством точек нескольких различных кривых. Например, петлеобразное множество на рис. 10, *а* обходится с самопересечением, на рис. 10, *б* — без самопересечения (обход показан стрелками). Таким образом задаются две различные кривые. В первой кривой точка *а* соседствует с точкой *б* и точка *с* — с точкой *д*, тогда как во второй кривой сосед-

ними являются точки a c s и b c d .

Точка e делит рассматриваемое точечное множество на две такие части, что отображение, задающее соответствующий кусок кривой, имеет обратное отображение, которое каждой точке куска ставит в соответствие одну точку из области параметра. Если отображение, задающее всю кривую, обратимо, то кривая не имеет самопересечений (самокасаний) и называется *простой дугой*. Понятие кривой синтезирует понятия точечного множества и отношения соседства на нем. Соседними считаются точки, чьи прообразы соседствуют в области параметра.

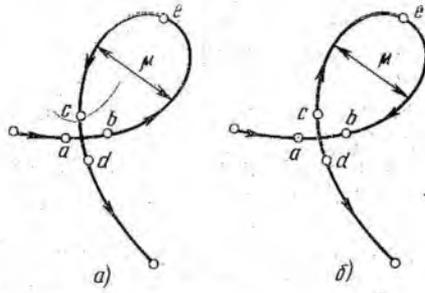


Рис. 10. Различные обходы петлеобразного множества

Одна и та же кривая может быть параметризована бесчисленным множеством способов. Параметризация осуществляется заданием декартовых координат «текущей» точки кривой как функций некоторого параметра: $x = x(t)$, $y = y(t)$, $z = z(t)$, $a \leq t \leq b$. Конкретный вид этих функций отличает данную параметризацию от всех прочих. Известно, что если $t = t(u)$ есть произвольная непрерывная и строго монотонная функция, то

$$x = x(t(u)) = \bar{x}(u); \quad y = y(t(u)) = \bar{y}(u); \quad z = z(t(u)) = \bar{z}(u) \quad (107)$$

есть другая параметризация той же кривой. И наоборот, если

$$x = x(t); \quad y = y(t); \quad z = z(t) \quad \text{и} \quad x = \bar{x}(u); \quad y = \bar{y}(u); \quad z = \bar{z}(u)$$

есть две различные параметризации одной кривой, то существует единственная непрерывная и строго монотонная функция $t = t(u)$, переводящая одну параметризацию в другую по формулам (107). Заметим, что область значений такой функции — тоже отрезок. Подобные функции реализуют непрерывное и взаимно однозначное отображение отрезка на отрезок. Ясно, что обратная функция $u = u(t)$ переводит вторую параметризацию в первую.

Мы будем пользоваться *параметрическим уравнением кривой* в векторной форме:

$$\mathbf{r} = \mathbf{r}(t) = x(t) \mathbf{e}_1 + y(t) \mathbf{e}_2 + z(t) \mathbf{e}_3,$$

где $\mathbf{r} = \{x, y, z\}$ — радиус-вектор текущей точки на кривой; $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ — орты координатных осей.

Кривая называется *m раз непрерывно дифференцируемой*, если она допускает m раз непрерывно дифференцируемую параметризацию. Непрерывно дифференцируемые кривые ($m = 1$), допускающие параметризацию, первая производная которой всюду отлична от

нуля ($|\mathbf{r}'| \neq 0$), называются *гладкими*. Для них в каждой точке кривой определен касательный вектор. Если при некотором t выполняется равенство $|\mathbf{r}'(t)| = 0$, то точка может быть особой. Кусочно-гладкие кривые в инженерной геометрии обычно рассматриваются как составные, т. е. для каждого гладкого куска вводится своя параметризация.

Рассматриваемые нами кривые будем предполагать гладкими и обладающими кусочно непрерывной ограниченной кривизной, т. е. допускающими такую параметризацию $\mathbf{r} = \mathbf{r}(t)$, что вторая производная $\mathbf{r}''(t)$ кусочно-непрерывна и ограничена.

Сформулированные выше условия выделяют из всех кривых сравнительно узкий класс Q , который включает кривые, используемые в инженерной геометрии. Несмотря на довольно жесткие математические ограничения, этот класс оказывается вполне достаточным для практических целей.

Дискретно-точечное задание кривой. Следуя инженерной традиции, этим термином будем называть конечную совокупность $\{\mathbf{r}_i, i = 1, 2, \dots, N\}$ точек кривой, упорядоченных по ее длине. В каждой или в некоторых из них могут быть заданы дополнительные сведения о свойствах кривой в окрестности соответствующей точки. Такой способ представления сложных кривых в инженерной практике является обычным. При его использовании предполагается наличие некоторого инструмента для интерполирования. Физической реализацией такого инструмента является, например, упругая рейка. В качестве математического «инструмента» используются различные функции, чаще всего полиномы Лагранжа, Ньютона, а в последние годы — сплайны. Известно, что для конкретного дискретно-точечного задания кривой и любой возрастающей числовой последовательности $t_1 < t_2 < \dots < t_N$ можно найти (и притом не единственным образом) такую параметризацию $\mathbf{r} = \mathbf{r}(t)$, $t \in [t_1, t_N]$, этой кривой, что $\mathbf{r}(t_i) = \mathbf{r}_i$ для всех i . Следовательно, задача об интерполировании кривой по ее дискретно-точечному заданию может быть сведена к интерполированию какой-либо ее параметризации, т. е. некоторой вектор-функции.

Параметризация поверхности. Под словом «поверхность» здесь будет пониматься расположенная в трехмерном евклидовом пространстве двусторонняя поверхность с краем, которую можно задать непрерывным отображением замкнутого прямоугольника $\Omega: \{a \leq t \leq b, c \leq s \leq d\}$, лежащего в плоскости вещественных переменных t, s . Отображение (параметризация) определяется заданием координат точек поверхности как функций $x = x(t, s)$; $y = y(t, s)$; $z = z(t, s)$, которые называются уравнениями поверхности в параметрической форме. Популярным примером односторонней поверхности с краем является лента Мебиуса. Модель этой поверхности можно получить за счет взаимного наложения двух противоположных сторон изогнутого в пространстве прямоугольника, т. е. двусторонней поверхности. Именно так моделировалась в ЭВМ лента Мебиуса, изображение которой было выведено затем на экран устройства визуализации (см. рис. 73).

В дальнейшем будем пользоваться параметрическим уравнением поверхности в векторной форме:

$$\mathbf{r} = \mathbf{r}(t, s) = x(t, s)\mathbf{e}_1 + y(t, s)\mathbf{e}_2 + z(t, s)\mathbf{e}_3,$$

где t и s принадлежат прямоугольнику Ω .

Поверхность называется m раз непрерывно дифференцируемой, если она допускает m раз непрерывно дифференцируемую параметризацию. Непрерывно дифференцируемые поверхности ($m = 1$), допускающие параметризацию, для которой ранг матрицы

$$C = \begin{bmatrix} x_t & y_t & z_t \\ x_s & y_s & z_s \end{bmatrix}$$

равен двум всюду в прямоугольнике Ω , называются *гладкими*. Геометрически это означает, что в каждой точке каждой координатной линии определен касательный вектор и в любой точке касательные векторы \mathbf{r}_t и \mathbf{r}_s пересекающихся координатных линий неколлинеарны. Таким образом, в каждой точке поверхности определен направляющий вектор нормали $\mathbf{N} = \mathbf{r}_t \times \mathbf{r}_s$. Если ранг матрицы C в точке (t, s) меньше двух, то точка может быть особой. Если это имеет место на линии в области Ω , то на поверхности этой линии может соответствовать ребро.

Если $\mathbf{r} = \mathbf{r}(u, v)$ — некоторая параметризация поверхности и $u = u(t, s)$, $v = v(t, s)$ — произвольные непрерывно дифференцируемые функции такие, что отличен от нуля определитель

$$\begin{vmatrix} u_t & u_s \\ v_t & v_s \end{vmatrix} \neq 0,$$

то $\mathbf{r} = \mathbf{r}(u(t, s), v(t, s))$ — параметризация той же поверхности.

Рассматриваемые поверхности подчинены еще одному требованию. Участки плоского сечения поверхности, где угол ϕ между нормалью \mathbf{N} и секущей плоскостью удовлетворяет неравенству $\phi \leq \theta < \pi/2$, являются кривыми класса Q . Это требование всегда выполняется на практике.

Каркас на поверхности. Аналогом дискретно-точечного задания кривой в случае поверхности является каркасное задание. В прямоугольнике Ω выберем прямоугольную сетку

$$\Delta: a = t_1 < t_2 < \dots < t_N = b; \quad c = s_1 < s_2 < \dots < s_M = d.$$

Совокупность координатных линий на поверхности, прообразами которых являются линии этой сетки, будем называть *каркасом поверхности*. Край поверхности составлен из кусков каркасных линий. Куски поверхности, на которые ее разбивает каркас, будем называть клетками, края клеток — ячейками каркаса, точки взаимопересечений каркасных линий — узлами каркаса. Открытая клетка — это клетка без окаймляющей ее ячейки. Мы будем считать, что края поверхностей, каркасные кривые и линии стыка дважды непрерывно дифференцируемых кусков поверхностей являются кривыми класса Q .

Сформулированные условия определяют класс T поверхностей, применяемых в инженерной геометрии. Эти поверхности являются *квадрируемыми*, т. е. имеют конечную площадь. В силу того, что ячейка образована кривыми класса Q , квадрируемыми являются также и отдельно взятые клетки, причем площадь поверхности складывается из площадей всех клеток, или, как говорят, обладает свойством аддитивности. Аддитивными оказываются интегралы по поверхности от некоторых функций, и этим пользуются при практических вычислениях (гл. IV, п. 1).

2. ПОНЯТИЕ БЛИЗОСТИ ГЕОМЕТРИЧЕСКИХ ОБЪЕКТОВ

Метризация пространства. Инженерная геометрия рассматривает кривые и поверхности, погруженные в евклидово пространство, которое является метрическим. Метризация пространства осуществляется заданием расстояния $\rho(a, b)$ как функции произвольной пары точек. Евклидово расстояние между точками выражается через декартовы координаты известной формулой

$$\rho(a, b) = \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2 + (z_b - z_a)^2}$$

и отражает наше естественное представление о близости точек.

Отсюда вытекает понятие *диаметра точечного множества*, характеризующее линейные размеры этого множества, его «габариты». Для ограниченного замкнутого множества диаметр есть максимальное расстояние между его точками — полная аналогия с диаметром шара. Наоборот, минимальное расстояние между двумя точками, из которых одна принадлежит некоторому множеству, а другая — другому, есть *кратчайшее расстояние между точками множеств*. Для пересекающихся множеств это кратчайшее расстояние равно нулю.

Внутренняя метрика. В понятия «близость точек», «соседство точек» может быть вложен различный смысл в зависимости от содержащего эти точки геометрического объекта. Так, любые две точки a и b поверхности соединяются лежащей на ней спрямляемой кривой длины σ_{ab} . Таких кривых (с различными и одинаковыми длинами) бесконечно много. Обозначим $\{\sigma_{ab}\}$ множество их длин, а $\inf \sigma_{ab}$ — нижнюю грань этого множества. Нижняя грань — *infimum* — числового множества есть наибольшее из всех чисел, которые не превышают каждое из чисел этого множества. Величина $\inf \sigma_{ab}$ есть внутреннее расстояние между точками a и b на поверхности. Для куска поверхности имеет смысл понятие, аналогичное понятию диаметра. А именно, для клетки \mathcal{K} , рассматриваемой вместе с ее краем, *внутренний диаметр* есть

$$d_{\mathcal{K}} = \max_{a, b \in \mathcal{K}} \{\inf \sigma_{ab}\} = \max_{a, b \in \mathcal{K}} \{\min \sigma_{ab}\}.$$

Здесь нижнюю грань можно заменить на минимум, поскольку в клетке целиком содержится кратчайшая кривая, связывающая любые две точки клетки.

На клетке с краем обязательно найдутся по крайней мере две точки (пусть это будут a и b), минимальное внутреннее расстояние между которыми равно диаметру клетки. Поскольку длина прямолинейного отрезка между ними (евклидово расстояние) не больше длины любой кривой, их соединяющей, замкнутую область на поверхности, имеющую внутренний диаметр d , можно поместить целиком внутрь сферы, описанной возле правильного тетраэдра с ребром длины d . Диаметр самой этой сферы равен $\sqrt{3/2}d$.

Рассмотрим область на поверхности, составленную из двух смежных замкнутых клеток с внутренними диаметрами d_1 и d_2 . Общая часть этих клеток (обозначим ее γ) есть либо отрезок каркасной линии, либо узел. Установим соотношение между внутренним диаметром $d_{1,2}$ этой области и внутренними диаметрами d_1 и d_2 составляющих ее клеток. Если точки лежат в одной клетке, например, первой, то найдется дуга ab такая, что $\sigma_{ab} \leq d_1$. Если они лежат в разных клетках, а c — одна из точек пересечения кривой ab с множеством γ , то $\sigma_{ac} \leq d_1$, $\sigma_{cb} \leq d_2$, $\sigma_{ab} \leq d_1 + d_2$. Значит, $\sigma_{ab} = d_{1,2} \leq d_1 + d_2$.

Если цепь из n попарно смежных клеток с диаметрами d_1, d_2, \dots, d_n имеет диаметр D , то, последовательно применяя это неравенство, получаем $D \leq d_1 + d_2 + \dots + d_n$. Если на поверхности с каркасом диаметры всех клеток ограничены сверху значением d , а любые две клетки можно соединить клеточной цепью не более чем из n клеток, то внутренний диаметр всей поверхности оценивается неравенством $D \leq nd$.

Такие предварительные оценки диаметра позволяют сократить время решения задач на взаимное расположение геометрических объектов. Например, при отыскании пересечения прямой линии с поверхностью одно найденное расстояние ρ от некоторого узла каркаса до прямой может оказаться достаточным, чтобы сделать заключение об отсутствии пересечения. Действительно, если $(n + 1)d \leq \rho$, то никакой узел каркаса не может быть расположен ближе чем на d по отношению к данной прямой, т. е. прямая не пересекает ни одну клетку на поверхности.

По отношению к ячейке каркаса также применимы понятия диаметра и внутреннего диаметра. Ясно, что диаметр ячейки не больше ее внутреннего диаметра, диаметра клетки и внутреннего диаметра клетки. Диаметр и внутренний диаметр ячейки могут быть значительно меньше диаметра и внутреннего диаметра клетки, например в случае маленького выреза в большой сфере. Однако имеет место обратная зависимость: если каркас на поверхности крупноячеистый, то диаметры клеток заведомо велики.

Далее нам потребуется обобщение понятия расстояния между точками на более сложные геометрические объекты. Если в евклидовом пространстве расстояние между точками a и b есть $\rho(a, b)$, то расстояние между точкой a и некоторым множеством A из этого метрического пространства есть $\text{dist}(a, A) = \inf_{x \in A} \rho(a, x)$, где ниж-

няя грань берется по всем точкам x , принадлежащим A . Шаровая окрестность радиуса r (или r -окрестность) точки a представляет собой множество точек x таких, что $\rho(x, a) < r$. По аналогии, r -окрестность множества A , обозначаемая $v_r(A)$, определяется как множество точек x таких, что $\text{dist}(x, A) < r$.

Метрика Хаусдорфа. Для множества замкнутых подмножеств ограниченной области метрического пространства Хаусдорфом¹ была введена специальная метрика. Хаусдорфово расстояние $\chi(A, B)$ между двумя ограниченными замкнутыми подмножествами A и B из E^n есть нижняя грань множества чисел r таких, что $v_r(A)$ содержит B и $v_r(B)$ содержит A , т. е.

$$\chi(A, B) = \inf \{ r: v_r(A) \supset B, v_r(B) \supset A \}. \quad (108)$$

Это расстояние удовлетворяет аксиомам метрики. Его не следует смешивать с понятием кратчайшего расстояния между точками двух множеств, которое свойствами метрики не обладает и, к тому же, характеризует только «локальную удаленность». Хаусдорфово расстояние в значительной степени является глобальной характеристикой.

Найдем другое выражение для хаусдорфова расстояния между множествами A и B , эквивалентное выражению (108). Рассмотрим величины

$$\max_{x \in A} \text{dist}(x, B) \text{ и } \max_{x \in B} \text{dist}(x, A),$$

максимальную из них обозначим λ . По смыслу величины λ всякая r -окрестность множества B при $r > \lambda$ содержит в себе множество A , т. е. $v_r(B) \supset A$. Точно так же и наоборот: $v_r(A) \supset B$ при $r > \lambda$. Всякая r -окрестность множества B при $r < \lambda$ не содержит всего A ; например, точка $a \in A$, на которой достигается $\text{dist}(a, B) = \lambda$, не входит в $v_r(B)$ при $(r < \lambda)$. Либо всякая r -окрестность множества A при $r < \lambda$ не содержит всего B , если на некоторой точке $b \in B$ достигается $\text{dist}(b, A) = \lambda$. Следовательно, λ есть нижняя грань множества значений r таких, что $v_r(B)$ содержит A и $v_r(A)$ содержит B . Таким образом, λ это и есть хаусдорфово расстояние, как оно определено в (108) и можно записать

$$\chi = \lambda = \max \left\{ \max_{x \in A} \text{dist}(x, B), \max_{x \in B} \text{dist}(x, A) \right\}. \quad (109)$$

Теперь рассмотрим несколько примеров, когда χ легко вычисляется. Для двух кругов на плоскости хаусдорфово расстояние есть сумма $l + R$ (рис. 11, а). Для двух прямолинейных параллельных отрезков (рис. 11, б) хаусдорфово расстояние равно большему из расстояний между левыми или правыми концами. Для двух прямолинейных отрезков, расположенных на плоскости Т-образно, хаус-

¹ Хаусдорф Феликс (1868—1942) — немецкий математик, работал в таких математических областях, как теория множеств, топология, функциональный анализ, уточнил понятие близости или соседства точек (1927), дал современное определение топологического пространства.

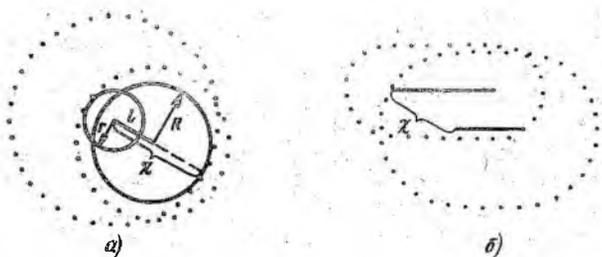


Рис. 11. «Близость» для кругов и параллельных отрезков

дорфово расстояние $\chi = (l + d)$ (рис. 12, а) либо $\chi = \sqrt{d^2 + (L/2)^2}$ (рис. 12, б). Точками на рисунках отмечены соответствующие окрестности v , сравниваемых точечных множеств.

Мера близости по Фреше¹. Расстояние между двумя кривыми f и g можно определить следующим образом. Будем обозначать всякую параметризацию кривой f символом $f(t)$, а множество всех параметризаций этой кривой — $\{f\}$. Для параметризаций $f(t)$ и $g(t)$ указанных кривых сравниваются изопараметричные точки, т. е. точки, соответствующие одному и тому же значению параметра t . Максимальное расстояние между изопараметричными точками есть $\max_t \rho(f(t), g(t))$. При варьировании параметризаций меняются изопараметрические соответствия между кривыми и получается бесконечное множество значений максимумов расстояний в изопараметрических парах. Нижняя грань множества, полученного по всевозможным параметризациям, и принимается за меру близости кривых:

$$\mu(f, g) = \mu(g, f) = \inf_{\{f, g\}} \left\{ \max_t \rho(f(t), g(t)) \right\}. \quad (110)$$

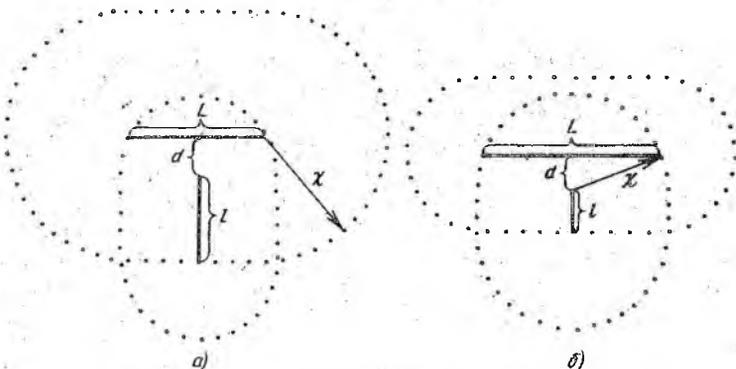


Рис. 12. «Близость» для T-образно расположенных отрезков

¹ Фреше Морис (1878—1973) — французский математик, один из основоположников теоретико-множественной топологии, первым дал общее определение метрического пространства (1906).

Эта мера близости удовлетворяет всем аксиомам метрики на классе Q . Покажем, что в определении метрики Фреше можно ограничиться варьированием параметризации только одной из кривых при фиксированной параметризации другой кривой. Действительно, для любых двух параметризаций $f_1(t)$ и $f_2(t)$ одной и той же кривой всегда найдется (см. п. 1) взаимно однозначное отображение $\varphi(t)$ отрезка $[a, b]$ на себя такое, что последовательное применение отображения φ и затем отображения f_1 (записывается $f_1 \circ \varphi$ и называется композицией отображений) дает тот же результат, что одно f_2 . Иными словами, параметризация $f_1 \circ \varphi$ и f_2 совпадают (записывается $f_1 \circ \varphi = f_2$). Наоборот, $f_1 \cong f_2 \circ \varphi^{-1}$.

Зафиксируем в определении метрики Фреше некоторую параметризацию g_1 . Множество $\{f\}$ всевозможных параметризаций первой кривой вместе с параметризацией g_1 второй кривой исчерпывает всевозможные соответствия между точками кривых. Чтобы убедиться в этом, рассмотрим произвольную параметризацию g_2 второй кривой, отличную от g_1 , и соответствие точек кривых, задаваемое парой параметризаций (f_1, g_2) , где f_1 — некоторая параметризация первой кривой. Между g_1 и g_2 существует связь через посредство некоторого взаимно однозначного отображения φ : $g_1 = g_2 \circ \varphi$. Применяя к отрезку $[a, b]$ отображение φ , а затем к преобразованному отрезку — отображения f_1 и g_2 , получим пару $(f_1 \circ \varphi, g_2 \circ \varphi)$ других параметризаций рассматриваемых кривых, но соответствие точек кривых здесь будет таким же, как в случае пары (f_1, g_2) . Композиция $f_1 \circ \varphi$, которую мы обозначим f_2 , принадлежит множеству $\{f\}$, и полученная пара параметризаций может быть записана в виде (f_2, g_1) . Таким образом, для произвольных g_2 и f_1 соответствие, задаваемое парой (f_1, g_2) , сводится к соответствию, задаваемому парой вида (f_2, g_1) , поэтому определение меры близости кривых согласно Фреше может быть записано в следующей несимметричной форме:

$$\mu(f, g) = \mu(g, f) = \inf_{(f)} \left\{ \max_t \rho(f(t), g(t)) \right\}. \quad (111)$$

Отношение между метриками Хаусдорфа и Фреше. Рассмотрим две кривые f и g , для которых мера близости по Фреше $\mu(f, g) = r$. Допустим существование на кривой f точки a_ε такой, что для любой точки $g(t)$ кривой g расстояние до a_ε больше r , т. е. $\rho(a_\varepsilon, g(t)) \geq r + \varepsilon$, где ε — произвольно малое положительное число. Это означало бы, что при любой параметризации кривых f и g максимальное значение величины $\rho(f(t), g(t))$ было бы не меньше числа $r + \varepsilon$. Тогда и нижняя грань множества таких максимумов (по всем параметризациям) не равнялась бы r . Таким образом, наше допущение неверно, и при любом ε для произвольной точки a кривой f найдется на кривой g , т. е. в точечном множестве G , такая точка b , что $\rho(a, b) < r + \varepsilon$. Следовательно, $\text{dist}(a, G) < r + \varepsilon$, и в силу произвольности точки a из точечного множества F кривой f все F лежит в $(r + \varepsilon)$ -окрестности множества G : $F \subset v_{r+\varepsilon}(G)$. Поскольку рассуждения симметричны относительно обеих кривых, справедливо и обратное включение $G \subset v_{r+\varepsilon}(F)$. Значит, $\chi(F, G) \leq r + \varepsilon$, следовательно,

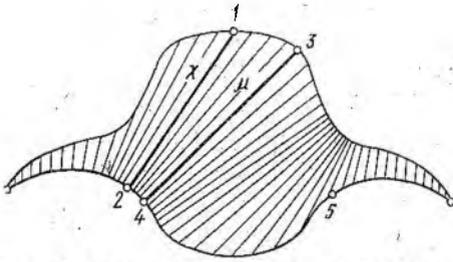


Рис. 13. «Близость» между гладкими графиками

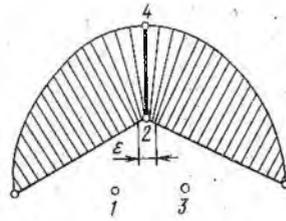


Рис. 14. Мера близости μ не достигается на отрезке изменения параметра

всегда хаусдорфово расстояние между множествами точек двух кривых не превышает меры близости этих кривых:

$$\chi(f, g) \leq \mu(f, g).$$

Мера близости двух кривых (параллельных отрезков) (см. рис. 11, б) равна расстоянию между левыми концами, так как при любой параметризации концы кривых обязательно соответствуют начальному или конечному значению параметра. Таким образом, в этом примере имеем равенство $\chi = \mu$. В примерах на рис. 12 мера близости Т-образно расположенных отрезков равна расстоянию между нижним концом вертикального отрезка и одним из концов горизонтального. В этих примерах $\chi < \mu$.

Наконец, рассмотрим пример (см. рис. 10), где представлены две кривые с одним и тем же множеством точек $F \equiv G$. В этом случае, очевидно, имеем $\chi(F, G) = \chi(F, F) = \chi(G, G) = 0$, тогда как значение μ здесь равно ширине петли, и поэтому $\chi < \mu$.

Если кривая f является простой дугой, т. е. не имеет самопересечений и самоприкосновений (самокасаний в случае гладкости), то ее множество точек не может служить множеством точек никакой другой кривой. Поэтому при сравнении кривых f и g , не имеющих самопересечений и самоприкосновений, из равенства $\chi(F, G) = 0$ вытекает равенство $\mu(f, g) = 0$.

Кривыми такого сорта являются графики функций $y = f(x)$, заданных на числовой оси, в частности дуги, которые взаимно однозначно проецируются на свои хорды. Однако и для таких кривых неравенство $\chi \leq \mu$ сохраняет силу, что иллюстрирует пример на рис. 13. Здесь для точки 1 верхней кривой f на нижней кривой g существует изолированная точка 2, отстоящая от точки 1 на расстояние $\chi(F, G)$. Допустим, что в этом примере $\mu(f, g) = \chi(F, G)$. Значение μ должно достигаться при таких параметризациях $f(t)$ и $g(t)$, когда точка 1 изопараметрична точке 2 (в силу симметрии картины вторую возможность, изопараметричность точек 1 и 5, можно не рассматривать). Параметризации непрерывные, и правой полуокрестности точки 1 должна соответствовать правая полуокрестность точки 2. Но это невозможно, так как любая точка правой полуокрестности точки 1 удалена от любой точки правой полуокрестности точки 2 больше, чем на χ . Следовательно, в этом примере

$\mu \neq \chi$. На рис. 13 тонкими прямыми показано изопараметрическое соответствие (одно из возможных), на котором достигается $\mu(f, g)$. Значение μ равно расстоянию между точками 3 и 4, образующими изопараметрическую пару в указанном соответствии, и примерно на 10 % превышает значение χ .

Мера близости $\mu(f, g)$ некоторых кривых не всегда может быть представлена длиной отрезка, один конец которого лежит на кривой f , а другой — на кривой g , как это имело место в предыдущем примере. Возможны случаи, когда изопараметрического соответствия между кривыми f и g , на котором бы достигалась $\mu(f, g)$, не существует. Рассмотрим пример.

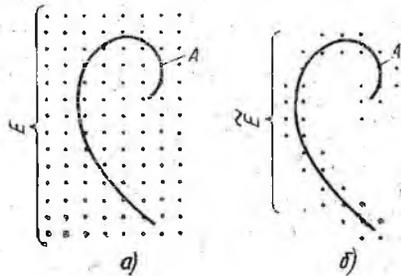
Верхняя кривая f на рис. 14 составлена из дуг трех окружностей с центрами 1, 2 и 3, а нижняя кривая g является двузвенной ломаной с точкой излома в центре второй окружности. Здесь величина $\mu(f, g)$ равна радиусу r второй окружности. Тонкими линиями на рис. 14 показано изопараметрическое соответствие между кривыми, при котором точки средней окружности изопараметричны точкам малой ε -окрестности точки излома на кривой g . Изопараметрические пары с максимальным расстоянием лежат чуть левее и чуть правее вертикального радиуса 2—4, и это расстояние больше радиуса r . Чем меньше ε , тем ближе к r значение максимального расстояния в изопараметрических парах, но вся дуга средней окружности ни при каких параметризациях $f(t)$ и $g(t)$ не может быть изопараметрична одной точке 2, которая здесь является точкой «инфимума максимумов» в том смысле, что к ней сходятся концы отрезков максимальных изопараметрических пар, причем длины этих отрезков, убывая, стремятся к $r = \mu$. Подобные примеры существуют и в случае, когда обе кривые гладкие. Приведенные примеры свидетельствуют о том, что наличие инфимума в определениях (110) и (111) является существенным.

Качество параметризаций. Ниже речь пойдет о приближении одних кривых другими кривыми, а именно — параметризуемыми вектор-сплайнами. Естественно было бы все оценки погрешности приближения получать в метрике μ , но, к сожалению, найти инфимум в выражениях (110), (111) по всем параметризациям не удается. Поэтому приходится выводить оценку максимума отклонений точек кривых при какой-то одной параметризации. Последняя может задавать такое изопараметрическое соответствие между точками кривых, что максимум расстояния будет значительно превышать μ . А поскольку оценивается именно он, то применительно к μ оценка оказывается более грубой. Напротив, при удачно выбранной параметризации можно получить точную оценку.

Конечные ε -сети. Суть численных методов состоит в переходе от бесконечных процедур над бесконечными множествами к процедурам, завершающимся в ограниченное время и имеющим своими результатами конечные множества. Эти последние должны в каком-то смысле аппроксимировать бесконечные множества. В метрическом пространстве такую функцию выполняют *конечные ε -сети*. Конечное множество E является конечной ε -сетью точечного множества A ,

Рис. 15. Конечная ε -сеть и локализованная конечная ε -сеть

если для любой точки a из A можно указать хотя бы одну точку e из E такую, что $\rho(a, e) \leq \varepsilon$ (рис. 15, а). Таким образом, исходя из данного определения можно заключить, что $\text{dist}(a, E) \leq \varepsilon$, и в силу произвольности точки a множество A входит



в ε -окрестность своей конечной ε -сети, т. е. $A \subset v_\varepsilon(E)$. Конечную ε -сеть множества, для которой справедливо и обратное включение $\bar{E} \subset v_\varepsilon(A)$, будем называть *локализованной конечной ε -сетью* множества A (рис. 15, б). Если A есть множество точек кривой и \bar{E} — его локализованная конечная ε -сеть, то хаусдорфово расстояние между ними удовлетворяет неравенству $\chi(A, \bar{E}) \leq \varepsilon$.

В соответствии с принципом численного подхода континуальное отображение $f(t)$, фигурирующее в определении кривой, следует заменить таким кусочно-постоянным отображением $\tilde{f}(t)$, которое переводит отрезок $[a, b]$ изменения параметра в конечную ε -сеть множества точек кривой, локализованную около этого множества. Практически достаточно оценить сверху значение шага, двигаясь с которым по отрезку $[a, b]$, нужно вычислять приближенные точки, образующие в совокупности локализованную конечную ε -сеть участка кривой между смежными узлами дискретно-точечного задания. Сама процедура получения точек этой локализованной конечной ε -сети может рассматриваться в качестве конструктивного определения кусочно-постоянного отображения $\tilde{f}(t)$.

Построенное таким путем кусочно-постоянное отображение, с одной стороны, как бы синтезирует концепции «близости» по Хаусдорфу и Фреше. С другой стороны, оно представляет собой аналог *сеточной функции* — математического понятия, столь широко используемого в численном анализе. Действительно, кусочно-постоянное отображение $\tilde{f}(t)$ аппроксимирует отображение $f(t)$ в том смысле, что образ любого интервала из отрезка $[a, b]$ при отображении $\tilde{f}(t)$ удален в хаусдорфовой метрике не более чем на ε от образа этого интервала при отображении $f(t)$. Сеточная функция, заданная на отрезке, есть кусочно-постоянное отображение его в числовую ось, т. е. в одномерное метрическое пространство. В указанном смысле сеточная функция также аппроксимирует обычную функцию, заданную на том же отрезке.

Все сказанное в этом параграфе относительно близости кривых переносится почти без изменения на случай сравнения взаимного расположения поверхностей. Только вместо отрезка изменения параметра здесь фигурирует прямоугольник, а роль концов дуги выполняет край куска поверхности.

3. НЕКОТОРЫЕ СВЕДЕНИЯ ИЗ ДИФФЕРЕНЦИАЛЬНОЙ ГЕОМЕТРИИ

В связи с необходимостью решения геометрических задач ниже дана сводка сведений о дифференцируемых кривых и поверхностях. Часть понятий основана на предположении существования в окрестности рассматриваемой точки непрерывных первых производных, другая — также непрерывных вторых производных, наконец, третья часть вводится в предположении непрерывности еще и третьих производных. Мы не будем каждый раз оговаривать эти условия. По изложению читатель сам увидит, когда какие из них предполагаются выполненными.

При операциях с векторами используются известные формулы векторного исчисления

$$\mathbf{r} \times (\mathbf{p} \times \mathbf{q}) = \mathbf{p}(\mathbf{q}\mathbf{r}) - \mathbf{q}(\mathbf{p}\mathbf{r}), \quad (\mathbf{p} \times \mathbf{q})(\mathbf{r} \times \mathbf{l}) = (\mathbf{p}\mathbf{r})(\mathbf{q}\mathbf{l}) - (\mathbf{p}\mathbf{l})(\mathbf{q}\mathbf{r}).$$

Пространственные кривые. Кривые задаются вектор-функциями $\mathbf{r} = \mathbf{r}(t)$. Орт касательной к кривой есть вектор $\boldsymbol{\tau} = \langle \mathbf{r}'(t) \rangle$. Орты главной нормали и бинормали суть

$$\mathbf{v} = \langle \mathbf{r}'(t) \times (\mathbf{r}'(t) \times \mathbf{r}''(t)) \rangle, \quad \boldsymbol{\beta} = \langle \mathbf{r}'(t) \times \mathbf{r}''(t) \rangle.$$

Тройка векторов $\boldsymbol{\tau}$, \mathbf{v} , $\boldsymbol{\beta}$ и попарно содержащие их плоскости образуют естественный трехгранник пространственной кривой.

Длина дуги кривой между точками, соответствующими значениям параметра t_1 и t_2 , определяется формулой

$$\sigma = \int_{t_1}^{t_2} |\mathbf{r}'(t)| dt.$$

При перемещении вдоль кривой из точки a в точку a_1 по дуге длиной $\Delta\sigma$ естественный трехгранник может поворачиваться. Если $\Delta\psi$ — угол поворота бинормали, то формулами

$$k = \lim_{\Delta\sigma \rightarrow 0} \left| \frac{\Delta\boldsymbol{\beta}}{\Delta\sigma} \right|; \quad T = \lim_{\Delta\sigma \rightarrow 0} \frac{\Delta\psi}{\Delta\sigma}$$

определяются соответственно кривизна и кручение кривой в точке a . Их вычисляют по формулам

$$k = \frac{|\mathbf{r}'(t) \times \mathbf{r}''(t)|}{|\mathbf{r}'(t)|^3}; \quad T = \frac{\mathbf{r}'(t) \cdot (\mathbf{r}''(t) \times \mathbf{r}'''(t))}{|\mathbf{r}'(t) \times \mathbf{r}''(t)|^2}.$$

Обратные величины — это радиус кривизны и радиус кручения.

В качестве параметра можно взять длину дуги σ (*лонгальный параметр*), отсчитываемую от некоторой точки. Такая параметризация называется *естественной (натуральной)*. Для нее

$$|\mathbf{r}'(\sigma)| = 1; \quad k = |\mathbf{r}''(\sigma)|; \quad T = \mathbf{r}'(\sigma) \cdot \mathbf{r}''(\sigma) \times \mathbf{r}'''(\sigma) \cdot |\mathbf{r}''(\sigma)|^{-2}.$$

Производные векторов $\boldsymbol{\tau}$, \mathbf{v} и $\boldsymbol{\beta}$ по σ выражаются *формулами Френе*

$$\boldsymbol{\tau}' = k(\sigma) \mathbf{v}; \quad \mathbf{v}' = -k(\sigma) \boldsymbol{\tau} + T(\sigma) \boldsymbol{\beta}; \quad \boldsymbol{\beta}' = -T(\sigma) \mathbf{v}.$$

Поверхности. Для задания поверхностей используются вектор-функции $\mathbf{r} = \mathbf{r}(t, s)$. Частные производные \mathbf{r}_t и \mathbf{r}_s определяют направления касательных к координатным линиям. Орт нормали к поверхности в точке a есть $\mathbf{n} = \langle \mathbf{r}_t \times \mathbf{r}_s \rangle$.

В теории поверхностей важную роль играют две квадратичные формы поверхности. В точке a поверхности проведем касательную плоскость. Любое смещение \mathbf{l} от точки a в этой плоскости можно разложить по векторам \mathbf{r}_t и \mathbf{r}_s следующим образом: $\mathbf{l} = \lambda_1 \mathbf{r}_t + \lambda_2 \mathbf{r}_s$. *Первая квадратичная форма* поверхности есть квадрат длины вектора \mathbf{l} , т. е. $I(\mathbf{l}) = |\mathbf{l}|^2$, или

$$I(\mathbf{l}) = E\lambda_1^2 + 2F\lambda_1\lambda_2 + G\lambda_2^2,$$

где коэффициенты $E = \mathbf{r}_t^2$; $F = \mathbf{r}_t \mathbf{r}_s$; $G = \mathbf{r}_s^2$ являются функциями точки на поверхности. Очевидно, форма $I(\mathbf{l})$ положительна и обращается в нуль только при $\lambda_1 = \lambda_2 = 0$. На единичном векторе $\boldsymbol{\tau}$ форма $I(\boldsymbol{\tau}) = 1$, на векторе $l\boldsymbol{\tau}$ форма $I(l\boldsymbol{\tau}) = l^2 I(\boldsymbol{\tau})$. Отметим, что $(\mathbf{r}_t \times \mathbf{r}_s)^2 = EG - F^2 > 0$.

Длина кривой, задаваемой на поверхности уравнением $\rho(\omega) = \mathbf{r}[t(\omega), s(\omega)]$, выражается через первую квадратичную форму интегралом

$$\sigma = \int_{\omega_1}^{\omega_2} \sqrt{E\lambda_1^2 + 2F\lambda_1\lambda_2 + G\lambda_2^2} d\omega.$$

Посредством первой квадратичной формы можно выразить косинус угла между пересекающимися кривыми на поверхности. Если направления кривых в точке пересечения заданы векторами $\mathbf{l}(\lambda_1, \lambda_2)$ и $\mathbf{m}(\mu_1, \mu_2)$, то

$$\cos \phi = \frac{|\mathbf{lm}|}{|\mathbf{l}||\mathbf{m}|} = \frac{E\lambda_1\mu_1 + F(\lambda_1\mu_2 + \lambda_2\mu_1) + G\lambda_2\mu_2}{\sqrt{I(\mathbf{l})I(\mathbf{m})}}.$$

Отсюда угол между координатными линиями $\lambda_2 = 0$ и $\mu_1 = 0$ (или $\lambda_1 = \mu_2 = 0$) определяется формулой $\cos \phi = F/\sqrt{EG}$. Условие $F = 0$ является необходимым и достаточным для того, чтобы координатная сеть на поверхности была ортогональной.

Кривизна линий на поверхности. Введем понятие *второй квадратичной формы*, определяемой выражением

$$II(\boldsymbol{\tau}) = L\lambda_1^2 + 2M\lambda_1\lambda_2 + N\lambda_2^2.$$

Здесь λ_1 и λ_2 — координаты вектора-смещения $\boldsymbol{\tau} = \lambda_1 \mathbf{r}_t + \lambda_2 \mathbf{r}_s$ в касательной плоскости в точке a' , а коэффициенты выражаются формулами $L = \mathbf{r}_{tt} \mathbf{n}$; $M = \mathbf{r}_{ts} \mathbf{n}$; $N = \mathbf{r}_{ss} \mathbf{n}$, где \mathbf{n} — орт нормали в точке касания. В каждой точке поверхности значение второй квадратичной формы на единичном векторе равно кривизне нормального сечения поверхности в направлении $\boldsymbol{\tau}$. Если смещение $\mathbf{l} = l\boldsymbol{\tau}$, то $II(\mathbf{l}) = l^2 II(\boldsymbol{\tau})$.

Существуют два взаимно перпендикулярных направления, которые называются *главными направлениями* поверхности в точке. В этих направлениях нормальная кривизна достигает экстремаль-

ных значений k_1 и k_2 , именуемых *главными кривизнами*. Нормальная кривизна k_0 в направлении, составляющем угол φ с первым главным направлением, дается *формулой Эйлера*

$$k_0 = k_1 \cos^2 \varphi + k_2 \sin^2 \varphi.$$

Если секущая плоскость наклонена к нормали под углом α , то кривизна наклонного сечения k в точке, где рассматривается нормаль, связана с нормальной кривизной в том же направлении *формулой Менье* $k \cos \alpha = k_0$.

Значения главных кривизн — это корни квадратного уравнения

$$k^2 (EG - F^2) - k (LG - 2MF + NE) + (LN - M^2) = 0.$$

[†] Полусумму главных кривизн принято называть *средней кривизной* поверхности в данной точке, а их произведение — *полной* или *гауссовой кривизной*. По свойству корней квадратного уравнения

$$H = \frac{1}{2} (k_1 + k_2) = \frac{LG - 2MF + NE}{EG - F^2};$$

$$K = k_1 k_2 = \frac{LN - M^2}{EG - F^2}. \quad (112)$$

Поверхность, у которой средняя кривизна всюду равна нулю, называется *минимальной*. Физическим примером такой поверхности является упругая пленка, натянутая на некоторую замкнутую пространственную проволочную рамку.

Знак гауссовой кривизны определяется знаком выражения $D = LN - M^2$. Если $D > 0$, то точка называется эллиптической, если $D < 0$ — гиперболической и $D = 0$ — параболической точкой. Поверхности, для которых гауссова кривизна постоянна во всех точках, называются *поверхностями постоянной кривизны*. Примером их служит сфера.

Если $\tau = \lambda_1 \tau_1 + \lambda_2 \tau_2$ — единичный вектор, совпадающий с главным направлением в некоторой точке поверхности, то коэффициенты λ_1 и λ_2 удовлетворяют условию

$$\begin{vmatrix} \lambda_2^2 & -\lambda_1 \lambda_2 & \lambda_1^2 \\ E & F & G \\ L & M & N \end{vmatrix} = 0.$$

Линейчатые и развертывающиеся поверхности. Построим поверхность следующим образом. В каждой точке пространственной кривой $\rho(t)$ — направляющей — зададим непрерывную вектор-функцию $l(t)$ — образующую. Поверхность, описываемая уравнением

$$r(t, s) = \rho(t) + sl(t),$$

где s — параметр вдоль образующей, называется *линейчатой* (рис. 16). В частном случае, если $l(t)$ — единичный вектор, то s имеет смысл расстояния вдоль образующей. Признаком линейчатой поверхности является равенство нулю коэффициента N второй квадратичной формы. Вследствие этого согласно (112) гауссова кривизна

$K \leq 0$, т. е. главные кривизны имеют разные знаки или одна из них равна нулю.

Вектор, нормальный к поверхности, можно записать в виде

$$\mathbf{N} = \mathbf{r}_t \times \mathbf{r}_s = (\rho' \times \mathbf{l}) + s(\mathbf{l}' \times \mathbf{l}).$$

В общем случае векторы $\rho' \times \mathbf{l}$ и $\mathbf{l}' \times \mathbf{l}$ неколлинеарны, поэтому при перемещении вдоль образующей вектор нормали меняет направление. Касательная плоскость при этом поворачивается вокруг образующей.

Если векторы $\rho' \times \mathbf{l}$ и $\mathbf{l}' \times \mathbf{l}$ коллинеарны, то нормаль сохраняет свое направление вдоль образующей. Поверхность в этом случае называется *развертывающейся*. Векторное произведение ортогонально плоскости, содержащей его сомножители. Поэтому коллинеарность векторов $\rho' \times \mathbf{l}$ и $\mathbf{l}' \times \mathbf{l}$ имеет место при условии $\rho'(t) \mathbf{l}'(t) \mathbf{l}(t) = 0$. Таким образом, компланарность векторов $\rho'(t)$, $\mathbf{l}'(t)$, $\mathbf{l}(t)$ является признаком развертывающейся поверхности. Условие это означает, что во второй квадратичной форме не только $N = 0$, но и $M = 0$. Поэтому гауссова кривизна развертывающейся поверхности $K = 0$, т. е. все ее точки параболические.

Поверхности вращения. Поверхность, которая получается путем вращения пространственной кривой $\rho = \rho(t)$ вокруг оси, определяемой ортом \mathbf{l} , называется поверхностью *вращения*. Ее уравнение в векторной форме имеет вид

$$\mathbf{r}(t, \varphi) = [(\mathbf{l} \times \Delta\rho) \times \mathbf{l}] \cos \varphi + (\mathbf{l} \times \Delta\rho) \sin \varphi + (\mathbf{l}, \Delta\rho) \mathbf{l} + \mathbf{r}_0.$$

Здесь $\Delta\rho = \rho(t) - \mathbf{r}_0$ — вектор, зависящий от t , а \mathbf{r}_0 — радиус-вектор некоторой точки, лежащей на оси вращения (рис. 17).

На практике наиболее часто встречается случай, когда кривая $\rho = \rho(t)$ — плоская и лежит в одной плоскости с вектором \mathbf{l} . Тогда линии $t = \text{const}$ представляют собой параллели поверхности вращения, а линии $\varphi = \text{const}$ — ее меридианы. Коэффициент F первой квадратичной формы и коэффициент M второй квадратичной формы равны нулю. А это означает, что параллели и меридианы образуют ортогональную сеть и идут по главным направлениям поверхности.

Геодезическая кривизна линии на поверхности. Геодезические линии. Пусть на поверхности $\mathbf{r} = \mathbf{r}(t, s)$ задана дважды непрерывно-

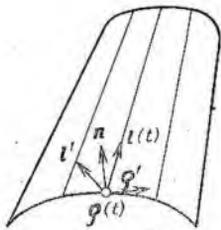


Рис. 16. Лиссажова поверхность

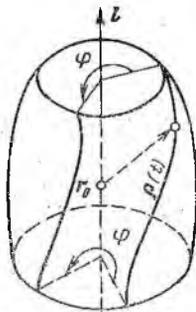


Рис. 17. Поверхность вращения

дифференцируемая кривая g уравнениями $t = t(\sigma)$, $s = s(\sigma)$, или $\rho(\sigma) = r(t(\sigma), s(\sigma))$. Если спроецировать окрестность точки a кривой g на касательную плоскость поверхности в этой точке, то проекция \bar{g} будет иметь в точке a кривизну κ , называемую *геодезической кривизной* кривой g в точке a . Кривизна k кривой g связана с геодезической кривизной κ соотношением $k = \kappa \cos \vartheta$, где ϑ — острый угол между дрямыми, на которых лежат орты главных нормалей ν кривой g и $\rho'(\sigma) \times \mathbf{n}$ кривой \bar{g} (\mathbf{n} — орт нормали к поверхности). Тогда с точностью до знака $\kappa = k\nu(\rho'(\sigma) \times \mathbf{n})$, или

$$\kappa = \rho''(\sigma) \rho'(\sigma) \cdot \mathbf{n}.$$

В случае произвольной параметризации $\rho = \rho(w)$ кривой g

$$\kappa = \frac{\rho''(w) \rho'(w) \cdot \mathbf{n}}{|\rho'(w)|^3}. \quad (113)$$

Кривая на поверхности называется *геодезической линией* (или просто — геодезической), если у нее в каждой точке геодезическая кривизна равна нулю. Всякий прямолинейный отрезок, принадлежащий поверхности, является геодезической линией. Характеристический признак криволинейной геодезической (с кривизной, отличной от нуля) состоит в том, что ее главная нормаль в каждой точке совпадает с нормалью \mathbf{n} поверхности. Через каждую внутреннюю точку поверхности в любом направлении проходит единственная геодезическая. Через две точки на поверхности могут проходить несколько геодезических линий. Например, на плоскости существует одна геодезическая, а на сфере любая дуга большого круга, проходящая через эти точки, есть геодезическая. На геодезических достигается кратчайшее внутреннее расстояние.

4. ВЕКТОР-СПЛАЙНЫ ОДНОЙ И ДВУХ ПЕРЕМЕННЫХ

В этом и последующих параграфах для описания кривых и поверхностей используются вектор-функции, компоненты которых являются полиномиальными сплайнами.

Вектор-сплайн одной переменной. Так будем называть вектор-функцию, определенную на отрезке $[a, b]$ вещественной переменной t в виде:

$$\mathbf{V}_{n,\nu}(t) = V_x(t) \mathbf{e}_1 + V_y(t) \mathbf{e}_2 + V_z(t) \mathbf{e}_3,$$

где $V_x(t)$, $V_y(t)$ и $V_z(t)$ — сплайн-функции степени n класса C^m ($\nu = n - m$), заданные на некоторой сетке $\Delta: a = t_1 < t_2 < \dots < t_N = b$.

Понятие вектор-сплайна можно распространить на векторное пространство произвольной конечной размерности.

Аппроксимация вектор-функции $\mathbf{r}(t)$ с компонентами $x(t)$, $y(t)$, $z(t)$ вектор-сплайном состоит в аппроксимации каждой из них сплайн-функциями $V_x(t)$, $V_y(t)$ и $V_z(t)$ соответственно. Это может быть интерполяция, локальная аппроксимация, аппроксимация со сглаживанием.

Погрешность приближения характеризуется остаточным членом $R(t) = V_{n,v}(t) - r(t)$. Его будем оценивать величиной

$$\|R(t)\| = \max_{t \in [a, b]} |V_{n,v}(t) - r(t)|,$$

которую можно рассматривать как норму в пространстве вектор-функций по аналогии с пространством вещественных (скалярных) непрерывных функций, введенных в гл. I, п. 1. Для вектор-функций классов гладкости $C^m[a, b]$ или $C^{(m)}[a, b]$ остается в силе буквально все, что было получено для оценок погрешности аппроксимации скалярных функций, и, следовательно, остаются в силе и сами оценки, если входящие в них величины заменить их векторными аналогами. В качестве примера приведем оценку погрешности интерполяции вектор-функции класса $C^{(2)}[a, b]$ вектор-сплайном первой степени. По аналогии с формулой (28) она имеет вид

$$\|R^{(p)}(t)\|_{[t_i, t_{i+1}]} \leq K_p (\Delta t_i)^{2-p} \|r''(t)\|_{[t_i, t_{i+1}]}, \quad p = 0, 1,$$

где $K_0 = 1/8$; $K_1 = 1/2$.

Кубические вектор-сплайны. По аналогии с кубическими сплайн-функциями на каждом промежутке $[t_i, t_{i+1}]$ их записывают в виде

$$V_{n,v}(t) = \varphi_1(u) V(t_i) + \varphi_2(u) V(t_{i+1}) + \Delta t_i [\varphi_3(u) V'(t_i) + \varphi_4(u) V'(t_{i+1})]. \quad (114)$$

Напомним, что u — локальный параметр сплайна — связан с переменной t соотношением $u = (t - t_i)/\Delta t_i$, функции $\varphi_k(u)$, $k = 1, 2, 3, 4$, определены формулами (21). Поскольку в дальнейшем в основном рассматриваются кубические вектор-сплайны, то для простоты обозначим их через $V(t)$ и будем называть *V-сплайнами*.

Для построения *V-сплайна* класса C^1 (*эрмитова V-сплайна*) требуется в узлах сетки Δ задать его значения V_i и значения его первой производной V'_i . *V-сплайн* класса C^2 определяется по значениям V_i в узлах сетки Δ с учетом граничных условий типов I—IV (гл. I, п. 6). Для построения его компонент $V_x(t)$, $V_y(t)$ и $V_z(t)$ нужно решить три системы уравнений.

Свойства сплайновых кривых. *V-сплайн*, вообще говоря, задает кривую в трехмерном евклидовом пространстве, которую в дальнейшем будем называть *сплайновой кривой*. Ее геометрия естественно определяется функциональной структурой *V-сплайна* и конкретными значениями его параметров. Рассмотрим некоторые свойства сплайновых кривых.

Свойство 1. Если векторы $\Delta V_i = V_{i+1} - V_i$, V'_i и V'_{i+1} коллинеарны, то $V(t)$ на интервале $[t_i, t_{i+1}]$ описывает кривую, лежащую в плоскости этих векторов.

Это утверждение становится очевидным, если выражение для *V-сплайна* с учетом равенства $\varphi_1(u) + \varphi_2(u) = 1$ представить в виде

$$V(t) - V_i = \varphi_2(u) \Delta V_i + \Delta t_i [\varphi_3(u) V'_i + \varphi_4(u) V'_{i+1}].$$

Справа стоит линейная комбинация компланарных векторов. Следовательно, вектор, стоящий слева в этом равенстве, лежит в той же плоскости.

Свойство 2. Если векторы ΔV_i , V_i , V_{i+1} коллинеарны, то точки $\mathbf{r} = \mathbf{V}(t)$, $t \in [t_i, t_{i+1}]$ лежат на прямой, проходящей через узлы V_i и V_{i+1} .

Действительно, если положить $\Delta V_i = \alpha V_i' = \beta V_{i+1}'$, то выражение (114) можно переписать в виде

$$\mathbf{V}(t) = \mathbf{V}_i + \Delta V_i \left[\varphi_2(u) + \frac{\Delta t_i}{\alpha} \varphi_3(u) + \frac{\Delta t_i}{\beta} \varphi_4(u) \right] = \mathbf{V}_i + \Delta V_i \Phi(u),$$

что и доказывает свойство. В частности, если $\alpha = \beta = \Delta t_i$, то $\Phi(u) = u$, и в этом случае $\mathbf{r} = \mathbf{V}(t)$ — уравнение хорды $[V_i, V_{i+1}]$.

Свойство 3. Сплайновая кривая при отличных от нуля значениях V_j имеет особую точку на i -м звене, если

$$\Delta V_i = \alpha V_i' + \beta V_{i+1}', \quad (3\alpha - \Delta t_i)(3\beta - \Delta t_i) = (3/4) \Delta t_i. \quad (115)$$

Эта особая точка есть точка возврата I или II рода.

В самом деле, необходимое условие того, что точка $\mathbf{V}(t^*)$ является особой точкой кривой, заданной уравнением $\mathbf{r} = \mathbf{V}(t)$, имеет вид

$$\mathbf{V}'(t^*) = \frac{1}{\Delta t_i} [\varphi_2'(u^*) \Delta V_i + \Delta t_i \varphi_3'(u^*) V_i' + \Delta t_i \varphi_4'(u^*) V_{i+1}'] = 0.$$

Отметим, что $u^* \neq 0; 1$, так как по условию V_i и V_{i+1} не равны нулю. Следовательно, $\varphi_2'(u^*) \neq 0$, и первое условие из (115) имеет место при

$$\alpha = -\Delta t_i \frac{\varphi_3'(u^*)}{\varphi_2'(u^*)} = \Delta t_i \frac{3u^* - 1}{6u^*};$$

$$\beta = -\Delta t_i \frac{\varphi_4'(u^*)}{\varphi_2'(u^*)} = \Delta t_i \frac{2 - 3u^*}{6(1 - u^*)}.$$

Исключив из этих равенств u^* , получим второе соотношение из (115).

Чтобы установить тип особой точки, рассмотрим значение второй производной в ней:

$$\mathbf{V}''(t^*) = \frac{1}{(\Delta t_i)^2} [\varphi_2''(u^*) \Delta V_i + \Delta t_i \varphi_3''(u^*) V_i' + \Delta t_i \varphi_4''(u^*) V_{i+1}'] =$$

$$= \frac{1}{(\Delta t_i)^2} [(\alpha \varphi_2''(u^*) + \Delta t_i \varphi_3''(u^*)) V_i' + (\beta \varphi_2''(u^*) + \Delta t_i \varphi_4''(u^*)) V_{i+1}'].$$

Опуская выкладки, приведем следующие сведения. Если $V_i' = \alpha V_{i+1}'$, то $\mathbf{V}''(t^*) \neq 0$ при всех $t^* \in [t_i, t_{i+1}]$. В противном случае $\mathbf{V}''(t^*) = 0$ при условии, что $\alpha + \beta = \Delta t_i$. Нетрудно убедиться, что это равенство несовместно со вторым соотношением из (115). Значит, $\mathbf{V}''(t^*) \neq 0$ и, как известно [18], особая точка есть точка возврата I или II рода.

На основании свойства 3 можно утверждать, что сплайновая кривая будет гладкой, если для ее плоских звеньев коэффициенты разложения α, β вектора хорды ΔV_i относительно репера (V_i, V_{i+1}) будут больше $\Delta t_i/3$.

Свойство 4. Сплайновая кривая не имеет точек с нулевой кривизной, если для каждого плоского ее звена выполняется соотношение

$$\Delta \mathbf{V}_i = \alpha \mathbf{V}'_i + \beta \mathbf{V}'_{i+1} \quad (116)$$

с коэффициентами $\alpha, \beta > \Delta t_i/3$.

Для доказательства этого утверждения рассмотрим выражение, стоящее в числителе формулы кривизны сплайновой кривой:

$$\mathbf{p}(t) = \mathbf{V}'(t) \times \mathbf{V}''(t) = \Delta t_i A(u) (\mathbf{V}'_i \times \Delta \mathbf{V}_i) + \Delta t_i B(u) (\mathbf{V}'_{i+1} \times \Delta \mathbf{V}_i) + \Delta t_i^2 D(u) (\mathbf{V}'_i \times \mathbf{V}'_{i+1}), \quad (117)$$

где $A(u) = 6(1-u)^2$; $B(u) = 6u^2$; $D(u) = 2[u(1-u) - u^2] - (1-u)^2$.

Пусть в некоторой точке $t^* \in [t_i, t_{i+1}]$ $\mathbf{p}(t^*) = 0$. Умножим скалярно это равенство на вектор $\Delta \mathbf{V}_i$:

$$\mathbf{p}(t^*) \Delta \mathbf{V}_i = \Delta t_i^2 D(u^*) \Delta \mathbf{V}_i \mathbf{V}'_i \mathbf{V}'_{i+1} = 0.$$

Так как $D(u) \neq 0$ для всех $u \in [0, 1]$ и векторы $\Delta \mathbf{V}_i, \mathbf{V}'_i, \mathbf{V}'_{i+1}$ не нулевые, то они компланарны и имеет место соотношение $a\mathbf{V}'_i + b\mathbf{V}'_{i+1} + c\Delta \mathbf{V}_i = 0$; $a^2 + b^2 + c^2 \neq 0$. Это означает, что кривизна сплайновой кривой может обратиться в нуль только на ее плоском звене.

Предположим, что $c \neq 0$. Представим вектор $\Delta \mathbf{V}_i$ в виде (116) при $\alpha = -a/c$, $\beta = -b/c$ и подставим его в (117):

$$\mathbf{p}(t) = [\beta A(u) - \alpha B(u) + \Delta t_i D(u)] \Delta t_i (\mathbf{V}'_i \times \mathbf{V}'_{i+1}).$$

Поскольку $\mathbf{V}'_i \times \mathbf{V}'_{i+1} \neq 0$ (иначе это противоречит принятому предположению), то $\mathbf{p}(t)$ может обратиться в нуль только за счет составляющей в квадратных скобках. Если ее преобразовать к виду

$$\psi(u) = (3\alpha - \Delta t_i) u^2 + (3\beta - \Delta t_i) (1-u)^2 + \Delta t_i u (1-u),$$

то становится очевидным, что $\psi(u)$ и, следовательно, $\mathbf{p}(t)$ не обращаются в нуль при $\alpha, \beta > \Delta t_i/3$.

Вектор-сплайн двух переменных. Это есть вектор-функция со значениями в трехмерном евклидовом пространстве, определенная в прямоугольной области $\Omega: [a, b] \times [c, d]$ вещественных переменных t и s в виде

$$\mathbf{W}_{n,\nu}(t, s) = W_x(t, s) \mathbf{e}_1 + W_y(t, s) \mathbf{e}_2 + W_z(t, s) \mathbf{e}_3,$$

где $W_x(t, s), W_y(t, s), W_z(t, s)$ — сплайн-функции степени n класса $C^{m,m}[\Omega]$, заданные на прямоугольной сетке $\Delta = \Delta_t \times \Delta_s$; $\Delta_t: a = t_1 < t_2 < \dots < t_N = b$; $\Delta_s: c = s_1 < s_2 < \dots < s_M = d$.

Аппроксимация вектор-функции $\mathbf{r}(t, s)$ вектор-сплайном $\mathbf{W}_{n,\nu}(t, s)$ состоит в аппроксимации каждой из ее компонент $x(t, s), y(t, s), z(t, s)$ сплайн-функциями $W_x(t, s), W_y(t, s)$ и $W_z(t, s)$ соответственно. Погрешность аппроксимации определяется остаточным членом $\mathbf{R}(t, s) = \mathbf{W}(t, s) - \mathbf{r}(t, s)$. Будем оценивать его, как и для V -сплайна, в евклидовой метрике, обобщая оценки приближения сплайнами функций различного класса гладкости на случай приближения вектор-функций.

Кубический вектор-сплайн двух переменных. По аналогии с кубической сплайн-функцией двух переменных в каждом прямоугольнике $\Omega_{ij}: [t_i, t_{i+1}] \times [s_j, s_{j+1}]$ кубический вектор-сплайн двух переменных представляется выражением

$$\mathbf{W}_{3,v}(t, s) = \varphi^T(u) \mathbf{F}_{ij} \varphi(v), \quad v = 1, 2, \quad (118)$$

где $\varphi^T(u) = [\varphi_1(u), \varphi_2(u), \Delta t_i \varphi_3(u), \Delta t_i \varphi_4(u)]$;

$$\mathbf{F}_{ij} = \begin{bmatrix} \mathbf{W}_{ij} & \mathbf{W}_{i, j+1} & \mathbf{W}_{ij}^{(0,1)} & \mathbf{W}_{i, j+1}^{(0,1)} \\ \mathbf{W}_{i+1, j} & \mathbf{W}_{i+1, j+1} & \mathbf{W}_{i+1, j}^{(0,1)} & \mathbf{W}_{i+1, j+1}^{(0,1)} \\ \mathbf{W}_{ij}^{(1,0)} & \mathbf{W}_{i, j+1}^{(1,0)} & \mathbf{W}_{i, j}^{(1,1)} & \mathbf{W}_{i, j+1}^{(1,1)} \\ \mathbf{W}_{i+1, j}^{(1,0)} & \mathbf{W}_{i+1, j+1}^{(1,0)} & \mathbf{W}_{i+1, j}^{(1,1)} & \mathbf{W}_{i+1, j+1}^{(1,1)} \end{bmatrix};$$

$$\varphi(v) = \begin{bmatrix} \varphi_1(v) \\ \varphi_2(v) \\ \Delta s_j \varphi_3(v) \\ \Delta s_j \varphi_4(v) \end{bmatrix};$$

$u = (t - t_i)/\Delta t_i$; $v = (s - s_j)/\Delta s_j$ — локальные переменные вектор-сплайна \mathbf{W} , φ_k , $k = 1, 2, 3, 4$ — скалярные функции, определенные ранее.

В дальнейшем кубический вектор-сплайн двух переменных будем обозначать $\mathbf{W}(t, s)$ и называть **W-сплайном**. Чтобы определить **W-сплайн** класса $C^{(2,2)}$ — эрмитов **W-сплайн** — требуется в узлах сетки Δ задать значения \mathbf{W}_{ij} , $\mathbf{W}_{ij}^{(1,0)}$, $\mathbf{W}_{ij}^{(0,1)}$, $\mathbf{W}_{ij}^{(1,1)}$. **W-сплайн** класса $C^{2,2}$ однозначно определяется по значениям \mathbf{W}_{ij} в узлах сетки Δ и граничным условиям, аналогичным I—IV (гл. I, п. 6). При этом недостающие элементы матрицы \mathbf{F}_{ij} : $\mathbf{W}_{ij}^{(1,0)}$, $\mathbf{W}_{ij}^{(0,1)}$, $\mathbf{W}_{ij}^{(1,1)}$, находятся путем решения соответствующих задач интерполяции для каждой из компонент **W-сплайна** $W_x(t, s)$, $W_y(t, s)$ и $W_z(t, s)$.

Свойства W-сплайнов. **W-сплайн** описывает некоторую поверхность в трехмерном евклидовом пространстве, которую в дальнейшем будем называть **сплайновой поверхностью**. Ее геометрические свойства определяются свойствами **W-сплайна** и в конечном счете зависят от его коэффициентов.

Сплайновая поверхность, как и сплайновая кривая, может иметь особые точки. Если для всех $t, s \in \Omega$ $\mathbf{W}^{(1,0)}(t, s) \times \mathbf{W}^{(0,1)}(t, s) \neq 0$, то сплайновая поверхность — гладкая. Это соотношение связывает 16 векторных величин: \mathbf{W}_{kl} , $\mathbf{W}_{kl}^{(1,0)}$, $\mathbf{W}_{kl}^{(0,1)}$, $\mathbf{W}_{kl}^{(1,1)}$, $k = i, i+1$, $l = j, j+1$ — значения **W-сплайна** в угловых точках области Ω_{ij} , $i = 1, 2, \dots, N-1$, $j = 1, 2, \dots, M-1$ и в отличие от аналогичного условия для сплайновых кривых имеет более сложный вид. Рассмотрим следующие свойства сплайновой поверхности.

Свойство 1. Если коэффициенты **W-сплайна** принадлежат некоторой плоскости P , т. е. имеют место равенства

$$\begin{aligned} (\mathbf{W}_{ij} - r_0) \mathbf{n} &= 0; \quad \mathbf{W}_{ij}^{(1,0)} \mathbf{n} = 0; \\ \mathbf{W}_{ij}^{(0,1)} \mathbf{n} &= 0; \quad \mathbf{W}_{ij}^{(1,1)} \mathbf{n} = 0, \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, M, \end{aligned}$$

где r_0 — некоторая точка плоскости P ; n — нормаль плоскости, то сплайн в этом случае описывает некоторую область с криволинейной границей в этой плоскости.

Действительно, разность $W(t, s) - r_0$ можно представить в виде

$$W(t, s) - r_0 = \varphi^T(u) F_{ij} \varphi(v) - \varphi^T(u) R_0 \varphi(v) = \varphi^T(u) (F_{ij} - R_0) \varphi(v),$$

где

$$R_0 = \begin{bmatrix} r_0 & r_0 & 0 & 0 \\ r_0 & r_0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Поскольку по условию $(F_{ij} - R_0) n = 0$, то $(W(t, s) - r_0) n = 0$ для всех $t, s \in \Omega$, а это и означает, что $W(t, s)$ принадлежит плоскости P .

Свойство 2. Для того чтобы координатные линии одного из семейств (для определенности пусть это будут линии $s = \text{const}$) сплайновой поверхности были плоско-параллельными сечениями, достаточны следующие условия:

$$\Delta W_i(s_j) n = [W(t_{i+1}, s_j) - W(t_i, s_j)] n = 0; \quad (119)$$

$$\Delta W_i^{(0,1)}(s_j) n = 0; \quad W_{ij}^{(1,0)} n = 0; \quad W_{ij}^{(1,1)} n = 0,$$

где n — нормаль плоскостей семейства.

Действительно, рассмотрим координатную линию $s = s^* \in [s_j, s_{j+1}]$:

$$V(t) = W(t, s^*) = \varphi_1(u) W(t_i, s^*) + \varphi_2(u) W(t_{i+1}, s^*) + \\ + \Delta t_i (\varphi_3(u) W^{(1,0)}(t_i, s^*) + \varphi_4(u) W^{(1,0)}(t_{i+1}, s^*)).$$

Если она принадлежит плоскости с нормалью n , то

$$\Delta W_i(s^*) n = 0; \quad W_i^{(1,0)}(s^*) n = 0; \quad W_{i+1}^{(1,0)}(s^*) n = 0,$$

где $W_i^{(1,0)}(s^*) = W^{(1,0)}(t_i, s^*)$. Используя выражение для W -сплайна, эти равенства представим в виде

$$\Delta W_i(s^*) n = [\varphi_1(v^*) \Delta W_i(s_j) + \varphi_2(v^*) \Delta W_i(s_{j+1}) + \\ + \Delta s_j (\varphi_3(v^*) \Delta W_i^{(0,1)}(s_j) + \varphi_4(v^*) \Delta W_i^{(0,1)}(s_{j+1}))] n = 0; \quad (120)$$

$$W_i^{(1,0)}(s^*) n = [\varphi_1(v^*) W_{ij}^{(1,0)} + \varphi_2(v^*) W_{i,j+1}^{(1,0)} + \\ + \Delta s_j (\varphi_3(v^*) W_{ij}^{(1,1)} + \varphi_4(v^*) W_{i,j+1}^{(1,1)})] n = 0.$$

Так как выражение $\alpha \varphi_1(v) + \beta \varphi_2(v) + \delta \varphi_3(v) + \gamma \varphi_4(v)$ не обращается в нуль тождественно ни при каких $\alpha, \beta, \gamma, \delta$, из которых хотя бы одно не нулевое, равенства (120) имеют место при условиях (119).

5. АППРОКСИМАЦИЯ КРИВЫХ V-СПЛАЙНАМИ

Решаемые на практике задачи математического моделирования кривых и поверхностей по их содержанию можно отнести к одному из следующих видов: приближенное описание этих геометрических

объектов, восстановление и генерация. Задачи первого вида состоят в подборе для заданной кривой или поверхности достаточно простого математического выражения, описывающего их с требуемой точностью. В задачах второго вида требуется восстановить кривую или поверхность по некоторым данным о них. При этом представляет интерес вопрос: какова будет точность восстановления? Задачи генерации — это задачи построения кривых и поверхностей с заданными свойствами.

О приближенном вычислении длин дуг кривых. Пусть $\mathbf{r} = \mathbf{r}(t)$, $t \in [a, b]$, есть параметризация кривой g . Выберем на отрезке $[a, b]$ равномерную сетку с шагом δt и рассмотрим ее i -й участок. Длина дуги участка и длина его хорды суть

$$\Delta\sigma_i = \int_{t_i}^{t_{i+1}} |\mathbf{r}'(t)| dt; \quad h_i = |h_i| = |\mathbf{r}(t_{i+1}) - \mathbf{r}(t_i)|, \quad t_i = a + (i-1)\delta t.$$

Рассмотрим разность

$$\Delta\sigma_i - h_i = \int_{t_i}^{t_{i+1}} \left(|\mathbf{r}'(t)| - \frac{|\mathbf{r}(t_{i+1}) - \mathbf{r}(t_i)|}{\delta t} \right) dt.$$

Используя стандартную технику (гл. I, п. 2), находим оценку

$$\Delta\sigma_i - h_i \leq (1/3) (\delta t)^2 \|\mathbf{r}''(t)\|_{[t_i, t_{i+1}]}. \quad (121)$$

Суммируя обе части неравенства по i и заменяя $\|\mathbf{r}''(t)\|_{[t_i, t_{i+1}]}$ на $\|\mathbf{r}''(t)\|_{[a, b]}$, получаем оценку разности длины дуги S и ломаной L :

$$S - L \leq (1/3) (b - a) \delta t \|\mathbf{r}''(t)\|_{[a, b]}.$$

Для того чтобы разность не превышала ε , достаточно взять шаг длиной

$$\delta t = \frac{3\varepsilon}{(b-a) \|\mathbf{r}''(t)\|_{[a, b]}}. \quad (122)$$

Приближенное описание кривой V-сплайном. Пусть g — заданная кривая и $\mathbf{r} = \mathbf{r}(t)$ ее параметризация. Требуется определить V-сплайн, приближающий кривую g с точностью ε .

На этот случай распространяются все рекомендации и алгоритмы, приведенные в гл. I, п. 5, 6 при рассмотрении интерполяции с заданной точностью эрмитовым кубическим или кубическим сплайном скалярной функции. Например, можно определить эрмитов V-сплайн на равномерной сетке, воспользовавшись для этой цели формулами, аналогичными (36) и (37), или определить его на неравномерной сетке, если возникает необходимость минимизировать число узлов. В любом случае результат зависит не только от дифференциальных свойств заданной кривой g , но и от ее конкретной параметризации $\mathbf{r}(t)$. Чтобы сетка не зависела от параметризации, можно в качестве параметра приближающего V-сплайна взять лонгальный параметр кривой. Это делается следующим образом.

Вспомним, что при построении неравномерной сетки область изменения параметра t — отрезок $[a, b]$ — разбивается с шагом δt точками $t_j = a + (j - 1) \delta t$, $j = 1, 2, \dots$. Некоторые из них выбирались в качестве узлов сетки сплайна, остальные использовались для контроля точности приближения. Алгоритм учитывает только дискретные значения аргумента. Поэтому, чтобы от параметра t перейти к лонгальному параметру, достаточно в точках t_j подсчитать значения $\sigma_j = \sigma(t_j)$. Шаг δt должен обеспечивать необходимую точность приближения длины дуги в точках t_j длиной ломаной с вершинами в этих точках. Для этого достаточно выбрать его по формуле (122).

Алгоритм модифицируется следующим образом:

1. Определяется шаг δt по правилу

$$\delta t = \min \left[\sqrt{\frac{3(\varepsilon - \varepsilon_1)}{A}}, \frac{3\varepsilon}{(b-a)A} \right], \quad \text{где } A \geq \|r''(t)\|_{[a, b]}.$$

2. В каждой точке t_j подсчитывается приближенное значение длины дуги по формуле

$$\sigma_j = \sum_{k=1}^{j-1} h_k, \quad \sigma_1 = 0.$$

3. V-сплайн строится относительно параметра σ .

4. Контроль точности интерполяции осуществляется не во всех точках t_j , а только в точках $t_i = (i - 1) \Delta_j \delta t$, $i = 1, 2, \dots$, где Δ_j — целое число.

$$\Delta_j = \text{ent} \left[(b - a) \sqrt{\frac{A(\varepsilon - \varepsilon_1)}{3\varepsilon^2}} \right] + 1.$$

Величина Δ_j определяется отношением шага, с которым осуществляется контроль точности, к шагу, обеспечивающему точность ε подсчета длины дуги.

Вычислительные затраты модифицированного алгоритма несколько больше, чем для исходного алгоритма, вследствие меньшего шага δt .

Пример. Рассмотрим приближение окружности и эллипса эрмитовым V-сплайном с предельной погрешностью ε . Их уравнения возьмем в виде

$$x = R \cos \varphi, \quad y = R \sin \varphi, \quad \varphi \in [0, 2\pi];$$

$$x = a \cos \varphi, \quad y = b \sin \varphi, \quad \varphi \in [0, 2\pi], \quad a > b.$$

Учитывая, что для окружности $\|r^{IV}(\varphi)\| = R$ и для эллипса $\|r^{IV}(\varphi)\| = a$, из оценки погрешности интерполяции (34) получаем шаг равномерной сетки, обеспечивающий заданную точность приближения ε : $h_{\text{окр}} = (384\varepsilon/R)^{1/4}$, $h_{\text{эл}} = (384\varepsilon/a)^{1/4}$. Например, окружность радиуса 1000 мм и эллипс с большой полуосью $a = 1000$ мм описываются эрмитовым V-сплайном при $\varepsilon = 0,1$ мм на равномерной сетке с шагом $h = 0,4427$. Такому шагу соответствует число узлов $N = \lceil 2\pi/h \rceil + 2 = 16$.

Для сравнения ниже приведены сетки эрмитовых V-сплайнов, интерполирующих первую четверть эллипса с полуосями $a = 1000$ мм, $b = 10$ мм, $\varepsilon = 0,1$ мм для следующих случаев:

а) равномерного разбиения области параметра φ :

$$\Delta_\varphi : \{0; 0,392; 0,784; 1,176; 1,57\};$$

б) неравномерного разбиения области параметра φ с учетом монотонности величины $|\mathbf{r}^{IV}(\varphi)|$:

$$\Delta_{\varphi} : \{0; 0,4427; 0,8967; 1,397; 1,57\};$$

в) неравномерного разбиения области лонгального параметра s по приведенному выше модифицированному алгоритму:

$$\Delta_s : \{0; 0,946; 10,366; 70,96; 339,81; 1000,2\}.$$

Большее число узлов сетки Δ_s по сравнению с числом узлов сеток Δ_{φ} объясняется тем, что естественная параметризация эллипса приближается V -сплайном хуже, чем параметризация относительно параметра φ .

Восстановление кривой V -сплайном. В практических приложениях встречается большое количество самых разнообразных задач, связанных с восстановлением кривой по некоторым неполным данным. Мы рассмотрим только наиболее общие задачи, к которым сводятся многие другие.

Задача 1. Кривая g задана в дискретно-точечной форме значениями $\mathbf{r}_i = \mathbf{r}(t_i)$, $\mathbf{r}'_i = \mathbf{r}'(t_i)$, t_i , $i = 1, \dots, N$, относящимися к некоторой параметризации $\mathbf{r}(t)$ этой кривой. Требуется восстановить кривую V -сплайном.

Решением этой задачи является эрмитов V -сплайн, интерполирующий вектор-функцию $\mathbf{r} = \mathbf{r}(t)$ по заданным значениям. Отметим, что существует бесчисленное множество V -сплайнов, восстанавливающих кривую по исходным данным. Действительно, при произвольной монотонной функции $t = t(\xi)$ вектор-функция $\mathbf{r} = \mathbf{r}[t(\xi)]$ также является параметризацией кривой g , и можно подсчитать соответствующие значения $\{\mathbf{r}_i, \mathbf{r}'_i(\xi_i), \xi_i\}$.

Если известно, к какому классу гладкости принадлежит кривая g , то можно оценить точность ее восстановления. В случае $g \in C^2$ по формулам типа (13) находим приближенные значения $\tilde{\mathbf{r}}_i = 2\mathbf{r}[t_{i-1}, t_i, t_{i+1}] = \mathbf{r}'_i + o(1)$, $i = 2, \dots, N-1$. Теперь можно воспользоваться оценками типа (33), подставляя в них $\max(|\tilde{\mathbf{r}}_i|, |\tilde{\mathbf{r}}_{i+1}|)$ вместо $\|\mathbf{r}''(t)\|_{[t_i, t_{i+1}]}$. При этом погрешность, вносимая в правую часть, будет малой более высокого порядка. Аналогично, для кривой $g \in C^4$ можно применять оценки (34) в том же смысле, заменяя четвертую производную четвертыми разделенными разностями.

Задача 2. Кривая g представлена в дискретно-точечной форме значениями $\mathbf{r}_i, \boldsymbol{\tau}_i, t_i$, $i = 1, \dots, N$. (Здесь и далее $\boldsymbol{\tau}_i$ — орт касательной к кривой в i -й точке.) Требуется восстановить кривую V -сплайном.

Задача совпадает с задачей 1, если $t = \sigma$ — лонгальный параметр. В противном случае нужно найти величины $\sigma'(t_i) > 0$, а затем $\mathbf{r}'_i = \sigma'(t_i) \boldsymbol{\tau}_i$. И снова приходим к задаче 1. Приведем два способа определения $\gamma_i = \sigma'(t_i)$.

1. Вычисляем $\sigma'(t_i)$ приближенно по формулам (11), заменяя в них еще $\Delta\sigma_i$ длинами хорд h_i . Получаем

$$\bar{\sigma}'(t_1) = (1 + \mu_2) \frac{h_1}{\Delta t_1} - \mu_2 \frac{h_2}{\Delta t_2}; \quad (123a)$$

$$\bar{\sigma}'(t_i) = \lambda_i \frac{h_{i-1}}{\Delta t_{i-1}} + \mu_i \frac{h_i}{\Delta t_i}, \quad i = 2, \dots, N-1; \quad (1236)$$

$$\bar{\sigma}'(t_N) = -\lambda_{N-1} \frac{h_{N-2}}{\Delta t_{N-2}} + (1 + \lambda_{N-1}) \frac{h_{N-1}}{\Delta t_{N-1}}, \quad (123B)$$

где $\lambda_i = \Delta t_i (\Delta t_{i-1} + \Delta t_i)^{-1}$; $\mu_i = 1 - \lambda_i$.

Согласно (12) формулы (11) обеспечивают точность приближения производных в узлах сетки порядка $O(\Delta t)$. Замена $\Delta \sigma_i$ на h_i в силу (121) вносит погрешность того же порядка. Вследствие приближенного определения величин \bar{g}_i точность восстановления кривой здесь может оказаться ниже, чем в задаче 1. В оценках (33) константы будут примерно в 4 раза большими. Оценки (34) в данном случае вообще места не имеют. Максимальный порядок приближения не $O(\Delta t^4)$, а только $O(\Delta t^3)$, и достигается он для кривых $g \in C^{(3)}$.

2. Определим величины γ_i из условий минимума функционала

$$J(\gamma_1, \dots, \gamma_N) = \int_a^b |\mathbf{V}''(t)|^2 dt.$$

Заменяя интеграл на сумму интегралов по отрезкам $[t_i, t_{i+1}]$ и подставляя в них выражение V-сплайна (114), приходим к формуле

$$J(\gamma_1, \dots, \gamma_N) = \sum_{i=1}^{N-1} [4(\Delta t_i)^{-1}(\gamma_i^2 + \gamma_i \gamma_{i+1} \tau_i \tau_{i+1} + \gamma_{i+1}^2) - 12(\Delta t_i)^{-2}(\gamma_i \mathbf{h}_i \tau_i + \gamma_{i+1} \mathbf{h}_i \tau_{i+1}) + (\Delta t_i)^{-3} h_i^2],$$

где \mathbf{h}_i — векторы хорд.

Необходимые условия минимума функционала дают систему уравнений для определения неизвестных γ_i :

$$c_i \gamma_{i-1} + 2\gamma_i + b_i \gamma_{i+1} = d_i, \quad i = 1, \dots, N,$$

где $c_i = \mu_i \tau_{i-1} \tau_i$; $b_i = \lambda_i \tau_i \tau_{i+1}$, $\mu_1 = \lambda_N = 0$; $\lambda_1 = \mu_N = 1$;

$$d_i = 3 \left(\frac{\mu_i}{\Delta t_{i-1}} \tau_i \mathbf{h}_{i-1} + \frac{\lambda_i}{\Delta t_i} \tau_i \mathbf{h}_i \right).$$

Матрица системы — трехдиагональная с диагональным преобладанием. Во-первых, для решения системы применим метод прогонки (гл. 1, п. 10). Во-вторых, сравнивая эту матрицу с матрицей системы (58) и повторяя проделанные ранее рассуждения, находим, что все $\gamma_i > 0$, если $2d_i - c_i d_{i-1} - b_i d_{i+1} > 0$, $i = 1, \dots, N$, $d_0 = d_{N+1} = 0$. Можно показать, что эти условия выполняются, если имеют место неравенства

$$2 \min \left(\frac{\tau_i \mathbf{h}_{i-1}}{\Delta t_{i-1}}, \frac{\tau_i \mathbf{h}_i}{\Delta t_i} \right) > \max \left(\frac{h_{i-2}}{\Delta t_{i-2}}, \dots, \frac{h_{i+1}}{\Delta t_{i+1}} \right).$$

В случае параметризации по суммарной длине хорд ($\Delta t_i = h_i$) эти неравенства имеют простой геометрический смысл, а именно:

$$\min [\cos(\widehat{\tau_i, \mathbf{h}_{i-1}}); \cos(\widehat{\tau_i, \mathbf{h}_i})] > (1/2),$$

т. е. угол наклона касательного вектора к соседним хордам не должен превышать $\pi/3$.

Задача 3. Кривая g задана в дискретно-точечной форме значениями \mathbf{r}_i , $i = 1, \dots, N$. Кроме того, в некоторых точках \mathbf{r}_k могут быть известны орты касательных $\boldsymbol{\tau}_k$, $k \in \{1, \dots, N\}$. Требуется восстановить кривую V -сплайном.

Во-первых, определим орты касательных в тех точках \mathbf{r}_i , где они не заданы. Через три точки \mathbf{r}_{i-1} , \mathbf{r}_i , \mathbf{r}_{i+1} проведем окружность. Орты касательных к окружности в точках \mathbf{r}_i :

$$\boldsymbol{\tau}_1 = \langle h_2(h_1 + h_2)\mathbf{h}_1 - h_1^2\mathbf{h}_2 \rangle; \quad (124a)$$

$$\boldsymbol{\tau}_i = \langle h_i^2\mathbf{h}_{i-1} + h_{i-1}^2\mathbf{h}_i \rangle, \quad i = 2, \dots, N-1; \quad (124б)$$

$$\boldsymbol{\tau}_N = \langle -h_{N-1}^2\mathbf{h}_{N-2} + h_{N-2}(h_{N-2} + h_{N-1})\mathbf{h}_{N-1} \rangle. \quad (124в)$$

В ряде случаев достаточно принять величины $\boldsymbol{\tau}_i$ в качестве ортов касательной к кривой. Но не всегда. Формулы (124) не учитывают изменения кривизны кривой в окрестности точки \mathbf{r}_i и потому применимы, только когда оно достаточно мало. Чтобы ослабить этот недостаток, усложним формулы (124). Используем дополнительно точки \mathbf{r}_{i-2} , \mathbf{r}_{i+2} и величины

$$\frac{1}{2\rho_{i-1}} = \frac{|\langle \mathbf{h}_{i-2} \rangle \times \langle \mathbf{h}_{i-1} \rangle|}{|\mathbf{h}_{i-2} + \mathbf{h}_{i-1}|}, \quad \frac{1}{2\rho_{i+1}} = \frac{|\langle \mathbf{h}_i \rangle \times \langle \mathbf{h}_{i+1} \rangle|}{|\mathbf{h}_i + \mathbf{h}_{i+1}|}.$$

Здесь ρ_{i-1} — радиус окружности, проходящей через точки \mathbf{r}_{i-2} , \mathbf{r}_{i-1} , \mathbf{r}_i . Аналогично ρ_{i+1} .

Вместо формулы (124б) введем

$$\boldsymbol{\tau}_i = \left\langle \frac{h_i^2}{\rho_{i+1}}\mathbf{h}_{i-1} + \frac{h_{i-1}^2}{\rho_{i-1}}\mathbf{h}_i \right\rangle, \quad i = 2, \dots, N-1, \quad (125)$$

полагая $\rho_1 = \rho_2$ и $\rho_N = \rho_{N-1}$.

Очевидно, что если пять точек $\mathbf{r}_{i-2}, \dots, \mathbf{r}_{i+2}$ лежат на одной окружности, то $\rho_{i-1} = \rho_{i+1}$, и формулы (124б) и (125) дают один и тот же результат. Величины $\boldsymbol{\tau}_1$ и $\boldsymbol{\tau}_N$ вычисляются по прежним формулам (124а) и (124в). Для гладкой замкнутой кривой формула (125) применяется во всех точках (с циклической заменой индексов). Если три точки \mathbf{r}_{i-2} , \mathbf{r}_{i-1} , \mathbf{r}_i лежат на одной прямой, то согласно (125) орты касательных в них одинаковы и равны $\langle \mathbf{h}_{i-1} \rangle$.

Второй шаг решения этой задачи заключается в присвоении заданным точкам \mathbf{r}_i значений некоторого параметра t : $t_1 < t_2 < \dots < t_N$. На практике для этой цели удобно использовать *суммарную длину хорд* вдоль ломаной с вершинами \mathbf{r}_i : $t_1 = 0$, $t_i = h_1 + \dots + h_{i-1}$, $i = 2, \dots, N$. Как было показано выше, этот параметр служит хорошим приближением логарифмического параметра кривой. В результате оказываемся в условиях задачи 2. Точность восстановления кривой примерно такая же, как и там.

Перечисленные задачи относятся к восстановлению кривой эрмитовым V -сплайном. Ниже приведены две задачи для V -сплайна класса C^2 .

Задача 4. Пусть $\mathbf{r} = \mathbf{r}(t)$, $t \in [a, b]$ — некоторая параметризация кривой g и известны значения \mathbf{r}_i в узлах некоторой сетки Δ . Требуется восстановить кривую V -сплайном.

Для определения V-сплайна необходимо ввести граничные условия одного из типов I—IV. В частности, по формулам (123а), (123в) и (124а), (124в) можно получить приближенные значения r_1^0 и r_N^0 , как в задаче 3, и использовать их в качестве граничных условий типа I. Оценка точности восстановления кривой по формулам типа (51) и (52) близка к истине.

Задача 5. Кривая g задана в дискретно-точечной форме значениями r_i , $i = 1, \dots, N$. Требуется восстановить ее V-сплайном. Эта задача сводится к предыдущей, если заданным точкам поставить в соответствие значения $t_1 < t_2 < \dots < t_N$ некоторого параметра t , например суммарной длины хорд.

Восстановление кривой со сглаживанием. В отличие от задач, уже рассмотренных, здесь предполагается, что исходные точки, по которым кривая должна быть восстановлена, заданы с некоторой погрешностью. Поэтому при определении восстанавливающего V-сплайна заданные точки можно варьировать в пределах вероятной погрешности.

Задача 6. Кривая g представлена в дискретно-точечной форме значениями r_i^0 , t_i , $i = 1, \dots, N$, относительно которых предполагается, что если $r(t)$, $t \in [a, b]$ — некоторая параметризация кривой, то $|r(t_i) - r_i^0| \leq \delta_i$. По этим условиям требуется восстановить кривую V-сплайном.

Решение ищем в виде V-сплайна класса C^2 , минимизирующего функционал

$$J(V) = \int_a^b |V''(t)|^2 dt$$

с ограничениями

$$|V(t_i) - r_i^0| \leq \delta_i, \quad i = 1, \dots, N, \quad (126)$$

и граничными условиями одного из типов: а) $V'(t_1) = r_1'$, $V'(t_N) = r_N'$, причем r_1' и r_N' , если они не заданы, можно определить приближенно, как в задаче 3; б) если кривая замкнутая, то используются периодические условия; в) условия вида $V''(t_1) = V''(t_N) = 0$ следует применять, если только известно, что кривизна кривой в окрестности концов близка к нулю.

Задача существенно упрощается, если ограничения (126) заменить тремя группами ограничений:

$$\begin{aligned} |V_x(t_i) - x_i^0| &\leq \delta_{xi}; & |V_y(t_i) - y_i^0| &\leq \delta_{yi}; \\ |V_z(t_i) - z_i^0| &\leq \delta_{zi}, & i &= 1, \dots, N, \end{aligned} \quad (127)$$

где δ_{xi} , δ_{yi} , δ_{zi} — неотрицательные числа такие, что $\delta_{xi}^2 + \delta_{yi}^2 + \delta_{zi}^2 = \delta_i^2$. Например, $\delta_{xi} = \delta_{yi} = \delta_{zi} = \delta_i/\sqrt{3}$. Вместо минимизации функционала $J(V)$ рассматривают три аналогичные задачи об определении скалярных функций $V_x(t)$, $V_y(t)$ и $V_z(t)$ с ограничениями (127). Их решение изложено в гл. 1, п. 7.

Сформулированную задачу можно решить также, используя алгоритм локального сглаживания, описанный в гл. I, п. 9. Для этого \mathbf{V} -сплайн представляется через B -сплайны:

$$\mathbf{V}(t) = \sum_{i=0}^{N+1} b_i B_i(t).$$

И здесь целесообразно рассматривать три задачи для сплайнов $V_x(t)$, $V_y(t)$, $V_z(t)$. Последние можно решать как при наличии ограничений (127), так и без таковых.

Задача 7. Кривая g представлена в дискретно-точечной форме значениями \mathbf{r}_v^0 , t_v , $v = 1, \dots, N_*$, заданными с некоторой погрешностью. Требуется восстановить кривую \mathbf{V} -сплайном.

Значение погрешности не известно. Поэтому компромисс между точностью восстановления и поведением сплайновой кривой достигается в процессе решения задачи с участием пользователя. Укажем два алгоритма, реализуемых в виде интерактивных процедур (в режиме диалога).

1. Вводится сетка Δ_T : $T_1 < T_2 < \dots < T_N$, ($N \ll N_*$), включающая характерные точки кривой, в окрестности которых происходит быстрое изменение кривизны, граничные точки и т. п. На отрезках $[T_{i-1}, T_{i+1}]$ строятся вектор-функции $\mathbf{P}_2^i(t)$, $i = 2, \dots, N-1$ по методу наименьших квадратов (гл. I, п. 1). Если отклонения $|\mathbf{r}_v^0 - \mathbf{P}_2^i(t_v)|$ недопустимо велики, то в сетку Δ_T вводят дополнительные узлы, и процедура повторяется. Значения вектор-функций $\mathbf{P}_2^2(T_1)$, $\mathbf{P}_2^i(T_i)$, $\mathbf{P}_2^{N-1}(T_N)$ принимают в качестве исходных данных для построения локально-аппроксимационного сплайна.

2. Сглаживающий эрмитов \mathbf{V} -сплайн строится на сетке Δ_T . Его коэффициенты \mathbf{V}_i , \mathbf{V}'_i , $i = 1, \dots, N$, в представлении (114) находят по методу наименьших квадратов с регуляризацией;

$$J(\mathbf{V}, \mathbf{V}') = \sum_{v=1}^{N_*} q_v [\mathbf{V}(t_v) - \mathbf{r}_v^0]^2 + \sum_{i=2}^{N_*-1} \beta_i [\mathbf{V}''(T_i + 0) - \mathbf{V}''(T_i - 0)]^2.$$

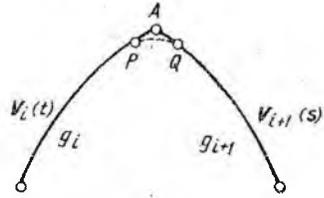
Варьируя число и расположение узлов сетки Δ_T и значения весовых коэффициентов q_v , β_i , добиваемся хорошего приближения при плавном характере сплайновой кривой. Отметим, что коэффициенты β_i влияют на плавность кривой. При $\beta_i \rightarrow \infty$ эрмитов \mathbf{V} -сплайн будет стремиться к сплайну класса C^2 .

Следующая задача является обобщением задач 6 и 7.

Задача 8. Кривая g приближенно представлена в дискретно-точечной форме значениями \mathbf{r}_i^0 , $i = 1, \dots, N$. Требуется восстановить ее \mathbf{V} -сплайном со сглаживанием.

Эта задача сводится к предыдущим, если заданным точкам \mathbf{r}_i^0 поставить в соответствие значения t_i^0 некоторой переменной t так, чтобы $t_1^0 < t_2^0 < \dots < t_N^0$. В качестве такой переменной можно использовать текущую длину ломаной с вершинами в точках \mathbf{r}_i^0 .

Рис. 18. Моделирование кусочно-гладких кривых V-сплайном



После получения V-сплайна значения t_i^0 можно уточнить, потребовав выполнения условий

$$|V(\bar{t}_i) - r_i^0| = \min_t |V(t) - r_i^0|,$$

$$i = 2, 3, \dots, N-1,$$

т. е. исходной точке r_i^0 приписывается значение параметра \bar{t}_i ближайшей к ней точки V-сплайна. Для полученных \bar{t}_i вновь определяется V-сплайн как решение предыдущей задачи и по нему подсчитывают новые значения параметра t , соответствующие точкам r_i^0 , и т. д. Этот процесс следует закончить, как только выполнится условие $(J^{(k)} - J^{(k+1)})/J^{(k)} \leq \bar{\epsilon}$, где $0 < \bar{\epsilon} < 1$.

Моделирование кусочно-гладких кривых. Рассмотрим практический прием, с помощью которого можно описывать V-сплайнами кривые, имеющие конечное число точек излома.

Пусть непрерывная кривая g состоит из M гладких дуг g_h . Каждую из них независимо от соседних приближаем V-сплайном класса C^1 или C^2 , применив тот или иной изложенный выше способ. Рассмотрим V-сплайны $V_i(t)$, $t \in [a, b]$ и $V_{i+1}(s)$, $s \in [c, d]$, соответствующие дугам g_i и g_{i+1} . Они имеют общую точку A : $V_i(b) = V_{i+1}(c)$ (рис. 18). Определим значения P и Q на расстоянии ϵ от точки A : $P = V_i(b - \delta t)$, $Q = V_{i+1}(c + \delta s)$, и построим звено V-сплайна, гладко сопрягающееся с $V_i(t)$ и $V_{i+1}(s)$ в этих точках. Ввиду малости ϵ значения δt и δs можно взять равными $\delta t = \epsilon/|V_i'(b)|$; $\delta s = \epsilon/|V_{i+1}'(c)|$. Это звено определим следующим образом:

$$V(t) = V_P \varphi_1(u) + V_Q \varphi_2(u) + \Delta t (V_P' \varphi_3(u) + V_Q' \varphi_4(u)), \quad (128)$$

где $V_P = V_i(b - \delta t) = V_i(t_P)$; $V_P' = V_i'(t_P)$;

$$V_Q = V_{i+1}(c + \delta s) = V_{i+1}(s_Q); \quad V_Q' = \alpha \langle V_{i+1}'(s_Q) \rangle,$$

$$\alpha = |V_P'|, \quad \Delta t = |V_P - V_Q|/|V_P'|.$$

В точке Q значение параметра будет $t_Q = t_P + \Delta t$. Введем вместо параметра s параметр t так, чтобы $t(s_Q) = t_Q$ и $V_{i+1}(t_Q) = V_{i+1}(s_Q)$. Эти условия будут выполнены, если положить $t = (s - s_Q)|V_{i+1}'(s_Q)|/|V_Q'| + t_Q$.

В результате всех этих преобразований мы получили гладкий V-сплайн, который совпадает с $V_i(t)$ и $V_{i+1}(t)$ на интервалах $[a, t_P]$ и $[t_Q, t(d)]$, а для $t \in [t_P, t_Q]$ он определен формулой (128). Выполнив эту процедуру последовательно для всех пар состыкованных с изломом дуг (g_h, g_{h+1}) , получим V-сплайн класса C^1 , приближающий кусочно-гладкую кривую g с точностью ϵ .

6. АППРОКСИМАЦИЯ ПОВЕРХНОСТЕЙ W-СПЛАЙНАМИ

Под аппроксимацией поверхности G W -сплайном будем понимать приближение какой-либо ее параметризации $\mathbf{r} = \mathbf{r}(t, s)$, $t, s \in \Omega$. Таким образом, задачи о приближении поверхностей сводятся к приближению вектор-функций W -сплайнами.

Приближенное описание поверхности W -сплайном. Пусть G — поверхность, заданная параметризацией $\mathbf{r} = \mathbf{r}(t, s)$, $t, s \in \Omega$. Требуется определить W -сплайн (эрмитов или класса $C^{2,2}$) приближающий эту поверхность с точностью ε .

Решение задачи будем искать среди W -сплайнов, интерполирующих вектор-функцию $\mathbf{r}(t, s)$. Тогда все сводится к определению в области Ω сетки Δ , которая обеспечивает заданную точность интерполяции. С этой целью можно воспользоваться рекомендациями и алгоритмами, изложенными в гл. I, п. 5, 6, при рассмотрении интерполяции с заданной точностью кубическим сплайном двух переменных.

Так, на основании оценок погрешности интерполяции вектор-функции W -сплайном (аналогичных, например, (42а) и (42б) при эрмитовой интерполяции) для заданного ε можно рассчитать допустимый шаг по каждой из переменных t и s (см. соответственно (43)). Для этого достаточно знать хотя бы мажоранты A и B старших производных:

$$\|D^{(p,0)}\mathbf{r}(t, s)\| \leq A; \|D^{(0,p)}\mathbf{r}(t, s)\| \leq B, \quad p=2 \text{ или } p=4.$$

От их точности зависят значения шагов Δt и Δs и, следовательно, размер всей сетки Δ .

Чтобы найти сетку по возможности минимального размера, следует применить алгоритм расчета неравномерной сетки. Здесь также требуется иметь оценки старших производных, но точность этих оценок определяет только время, затрачиваемое на решение задачи, и не влияет на размер получаемой сетки.

Восстановление поверхности W -сплайном. Задача 1. Поверхность G представлена в дискретно-точечной форме значениями одной из ее параметризаций $\mathbf{r} = \mathbf{r}(t, s)$ в узлах сетки $\Delta = \Delta_t \times \Delta_s$: $\mathbf{r}_{ij}, \mathbf{r}_{ij}^{(1,0)}, \mathbf{r}_{ij}^{(0,1)}, \mathbf{r}_{ij}^{(1,1)}$, $i = 1, \dots, N$; $j = 1, \dots, M$. Требуется восстановить поверхность G по этим данным W -сплайном.

Решение в виде эрмитова W -сплайна получим, приняв

$$W^{(p,q)}(t_i, s_j) = \mathbf{r}_{ij}^{(p,q)}, \quad p, q = 0, 1$$

и подставив эти значения в выражение для W -сплайна (118).

Если известны оценки производных $D^{(2,0)}\mathbf{r}(t, s)$ и $D^{(0,2)}\mathbf{r}(t, s)$ о точности восстановления поверхности, можно судить как о погрешности приближения вектор-функции $\mathbf{r} = \mathbf{r}(t, s)$ W -сплайном, воспользовавшись для этого соотношениями, аналогичными (42).

Задача 2. Поверхность G задана в дискретно-точечной форме значениями $\mathbf{r}_{ij}, \boldsymbol{\tau}_{ij}^{(t)}, \boldsymbol{\tau}_{ij}^{(s)}$ в узлах сетки $\Delta = \Delta_t \times \Delta_s$, где $\boldsymbol{\tau}_{ij}^{(t)}, \boldsymbol{\tau}_{ij}^{(s)}$ — единичные касательные векторы к координатным линиям $s = s_j$,

и $t = t_i$ на поверхности G в точке r_{ij} . Требуется восстановить поверхность W -сплайном.

Воспользовавшись алгоритмами решения задачи 2 из п. 5 восстанавливаем V -сплайном каждую из координатных линий $s = s_j$, $j = 1, \dots, M$ и $t = t_i$, $i = 1, \dots, N$. В результате в узлах сетки Δ получим приближенные значения $r_{ij}^{(1,0)}$, $r_{ij}^{(0,1)}$. Чтобы эту задачу окончательно свести к предыдущей, достаточно подсчитать значения смешанной производной $r_{ij}^{(1,1)} = r^{(1,1)}(t_i, s_j)$. Для этого восстановим вектор-функцию $r_i^*(s) = D^{(0,1)} r(t_i, s)$, $i = 1, \dots, N$ (или $r_j^*(t) = D^{(1,0)} r(t, s_j)$, $j = 1, \dots, M$) V -сплайнами по значениям $r_i^*(s_j)$ (или $r_j^*(t_i)$) на сетке Δ_s (или Δ_t).

Производные от этих V -сплайнов в узлах сетки, на которой они получены, принимаются в качестве приближенных значений смешанной производной от параметризации $r(t, s)$ поверхности G . При этом следует использовать ту из вектор-функций $r_i^*(s)$ и $r_j^*(t)$, относительно параметра которой вектор нормали изменяется меньше.

З а д а ч а 3. Поверхность G задана в дискретно-точечной форме значениями r_{ij} , $i = 1, \dots, N$; $j = 1, \dots, M$; $\tau_{i_l j_h}^{(t)}$, $\tau_{i_l j_l}^{(s)}$, $n_{i_m j_m}$, $i_h, i_l, i_m \in \{1, \dots, N\}$; $j_h, j_l, j_m \in \{1, \dots, M\}$, относящимися к узлам ее некоторого каркаса. (Здесь и в дальнейшем n_{ij} — единичный вектор положительного направления нормали к поверхности в точке r_{ij}). Требуется восстановить поверхность W -сплайном.

Не представляется сложным свести эту задачу к предыдущей. Сначала восстанавливаем каждую каркасную линию обоих семейств V -сплайнами $V_j(t)$ и $V_i(s)$, воспользовавшись алгоритмом решения задачи 3 из п. 5, и на этих V -сплайнах определяем приближенные значения недостающих векторов $\tau^{(t)}$ и $\tau^{(s)}$. Затем, чтобы обеспечить принадлежность векторов $\tau_{i_m j_m}^{(t)}$ и $\tau_{i_m j_m}^{(s)}$ заданной касательной плоскости в точках $r_{i_m j_m}$ поверхности G , преобразуем их по формуле

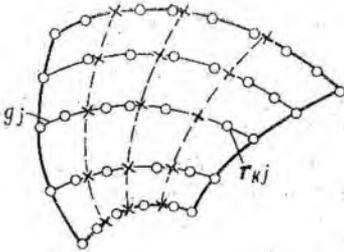
$$\tau_{i_m j_m} = \langle n_{i_m j_m} \times \tau_{i_m j_m} \times n_{i_m j_m} \rangle.$$

Окончательно из сеток $(\Delta_t)_j$ и $(\Delta_s)_i$, на которых определены были V -сплайны, восстанавливающие каркасные кривые, выбираем любые две сетки Δ_t и Δ_s . Таким образом приходим к условиям задачи 2.

З а д а ч а 4. Поверхность G задана в дискретно-точечной форме значениями r_{ij} , $i = 1, \dots, N$; $j = 1, \dots, M$, ее некоторой параметризации $r(t, s)$, $t, s \in \Omega$, в узлах сетки $\Delta = \Delta_t \times \Delta_s$. Требуется восстановить поверхность G W -сплайном.

Решение этой задачи можно получить в виде W -сплайна класса $C^{2,2}$, интерполирующего вектор-функцию $r(t, s)$ на сетке Δ по значениям r_{ij} , если дополнительно задать граничные условия типа I—IV на границах области Ω . На основании оценок интерполяции W -сплайном можно судить о погрешности восстановления поверхности G . Для этого необходимо иметь оценки старших производных $D^{(p,0)} r(t, s)$ и $D^{(0,p)} r(t, s)$, $p = 2$ или 4.

Рис. 19. Задание поверхности каркасными линиями одного семейства



Чтобы получить решение в виде эрмитова W -сплайна, достаточно приближенно подсчитать значения касательных векторов $\tau_{ij}^{(t)}$ и $\tau_{ij}^{(s)}$ к координатным линиям $s = s_j$ и $t = t_i$ в точках r_{ij} . В результате мы придем к условиям задачи 2.

Задача 5. Поверхность G задана в дискретно-точечной форме узлами r_{ij} , $i = 1, \dots, N$; $j = 1, \dots, M$ ее некоторого каркаса. Требуется восстановить поверхность G W -сплайном.

Если заданным точкам поставить в соответствие значения t_i, s_j из прямоугольной области Ω : $a \leq t \leq b$, $c \leq s \leq d$ вещественных переменных t и s , то задача будет сведена к предыдущей. Для этого выберем две каркасные линии с номерами i^* и j^* из разных семейств. Они представлены в дискретно-точечной форме значениями r_{ij^*} и r_{i^*j} . Принимая в качестве переменных t и s текущую длину вписанной ломаной с вершинами в заданных точках, для этих линий подсчитаем сетки Δt и Δs . Следовательно, получим сетку $\Delta = \Delta_t \times \Delta_s$ в области Ω .

Задача 6. Известна последовательность каркасных линий g_j , $j = 1, \dots, M$, одного семейства (например 1-го) поверхности G , каждая из которых представлена в дискретно-точечной форме значениями r_{kj} , $k = 1, \dots, N_j$. Требуется восстановить поверхность G W -сплайном.

Если на заданных кривых g_j указать «следы» каркасных линий второго семейства (на рис. 19 отмечены крестиком), то придем к условиям предыдущей задачи. Приведем несколько наиболее общих приемов, позволяющих это сделать.

1. Предположим, что удастся указать последовательность поверхностей G_i , $i = 1, \dots, N$ таких, что линии пересечения их с поверхностью G можно принять в качестве каркасных линий второго семейства. Тогда восстанавливаем исходные линии g_j V -сплайнами и определяем их точки пересечения v_{ij} , $i = 1, \dots, N_j$; $j = 1, \dots, M$ с поверхностями G_i . Последовательность $\{G_i\}$ необходимо выбирать с учетом допустимой погрешности восстановления поверхности G .

2. Восстанавливаем каждую каркасную линию g_j V -сплайном $V_j(t)$, $t \in [a, b]$. В качестве переменной удобно использовать нормированную текущую длину вдоль вписанной ломаной. Ее значения в заданных точках r_{hj} кривой g_j определяются следующим образом:

$$t_1 = 0, \quad t_k = \sum_{l=1}^{k-1} |r_{l+1,j} - r_{lj}| L_j^{-1}, \quad k = 2, \dots, N_j,$$

где L_j — длина всей ломаной. При таком определении переменной t область ее изменения есть отрезок $[0, 1]$. Выберем на этом отрезке сетку Δ_t . Последовательность значений $v_{ij} = V_j(t_i)$, $i = 1, \dots,$

..., N ; $j = 1, \dots, M$, в узлах этой сетки принимаем в качестве приближения узлов каркаса поверхности G .

Каждой исходной кривой g_j соответствует свое разбиение $\Delta_t^{(j)}$ отрезка $[0, 1]$. В качестве общей сетки Δ_t можно взять сетку наиболее сложной по характеру линии. Как правило, такая линия задается максимальным числом точек, что позволяет легко выбрать ее, не проводя дополнительного анализа. В некоторых случаях эта линия может быть указана заранее.

3. Может оказаться, что известны несколько каркасных линий Π семейства и их «следы» присутствуют в исходных точках. Эти линии разбивают поверхность G на сегменты, к каждому из которых можно применить приемы 1 или 2.

З а д а ч а 7. Известны линии g_l , $l = 1, \dots, p$, лежащие на поверхности G , каждая из которых задана в дискретно-точечной форме значениями r_{kl} , $k = 1, \dots, N_l$. Требуется восстановить поверхность W -сплайном (рис. 20).

В отличие от условий предыдущей задачи в рассматриваемой задаче указанные линии не могут быть приняты в качестве каркасных. Если удастся указать некоторый набор поверхностей G_j , $j = 1, \dots, M$, таких, что линии их пересечения с поверхностью G можно принять в качестве каркасных линий одного из семейств, рассматриваемая задача сводится к предыдущей. Для этого каждая из линий g_l приближается V -сплайном и ищутся точки его пересечения с поверхностями G_j . Полученные в результате точки группируются по принадлежности поверхностям G_j , а затем упорядочиваются в соответствии с выбранными на искомым каркасных линиях направлением. Массив значений v_{ij} , $i = 1, \dots, N_j$; $j = 1, \dots, M$ принимается в качестве дискретно-точечного приближения каркасных линий одного семейства на поверхности G .

Восстановление линейчатой поверхности. Ограниченный сегмент линейчатой поверхности можно представить в виде

$$r(t, s) = \rho_1(t)(1-s) + \rho_2(t)s, \quad t \in [a, b], s \in [0, 1],$$

где $r = \rho_1(t)$ и $r = \rho_2(t)$ — параметризации крайних дуг g_1 и g_2 сегмента. Предполагается, что каждая из этих дуг задана в дискретно-точечной форме значениями, по которым они могут быть восстановлены V -сплайнами $V_1(t)$ и $V_2(t)$ на некоторой сетке Δ_t . Кроме того, пары исходных точек ρ_{1i}, ρ_{2i} , $i = 1, \dots, N$ принадлежат одной образующей.

Тогда линейчатая поверхность восстанавливается W -сплайном частного вида

$$W(t, s) = V_1(t)(1-s) + V_2(t)s.$$

Допускаемая при этом погрешность не превышает наибольшей погрешности восстановления крайних дуг g_1 и g_2 .

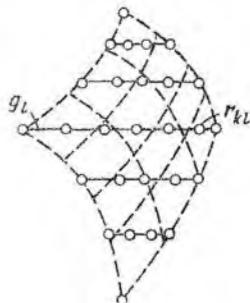


Рис. 20. Задание поверхности совокупностью линий

Восстановление поверхности вращения. В этом случае достаточно восстановить образующую. Допускаемая при этом погрешность не превышает погрешности восстановления образующей.

Восстановление поверхности со сглаживанием исходных данных. Эта задача существенно более трудная, чем подобная задача для кривых. Сложность проявляется и в ее формальной постановке и тем более в численной реализации решения. Не претендуя на исчерпывающую полноту решения проблемы, рассмотрим несколько возможных постановок.

Задача 8. Поверхность G задана в дискретно-точечной форме значениями $r_{ij}^0, t_i, s_j, i = 1, \dots, N, j = 1, \dots, M$, относительно которых предполагается, что если $r(t, s), t, s \in \Omega$ — некоторая параметризация поверхности G , то $|r(t_i, s_j) - r_{ij}^0| \leq \delta_{ij}$. По этим значениям требуется восстановить поверхность W -сплайном.

Решение в виде W -сплайна класса $C^{2,2}$, определенного на заданной сетке $\Delta = \Delta_t \times \Delta_s$, можно получить, если воспользоваться алгоритмом локального сглаживания функции двух переменных (см. гл. I, п. 9), применив его к компонентам W -сплайна $W_x(t, s), W_y(t, s)$ и $W_z(t, s)$. При этом ограничения на величину отклонений W -сплайна от заданных точек следует взять в виде

$$\begin{aligned} |W_x(t_i, s_j) - x_{ij}^0| &\leq \delta_{ij}^{(x)}; & |W_y(t_i, s_j) - y_{ij}^0| &\leq \delta_{ij}^{(y)}; \\ |W_z(t_i, s_j) - z_{ij}^0| &\leq \delta_{ij}^{(z)}, \end{aligned}$$

где $(\delta_{ij}^{(x)})^2 + (\delta_{ij}^{(y)})^2 + (\delta_{ij}^{(z)})^2 = \delta_{ij}^2$.

Как и в случае кривой, можно принять $\delta_{ij}^{(x)} = \delta_{ij}^{(y)} = \delta_{ij}^{(z)} = \delta_{ij} / \sqrt{3}$. Часто на практике для решения подобной задачи используют следующий прием. Последовательно сглаживаются данные вдоль координатных линий: сначала $s = s_j, j = 1, \dots, M$, а затем полученные значения — вдоль линий $t = t_i, i = 1, \dots, N$. Известно [8], что при определенной постановке одномерных задач такая процедура равносильна нахождению W -сплайна, компоненты которого $W_x(t, s), W_y(t, s)$ и $W_z(t, s)$ минимизируют некоторый функционал.

Задача 9. Поверхность G задана в дискретно-точечной форме значениями $r_k, t_k, s_k, k = 1, \dots, K$, где (t_k, s_k) — точки в прямоугольной области Ω вещественных переменных t и s . При этом предполагается, что значения r_k принадлежат некоторой δ -окрестности поверхности G . Требуется восстановить поверхность W -сплайном.

Так как допустимое отклонение δ_k заданных точек r_k от искомого W -сплайна не определено, представляется целесообразным искать решение последовательными приближениями, привлекая проектировщика после каждого шага к оценке результата.

В целом алгоритм поиска решения выглядит следующим образом. В области Ω выбираем сетку $\Delta = \Delta_t \times \Delta_s$ с учетом особенностей поверхности G и по возможности минимального размера. Восстанавливающий W -сплайн ищем в представлении через B -сплайны:

$$W(t, s) = \sum_{i=0}^{N+1} \sum_{j=0}^{M+1} b_{ij} B_i(t) \bar{B}_j(s)$$

из условия минимума функционала:

$$J(\mathbf{b}_{ij}) = \sum_{k=1}^K q_k [W(t_k, s_k) - r_k]^2, \quad q_k > 0.$$

Определив W -сплайн, подсчитываем его отклонения от заданных точек r_k . После анализа полученного результата вносим коррективы в значения весов q_k и в сетку Δ , изменяя число и расположение ее узлов. После этого процедура повторяется. Сетку Δ следует выбирать и изменять так, чтобы выполнялось условие $(N+2)(M+2) \ll K$ и исходные точки достаточно равномерно были распределены по клеткам и внутри них. В противном случае система для определения коэффициентов b_{ij} может оказаться неустойчивой к погрешностям округления или даже вырожденной.

Приведенный алгоритм следует применять в тех случаях, когда поверхность G хорошо описывается W -сплайном на сетке сравнительно малых размеров. Для сложных по форме поверхностей неизбежно приходится выбирать довольно густую сетку. В этом случае предпочтительнее алгоритм локального сглаживания. Рассмотрим основные его моменты. Выбираем сетку Δ так, чтобы в каждую клетку Ω_{ij} попало не менее трех исходных точек (t_k, s_j) . В каждой области $[t_{i-1}, t_{i+1}] \times [s_{j-1}, s_{j+1}]$ методом наименьших квадратов определяем векторный многочлен

$$P_2^{ij}(t, s) = \sum_{m=0}^2 \sum_{l=0}^2 c_{ml}^{ij} \Phi_m(t) \Phi_l(s).$$

Функции $\Phi_m(t)$ и $\Phi_l(s)$ приведены в гл. I, п. 2 и представляют собой многочлены второй степени. Варьируя сетку Δ и веса, добиваемся удовлетворительного приближения исходных данных многочленом $P_2^{ij}(t, s)$. В качестве приближенных значений r_{ij} поверхности в точках (t_i, s_j) принимаем $c_{11}^{ij} = P_2^{ij}(t_i, s_j)$ для внутренних узлов сетки и соответствующие коэффициенты для граничных узлов. И к ним применяем процедуру локальной аппроксимации, изложенную в гл. I, п. 9.

В обоих предлагаемых алгоритмах используется метод наименьших квадратов. Если условия его устойчивости не выполняются, то следует применять модифицированный метод с регуляризацией (гл. I, п. 1).

Задача 10. Поверхность G задана в дискретно-точечной форме значениями r_k , $k = 1, \dots, K$, принадлежащими ее некоторой δ -окрестности. Требуется восстановить поверхность W -сплайном.

Имеет смысл рассмотреть следующие приемы, сводящие эту задачу к предыдущим.

1. Предположим, что исходные точки можно отнести к множеству $\{g_j\}$ некоторых линий на поверхности G . Восстанавливая каждую из них со сглаживанием V -сплайнами $V_j(t)$, $t \in [a_j, b_j]$, приходим к условиям одной из задач 6 или 7 о восстановлении поверхности W -сплайном. Если отклонения полученного W -сплайна от исходных точек r_k окажутся неприемлемо велики или сама

сплайновая поверхность недостаточно плавной, то имеет смысл подвергнуть сглаживанию значения $W_{ij} = W(t_i, s_j)$ в узлах некоторой сетки Δ , т. е. решить задачу 8.

2. Предположим, что удается подобрать поверхность \tilde{G} : $r = R(t, s)$, $t, s \in \Omega$, между точками которой и точками поверхности G устанавливается взаимно однозначное соответствие. Тогда, присваивая исходным точкам r_k значения параметров t_k, s_k , соответствующих точкам R_k поверхности \tilde{G} , приходим к условиям задачи 9. Отметим, что поверхность \tilde{G} и ее параметризацию $R(t, s)$ следует выбирать такими, чтобы края восстанавливаемой поверхности G соответствовали координатные линии параметризации $R(t, s)$.

В некоторых простых случаях вспомогательную поверхность \tilde{G} можно определить следующим образом. Из множества $\{r_k\}$ выбираем точки, соответствующие крайним дугам g_1, g_2, g_3, g_4 поверхности G . Восстанавливаем эти дуги V -сплайнами $V_1(t), V_2(t), V_3(s), V_4(s)$, $t \in [0, 1], s \in [0, 1]$, естественно полагая

$$\begin{aligned} V_1(0) = V_3(0) &\equiv r_{00}; & V_2(0) = V_3(1) &\equiv r_{01}; & V_1(1) = V_4(0) &\equiv r_{10}; \\ V_2(1) = V_4(1) &\equiv r_{11}, \end{aligned}$$

где $r_{00}, r_{01}, r_{10}, r_{11}$ — угловые точки поверхности G . Поверхность \tilde{G} определим в виде

$$R(t, s) = -[-1, 1 - t, t] \begin{bmatrix} 0 & V_1(t) & V_2(t) \\ V_3(s) & r_{00} & r_{01} \\ V_4(s) & r_{10} & r_{11} \end{bmatrix} \begin{bmatrix} -1 \\ 1 - s \\ s \end{bmatrix}.$$

Очевидно, что $R(t, 0) = V_1(t)$; $R(t, 1) = V_2(t)$; $R(0, s) = V_3(s)$; $R(1, s) = V_4(s)$. Теперь каждой точке r_k ставим в соответствие ближайшую точку поверхности \tilde{G} , решая уравнения

$$(R(t_k, s_k) - r_k) R_t(t_k, s_k) = 0; \quad (R(t_k, s_k) - r_k) R_s(t_k, s_k) = 0$$

относительно неизвестных t_k, s_k .

О генерировании кривых и поверхностей. В практике проектирования объектов сложной формы приходится решать разнообразные задачи построения кривых и поверхностей с заданными свойствами. Требования, предъявляемые к кривой или поверхности как к элементу поверхности создаваемого объекта, очерчивают конструктору или дизайнеру довольно широкую область для поиска. Некоторые из этих требований, например требования эстетики, формализовать нельзя.

Из множества задач проектирования кривых и поверхностей трудно выделить какие-либо общие задачи, не зависящие от конкретной предметной области. Поэтому отметим только принципиальную возможность применения сплайнов вообще и кубических вектор-сплайнов в частности для решения подобных задач. Действительно, если существует кривая или поверхность, удовлетворяющая заданным требованиям, то эти геометрические объекты могут быть описаны сплайном сколь угодно точно.

Другая особенность сплайнов, обеспечивающая их эффективное применение для генерирования кривых и поверхностей, состоит в том, что изменение коэффициента сплайна ведет к локальному изменению объекта. Для эрмитовых вектор-сплайнов каждый коэффициент влияет только на два соседних звена. Изменение коэффициента кубического вектор-сплайна класса C^3 , представленного через B -сплайны, вызывает изменение четырех соседних звеньев. Окрестность изменения сплайна при варьировании одного из его коэффициентов (соответствующего, например, узлу t_i сетки Δ) можно сделать как угодно малой, если к сетке добавить дополнительные узлы в соответствующие окрестности узла t_i : по одному справа и слева от t_i для эрмитова сплайна, и по два — для B -сплайна.

Процедуры поиска решений задач проектирования кривых и поверхностей представляют собой процесс последовательного приближения и реализуются, как правило, в интерактивном режиме — режиме диалога. На каждом шаге результат представляется в виде графического изображения на дисплее или графопостроителе и сопровождается необходимой числовой информацией. Пользователь анализирует этот результат, изменяет соответствующие параметры и передает управление ЭВМ для выполнения очередного шага.

Нередко бывает задано начальное приближение искомого инженерного решения, в роли которого обычно выступает известный прототип. Тогда первый и необходимый шаг — построение математической модели прототипа и ее анализ. Затем, используя установленные свойства сплайнов, математическую модель прототипа модифицируют в направлении искомого решения.

МОДЕЛИРОВАНИЕ КРИВЫХ И ПОВЕРХНОСТЕЙ НЕЛИНЕЙНЫМИ СПЛАЙНАМИ

В главе вводится один тип непolynomialных сплайнов. На основе одномерных сплайнов этого типа осуществляется аппроксимация кривых и поверхностей. Для представления результатов аппроксимации используется понятие конечной ε -сети.

1. ОКРУЖНОСТНАЯ ИНТЕРПОЛЯЦИЯ ПЛОСКОЙ КРИВОЙ

Задача интерполяции. На плоскости имеется гладкая простая дуга с кусочно-непрерывной ограниченной кривизной $k(\sigma) \leq K$. Пусть на дуге задана декартовыми координатами (X, Y) упорядоченная по длине дуги конечная система узлов, разбивающая эту дугу на участки с длинами, не превышающими положительного числа S такого, что выполняется условие

$$KS \leq \theta < \pi/2. \quad (129)$$

Дугу с системой узлов будем считать интерполируемой кривой.

Введем на участке между i -м и $(i+1)$ -м узлами местную систему декартовых координат (x, y) с началом в i -м узле и осью абсцисс, направленной по хорде этого участка. Узлы $(i-1)$, i , $(i+1)$ единственным образом определяют окружность (в частном случае — прямую), узлы i , $(i+1)$, $(i+2)$ определяют вторую окружность (рис. 21). Меньшие дуги этих окружностей, содержащие по три узла каждая, будем называть *интерполирующими дугами* (участок интерполируемой кривой показан на рис. 21 точечной линией). Пусть $y = g_1(x)$, $y = g_2(x)$ — уравнения кусков интерполирующих дуг на отрезке $[x_i, x_{i+1}]$. Введем на этом отрезке линейные весовые функции

$$P_1(x) = \frac{x_{i+1} - x}{x_{i+1} - x_i};$$

$$P_2(x) = \frac{x - x_i}{x_{i+1} - x_i}.$$

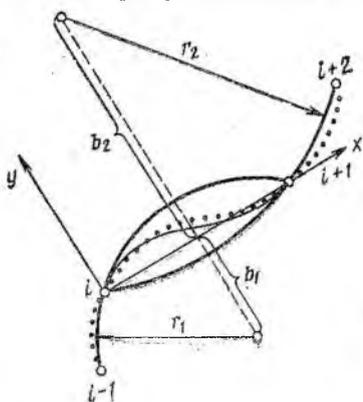


Рис. 21. Построение окружностной интерполирующей кривой

которые в сумме тождественно равны единице, а на концах отрезка принимают значения $\mathcal{P}_1(x_i) = \mathcal{P}_2(x_{i+1}) = 1$; $\mathcal{P}_1(x_{i+1}) = \mathcal{P}_2(x_i) = 0$.

Возьмем функцию

$$y = \mathcal{P}_1(x) g_1(x) + \mathcal{P}_2(x) g_2(x), \quad (130)$$

или в развернутом виде

$$y(x) = \left(\frac{x_{i+1} - x}{x_{i+1} - x_i} \right) \left[b_1 + \text{sign}(y_i - b_1) \sqrt{r_1^2 - \left(x - \frac{x_i + x_{i+1}}{2} \right)^2} \right] + \\ + \left(\frac{x - x_i}{x_{i+1} - x_i} \right) \left[b_2 + \text{sign}(y_i - b_2) \sqrt{r_2^2 - \left(x - \frac{x_i + x_{i+1}}{2} \right)^2} \right],$$

где b_1, b_2 — ординаты центров; r_1, r_2 — радиусы первой и второй окружностей. График этой функции примем в качестве кривой, интерполирующей участок между i -м и $(i+1)$ -м узлами. Поскольку на концах участка для производных весовых функций справедливы соотношения $\mathcal{P}'_1(x_i) = -\mathcal{P}'_2(x_i)$; $\mathcal{P}'_1(x_{i+1}) = -\mathcal{P}'_2(x_{i+1})$, касательная интерполирующей кривой в i -м узле совпадает с касательной первой окружности, а в $(i+1)$ -м узле — с касательной второй окружности. Каждая окружность выполняет роль первой на последующем участке и роль второй — на предшествующем. Касательная интерполирующей кривой не имеет скачков при переходе через узлы, и интерполирующая кривая, построенная по формуле (130) на всех участках, является гладкой.

Построенную таким способом *окружностную интерполяционную кривую* можно рассматривать как вектор-сплайн, где роль параметра t выполняет текущая длина вдоль ломаной с вершинами в заданных узлах.

Действительно, обозначим φ_i угол наклона хорды участка $[i, i+1]$ к оси OX общей системы координат (X, Y) . Координаты точки участка $[i, i+1]$ интерполирующей кривой в общей системе можно выразить через местные координаты в виде

$$X^i = x^i \cos \varphi_i - y^i \sin \varphi_i; \quad Y^i = x^i \sin \varphi_i + y^i \cos \varphi_i.$$

Местная абсцисса $x = t - t_i$, где t_i — длина ломаной между первым и i -м узлами. Отсюда с учетом (130) имеем выражение общих координат точки интерполирующей кривой через параметр t и константы, определяемые расположением узлов:

$$X^i(t) = (t - t_i) \cos \varphi_i - [\mathcal{P}_1(t - t_i) g_1(t - t_i) + \mathcal{P}_2(t - t_i) g_2(t - t_i)] \sin \varphi_i; \\ Y^i(t) = (t - t_i) \sin \varphi_i + [\mathcal{P}_1(t - t_i) g_1(t - t_i) + \mathcal{P}_2(t - t_i) g_2(t - t_i)] \cos \varphi_i.$$

Окружностная интерполяция интересна тем, что с нее началось широкое использование вектор-сплайнов в машинных вычислениях. Кроме того, любопытен геометрический подход к оценкам погрешности интерполяции. Если для V -сплайнов (гл. II) они определялись через производные приближаемой вектор-функции, то здесь более естественно это делать через кривизну.

Сравнение кривых через их кривизны. Рассмотрим соотношение, связывающее кривизну $k(x)$ графика функции $y = f(x)$ с производными этой функции в декартовых координатах (x, y) :

$$f''(x)/[1 + (f'(x))^2]^{3/2} = k(x). \quad (131)$$

При заданной правой части это равенство является нелинейным дифференциальным уравнением 2-го порядка относительно неизвестной функции $f(x)$. Поставим для уравнения (131) с ограниченной функцией $k(x)$ следующую краевую задачу.

Найти на отрезке $[0, 1]$ непрерывную, непрерывно дифференцируемую функцию $y = f(x)$, удовлетворяющую уравнению (131) всюду, кроме точек разрыва $k(x)$, и принимающую на концах отрезка заданные значения $y_0 = f(0)$, $y_1 = f(1)$.

Решение этой задачи представимо в виде

$$f(x) = y_0 + \int_0^x \frac{[B(z) + C] dz}{\sqrt{1 - [B(z) + C]^2}}, \quad B(x) = \int_0^x k(\xi) d\xi, \quad (132)$$

где константа C определяется из условия на правом конце отрезка. Необходимое условие существования решения состоит в том, что колебание функции $B(x)$ на отрезке $[0, 1]$ меньше двух: $\omega(B) < 2$. Достаточное условие имеет вид $\omega(B) < 1 - |y_0 - y_1|/\sqrt{1 - (y_0 - y_1)^2}$.

На основании этих результатов может быть доказана теорема о сравнении функций через кривизны их графиков: *если графики двух различных непрерывно дифференцируемых функций $f_1(x)$ и $f_2(x)$ совпадают на концах отрезка, а кривизны этих графиков удовлетворяют на всем отрезке неравенству $k_2(x) \leq k_1(x)$, то всюду внутри отрезка $f_1(x) \leq f_2(x)$.*

Поскольку правая часть уравнения (131) входит в представление (132) только под знаком интеграла, утверждение теоремы о сравнении справедливо для весьма широкого класса интегрируемых кривизн $k(x)$.

Погрешность окружностной интерполяции. Минимальное и максимальное значения кривизны интерполируемой кривой на участке между $(i - 1)$ -м и $(i + 2)$ -м узлами обозначим соответственно k_{\min} и k_{\max} . Дугу окружности с кривизной k_{\min} , проходящую через i -й и $(i + 1)$ -й узлы, назовем минимальной дугой, аналогичную дугу с кривизной k_{\max} назовем максимальной дугой. Через k_1 и k_2 обозначим кривизны первой и второй интерполирующих окружностей на участке $[i, i + 1]$.

Дугу первой интерполирующей окружности и участок интерполируемой кривой между $(i - 1)$ -м и $(i + 1)$ -м узлами по условию (129) рассматривать как графики функций $y = f(x)$ в местных координатах этого участка. Внутри участка оба графика имеют общую точку — i -й узел. Кривизна $k(x)$ интерполируемой кривой не может быть всюду больше или всюду меньше величины k_1 , так как в противном случае в силу теоремы о сравнении дуга кривой не имела бы

общих точек с первой окружностью. Поэтому кривизна первой интерполирующей окружности удовлетворяет неравенству

$$k_{\min} \leq k_1 \leq k_{\max}. \quad (133)$$

Аналогичное неравенство справедливо для кривизны второй интерполирующей окружности:

$$k_{\min} \leq k_2 \leq k_{\max}. \quad (134)$$

В местной системе координат участка $[i, i + 1]$ кусок интерполируемой кривой, минимальную и максимальную дуги можно рассматривать как графики функций соответственно $y = f(x)$; $y = r(x)$; $y = R(x)$. В силу теоремы о сравнении для всех x между x_i и x_{i+1} имеем следующую цепочку неравенств:

$$R(x) \leq f(x) \leq r(x). \quad (135)$$

Исходя из неравенств (133) и (134), на основании теоремы о сравнении можно утверждать, что обе интерполирующие дуги на участке $[i, i + 1]$ заключены между минимальной и максимальной дугами. Участок интерполирующей кривой по построению заключен между интерполирующими дугами, следовательно, имеет место неравенство $R(x) \leq g(x) \leq r(x)$.

Поскольку $R(x)$ и $r(x)$ ограничивают соответственно сверху и снизу как функцию $f(x)$, так и функцию $g(x)$, то модуль разности последних не может быть больше толщины «линзы», образованной минимальной и максимальной дугами. Если обозначить l длину хорды участка $[i, i + 1]$, то искомая оценка примет вид

$$|g(x) - f(x)| \leq \frac{1 - \sqrt{1 - (lk_{\max}/2)^2}}{k_{\max}} - \frac{1 - \sqrt{1 - (lk_{\min}/2)^2}}{k_{\min}}. \quad (136)$$

Полученная оценка носит локальный характер, так как от участка к участку входящие в правую часть величины могут изменяться.

Угловая погрешность. Расхождение направлений касательных интерполируемой и интерполирующей кривых тоже оценивается с использованием теоремы о сравнении. Опуская подробные рассуждения, приведем лишь результаты.

В местных координатах участка производная интерполируемой функции $f(x)$ оценивается неравенством

$$|f'(x)| \leq \max \left\{ \max_x |r'(x)|, \max_x |R'(x)| \right\} = c. \quad (137)$$

С другой стороны, производная линейно-весовой суммы (131) оценивается неравенством

$$|g'(x)| \leq \max \left\{ \max_x |g_1'(x)|, \max_x |g_2'(x)| \right\} = d.$$

Используя теорему о сравнении, можно показать, что $d \leq c$.

Обозначим α угол между касательным вектором интерполируемой кривой на участке $[i, i + 1]$ и хордой этого участка, β — ана-

логичный угол для интерполирующей кривой. Поскольку угол не может быть больше своего тангенса, то

$$\alpha \leq |f'(x)| \leq c, \quad \beta \leq |g'(x)| \leq d.$$

Угол γ между касательными векторами кривых f и g на участке между соседними узлами не превышает $\alpha + \beta$, т. е. оценка угловой погрешности имеет вид

$$\gamma \leq \alpha + \beta \leq c + d \leq 2c = lk_i \left[1 - \left(\frac{lk_i}{2} \right)^2 \right]^{-1/2}, \quad (138)$$

где lk_i — максимум модуля кривой $g(x)$ на участке между $(i-1)$ -м и $(i+1)$ -м узлами. Эта угловая оценка так же, как и оценка погрешности интерполяции (136), локальна. Асимптотическое свойство оценки (136) лучше, чем аналогичное свойство у оценки (138), но это будет показано ниже.

2. ОКРУЖНОСТНАЯ ИНТЕРПОЛЯЦИЯ ПРОСТРАНСТВЕННОЙ КРИВОЙ

Постановка задачи интерполяции. Рассматриваются гладкие пространственные кривые, составленные из конечного числа дважды непрерывно дифференцируемых дуг ограниченной кривизны $k \leq K < \infty$. Кривизна пространственной кривой является величиной неотрицательной. Вектор длины k , направленный вдоль главной нормали от точки кривой к центру кривизны, называется *вектором кривизны*. Для кривых рассматриваемого класса вектор кривизны непрерывно зависит от точки на каждой из составляющих дуг, а при переходе через точку стыка может испытывать скачки. Возможны поворот на конечный угол, конечное приращение длины (разрыв кривизны) либо то и другое одновременно. Допустимы также отрезки кривой, на которых кривизна равна нулю. На таких отрезках нормаль считается доопределенной постоянным вектором, совпадающим, например, с нормалью в одном из концов отрезка.

Пусть на кривой задана система узлов, разбивающая кривую на участки с длинами, не большими положительного числа S такого, что

$$\int_0^S k(\sigma) d\sigma \leq KS \leq \theta < \frac{\pi}{2}. \quad (139)$$

Здесь σ — лонгальный параметр.

В силу условия (139) дуга на каждом участке имеет однозначную проекцию на свою хорду. То же самое можно сказать о дуге на двух смежных участках.

Для дважды непрерывно дифференцируемых дуг кручение $T(\sigma) = \psi'(\sigma)$, где угол $\psi(\sigma)$ будем называть *углом скручивания*. Угол скручивания отсчитывается от положения нормали в начальной точке дуги и характеризует поворот (в положительном или отрицательном направлении) вектора главной нормали в нормальной плоскости. Если принять в начальной точке участка скручивание равным

ψ_0 , то в точке, соответствующей длине дуги σ , значение угла скручивания дает формула

$$\psi(\sigma) = \psi_0 + \int_0^{\sigma} T(\xi) d\xi + \sum_{j=1}^m \Delta\psi_j,$$

где $\Delta\psi_j$ — конечное приращение угла скручивания в точке j -го стыка дважды дифференцируемых дуг. Разность $\psi_{\max} - \psi_{\min}$ на участке кривой будем называть диапазоном скручивания этого участка и обозначать $\Delta\psi$. Наложим на выбор узлов интерполяции еще одно условие, учитывающее пространственный характер кривой: на протяжении любых трех последовательных участков между $(i-1)$ -м и $(i+1)$ -м узлами

$$\Delta\psi \leq \Psi \leq \pi. \quad (140)$$

Выбранные узлы интерполяции не обязательно должны совпадать с точками стыка дважды дифференцируемых дуг.

Окружностная интерполяция отрезка данной кривой на участке между i -м и $(i+1)$ -м узлами осуществляется по аналогии с плоским случаем. Местная декартова система координат (x, y, z) выбирается так, что начало ее совпадает с i -м узлом, ось ox направлена вдоль хорды в сторону $(i+1)$ -го узла, а остальные оси направлены произвольно. Через узлы $(i-1)$, i , $(i+1)$ проходит единственная окружность (в частном случае прямая), дугу которой между узлами i -м и $(i+1)$ -м описывает конец радиус-вектора $g_1(x)$, являющийся функцией абсциссы. Узлы i , $i+1$, $i+2$ определяют дугу окружности, описываемую радиус-вектором $g_2(x)$ на участке между i -м и $(i+1)$ -м узлами. На этом участке дуга интерполирующей кривой описывается вектор-функцией

$$g(x) = \mathcal{P}_1(x) g_1(x) + \mathcal{P}_2(x) g_2(x), \quad (141)$$

где $\mathcal{P}_1(x)$ и $\mathcal{P}_2(x)$ — линейные весовые функции на отрезке $[x_i, x_{i+1}]$. Если l — длина хорды участка между i -м и $(i+1)$ -м узлами, то $\mathcal{P}_1(x) = (l-x)/l$; $\mathcal{P}_2(x) = x/l$.

Интерполирующая кривая в целом описывается радиус-вектором относительно некоторой точки, взятой в качестве полюса $r(t) = r_i + \rho(t)$. Здесь t — параметр, роль которого выполняет текущая длина вдоль вписанной в узлы ломаной. С абсциссой на i -м участке этот параметр связан линейной зависимостью $x^i = t - t_i$.

Одному и тому же значению параметра отвечают точка данной кривой и точка интерполирующей кривой, находящиеся в одной плоскости, ортогональной хорде участка. Максимальное расстояние между такими парами точек и является мерой погрешности интерполяции на конкретном участке. Мера близости μ (см. гл. II, п. 2) для дуг данной и интерполирующей кривых на этом участке не превышает указанную погрешность.

Зададимся некоторой кусочно-непрерывной ограниченной функцией $k(\sigma)$, определенной на отрезке $[0, S]$, удовлетворяющей условию (139). Существует единственная плоская выпуклая дуга R дли-

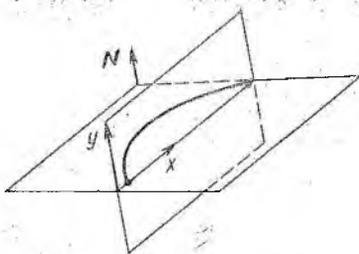


Рис. 22. Пространственная дуга, опирающаяся на плоскость

ны S , для которой $k(\sigma)$ есть кривизна. Существует бесконечное множество неконгруэнтных (неналожимых) между собой дуг с длинами S и тождественно совпадающими кривизнами $k(\sigma)$. Эти дуги отличаются друг от друга только скручиванием $\psi(\sigma)$. Для дуги R $\psi(\sigma) = 0$, поэтому будем

называть эту единственную в своем классе дугу — *раскруткой*. Класс D_{hS} кривых с длинами S и кривизнами $k(\sigma)$ порождается раскруткой в том смысле, что применением к ней различных скручиваний $\psi(\sigma)$ можно получить все кривые из этого класса.

Раскрутка R обладает несколькими экстремальными свойствами, для формулирования которых требуется ввести понятия, характеризующие дугу кривой. В силу условия (139) любая дуга из D_{hS} однозначно проецируется на свою хорду. Расстояние от точки дуги до хорды назовем высотой точки, максимальную высоту по всем точкам высотой дуги, максимальный угол наклона касательной к хорде — крутизной дуги.

Раскрутка R имеет: 1) минимальную длину хорды, 2) максимальную высоту, 3) максимальную крутизну, 4) максимальные высоты точек, соответствующих одному значению натурального параметра σ среди всех дуг порождаемого ею класса D_{hS} . Эти свойства раскрутки интуитивно ясны, но доказательства их довольно громоздки, поэтому здесь не приведены.

Теорема о пространственной дуге, опирающейся концами на плоскость. Рассмотрим горизонтальную плоскость. Пусть нормальный к ней вектор N направлен вверх и имеется дуга пространственной кривой, удовлетворяющая условию [139].

У т в е р ж д е н и е. Если концы дуги лежат в плоскости, а проекция вектора кривизны на вектор N всюду неположительна, то вся дуга (кроме концов) лежит над плоскостью.

Д о к а з а т е л ь с т в о. Спроецируем дугу на вертикальную плоскость, проходящую через хорду дуги (рис. 22). Полагая ось x направленной по хорде, а ось y направленной вверх, применим к полученной проекции и хорде теорему о сравнении. Кривизна хорды всюду равна нулю, а кривизна проекции дуги по условию доказываемой теоремы всюду неположительна, следовательно, вся проекция (кроме концов) лежит над хордой и сама кривая расположена над плоскостью.

С л е д с т в и е. Если дуга, удовлетворяющая условию (139), лежит над плоскостью, опираясь на нее только концами, то на дуге найдется по крайней мере одна точка (а в силу непрерывности кривизны — и целый интервал), где вектор кривизны имеет отрицательную проекцию на нормаль к плоскости.

Действительно, в противном случае согласно утверждению теоремы вся дуга должна была бы лежать под плоскостью.

Оценка погрешности интерполяции пространственной кривой.

Рассмотрим участок интерполируемой кривой между i -ми и $(i + 1)$ -ми узлами в проекции на плоскость, ортогональную его хорде. В силу условия (139) нормальная плоскость кривой в любой точке этого участка составляет с плоскостью, ортогональной хорде, угол не превышающий θ . Поэтому проекцию ψ любого угла Ψ в нормальной плоскости на плоскость, ортогональную хорде, можно оценить неравенством

$$\operatorname{tg}(\tilde{\psi}/2) \leq \operatorname{tg}(\psi/2)/\cos \theta. \quad (142)$$

Пусть некоторая плоскость, проходящая через хорду, сечет дугу участка кривой в точке a (рис. 23). Согласно следствию из теоремы о дуге, опирающейся на плоскость, на участке кривой найдутся точки b и c такие, что векторы кривизны в них будут иметь противоположные проекции на нормаль секущей плоскости. Следовательно, все секущие плоскости заполняют двугранный угол δ , который в силу (142) и с учетом условия (140) оценивается неравенством

$$\delta \leq 2 \operatorname{arctg} \left(\frac{\operatorname{tg}(\Delta\psi/2)}{\cos \theta} \right) \leq 2 \operatorname{arctg} \left(\frac{\operatorname{tg}(\Psi/2)}{\cos \theta} \right) = \Phi, \quad (143)$$

где $\Delta\psi$ — диапазон скручивания (140) участка кривой.

Найдем связь между кривизной кривой k и кривизной k' ортогональной проекции этой кривой на плоскость. По определению, кривизна есть предел отношения угла поворота $\Delta\varphi$ касательного вектора при смещении по кривой на длину дуги $\Delta\sigma$ к этому смещению; $k = d\varphi/d\sigma$. Будем считать, что плоскость проекции расположена горизонтально. В каждой точке кривой введем в рассмотрение угол между главной нормалью и линией пересечения нормальной плоскости с плоскостью проекции α и угол между касательным вектором и плоскостью проекции β . Зафиксируем на кривой некоторую точку, а другую точку, отстоящую от первой на $\Delta\sigma$ по длине дуги, будем считать подвижной при стягивании дуги к фиксированной точке. В проекции кривой на горизонтальную плоскость имеем соответственно дугу-приращение длины $\tilde{\Delta\sigma}$ и угол-приращение $\tilde{\Delta\varphi}$.

Перенесем начала касательного и нормального векторов в фиксированной точке и касательного вектора в подвижной точке в центр сферы, расположенный на линии пересечения нормальной плоскости в фиксированной точке с плоскостью проекции (рис. 24). Длины

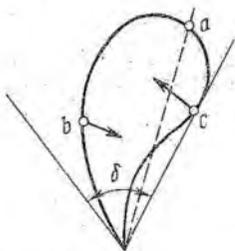


Рис. 23. Проекция участка пространственной кривой на ортогональную к хорде плоскость

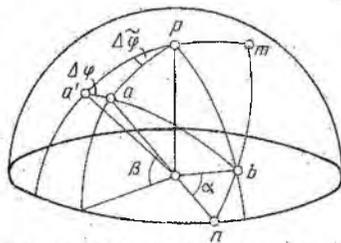


Рис. 24. Сферические треугольники

векторов и радиус сферы считаем равными единице. Введем обозначения: p — полюс сферы; n — след на сфере от линии пересечения нормальной плоскости с плоскостью проекции; a — конец фиксированного касательного вектора; a' — конец подвижного касательного вектора; b — конец фиксированного вектора главной нормали; m — точка пересечения меридиана точки a с нормальной плоскостью.

В сферическом треугольнике $aa'p$ имеет место соотношение

$$\frac{\sin \widehat{apa'}}{\sin \sphericalangle aa'} = \frac{\sin \widehat{aa'p}}{\sin \sphericalangle ap} \quad \text{или} \quad \frac{\sin \widetilde{\Delta\varphi}}{\sin \Delta\varphi} = \frac{\sin \widehat{aa'p}}{\sin ((\pi/2) - \beta)}.$$

Угол $aa'p$ с точностью до бесконечно малых высших (относительно $\Delta\sigma$) порядков равен углу \widehat{bap} . Синус последнего находится из прямоугольного сферического треугольника abm :

$$\sin \widehat{bap} = \sin \widehat{bam} = \sin \sphericalangle bm = \sin ((\pi/2) - \alpha).$$

Учитывая это, из предыдущей пропорции находим

$$\sin \Delta\widetilde{\varphi} \approx \frac{\sin ((\pi/2) - \alpha)}{\sin ((\pi/2) - \beta)} \sin \Delta\varphi.$$

Заменяя (с точностью до малых высших порядков) синусы на углы, получаем $\Delta\widetilde{\varphi} \approx (\cos \alpha / \cos \beta) \Delta\varphi$.

Длина проекции дуги между фиксированной и подвижной точками с такой же точностью выражается формулой $\Delta\widetilde{\sigma} \approx \Delta\sigma \cos \beta$.

Составив отношение $\Delta\widetilde{\varphi} / \Delta\widetilde{\sigma}$, в пределе получаем

$$k' = \lim_{\widetilde{\Delta\sigma} \rightarrow 0} \frac{\Delta\widetilde{\varphi}}{\Delta\widetilde{\sigma}} = \frac{\cos \alpha}{\cos^2 \beta} \lim_{\Delta\sigma \rightarrow 0} \frac{\Delta\varphi}{\Delta\sigma} = \frac{\cos \alpha}{\cos^2 \beta} k. \quad (144)$$

Это и есть искомое выражение для кривизны проекции кривой.

По построению (141) дуга интерполирующей кривой на участке $[i, i + 1]$ заключена в двугранный угол, образованный плоскостями первой и второй интерполирующих окружностей. В силу теоремы о дуге, опирающейся концами на плоскость, угол α для вектора главной нормали в любой точке участка $[i - 1, i + 1]$ данной кривой по отношению к плоскости первой интерполирующей окружности не может превышать величины Ψ (140):

$$\alpha \leq \Psi. \quad (145)$$

По условию (139) и по третьему экстремальному свойству раскрутки касательный вектор в любой точке участка кривой наклонен к хорде своего участка не более чем на угол θ . Поэтому с любой плоскостью, проходящей через эту хорду, касательные векторы образуют углы, не превышающие θ , т. е.

$$\beta \leq \theta. \quad (146)$$

Проекция участка $[i, i + 1]$ данной кривой на плоскость первой интерполирующей окружности имеет кривизну, определяемую фор-

мулой (144), откуда с учетом неравенств (145) и (146) получаем оценку

$$k'_{\min} = k_{\min} \cos \Psi \leq k' \leq \frac{k_{\max}}{\cos^2 \theta} = k'_{\max}. \quad (147)$$

Кривизна плоской кривой есть величина алгебраическая, со знаком, поэтому в случае, когда $\Psi > \pi/2$, вместо (147) нужно пользоваться оценкой

$$k'_{\min} = -\frac{k_{\max}}{\cos^2 \theta} \leq k' \leq \frac{k_{\max}}{\cos^2 \theta} = k'_{\max}. \quad (148)$$

В силу теоремы о сравнении оценкам (147) и (148) подчиняется кривизна k_1 первой интерполирующей окружности. Точно так же доказывается, что кривизна k_2 второй интерполирующей окружности удовлетворяет оценкам (147), (148). Аналогичными рассуждениями можно показать, что неравенствами (147), (148) оценивается и кривизна проекции участка $[i, i+1]$ интерполируемой кривой на секущую плоскость, т. е. плоскость, проходящую через хорду участка и любую точку кривой.

Поскольку векторы главных нормалей кривой на рассматриваемом участке отклоняются на угол, не больший Ψ , от плоскостей интерполирующих окружностей, такое же ограничение справедливо и для отклонения этих векторов от любой плоскости из двугранного угла, образованного плоскостями интерполирующих окружностей. Учитывая оценку (143), мы можем утверждать, что угол между любой секущей плоскостью и любой плоскостью из упомянутого выше двугранного угла не превышает величины Φ . Отсюда расстояние между точкой a на интерполируемой кривой и изопараметричной точкой b на интерполирующей кривой не превышает большей из длин сторон $a'b'$ или $a'b''$, лежащих против угла при вершине o в треугольниках $oa'b'$ и $oa'b''$ (рис. 25). Плоскости обоих этих треугольников ортогональны хорде, общий угол при вершине равен Φ , стороны oa' и ob' равны высоте H_1 дуги окружности кривизны k'_{\max} из (147), (148), опирающейся на хорду участка, а сторона ob'' равна высоте H_2 дуги окружности кривизны k'_{\min} , опирающейся на ту же хорду.

Если кривизна дуги окружности есть k , а хорда имеет длину l , то высота дуги выражается формулой

$$H = \frac{1 - \sqrt{1 - (lk/2)^2}}{|k|}.$$

Оценку погрешности окружностной интерполяции для участка между соседними узлами можно записать следующим образом:

$$|f(x) - g(x)| \leq \sqrt{H_1^2 + H_2^2 - 2H_1H_2 \cos \Phi}. \quad (149)$$

При $\Phi = 0, \pi$ неравенство (149) переходит в оценку (136) погрешности интерполяции плоской кривой.

Угловая погрешность. Расхождение направлений касательных векторов интерполируемой и интерполирующей кривых будем оценивать на участке между i -м и $(i+1)$ -м узлами с длиной хорды l .

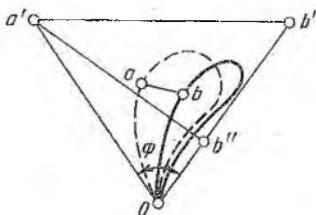


Рис. 25. Расхождение кривых при пространственной интерполяции

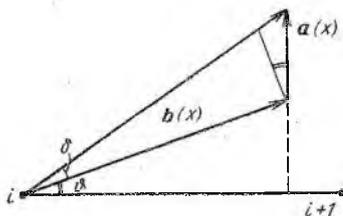


Рис. 26. Добавочный угол δ

В местной системе координат участка кривой производная от радиус-вектора текущей точки на интерполирующей кривой имеет следующее выражение:

$$g'(x) = \frac{1}{l} [g_2(x) - g_1(x)] + \left[\frac{l-x}{l} g'_1(x) + \frac{x}{l} g'_2(x) \right]. \quad (150)$$

Векторы $g'_1(x)$ и $g'_2(x)$ направлены по касательным к соответствующим окружностям, а по длине они не меньше единицы, так как $dx/d\sigma \leq 1$. Если начала этих векторов поместить, например, в i -й узел, то выражение во второй квадратной скобке (150) для каждого x задает исходящий из i -го узла вектор $b(x)$, конец которого лежит на прямолинейном отрезке, соединяющем концы векторов g'_1 и g'_2 . Каждый из этих последних векторов образует с хордой угол, не больший θ , поэтому и любой вектор $b(x)$ образует с хордой угол φ не больший θ . Так как угол между векторами $g'_1(x)$ и $g'_2(x)$ не больше 2θ , то длина любого вектора $b(x)$ оценивается снизу неравенством

$$|b(x)| \geq \cos \theta. \quad (151)$$

Поскольку угол между плоскостями интерполирующих окружностей не превышает величину Φ (143), длина вектора $a(x)$, стоящего в первой квадратной скобке выражения (150), точно оценивается правой частью неравенства (149):

$$|g_2(x) - g_1(x)| = |a(x)| \leq \sqrt{H_1^2 + H_2^2 - 2H_1H_2 \cos \Phi} = w(\Phi). \quad (152)$$

Вектор $\{(a(x)/l + b(x))\}$ образует с хордой участка максимальный угол тогда, когда хорда и оба вектора расположены в одной плоскости так, как показано на рис. 26. Тангенс добавочного угла δ $\text{tg } \delta \leq |a(x)| \cos \varphi / l |b(x)|$, откуда с учетом (151) и (152) имеем $\text{tg } \delta \leq w(\Phi) / (l \cos \theta)$. Следовательно, угол между касательным вектором участка интерполирующей дуги и хордой этого участка не превышает величины $\theta + \arctg(w(\Phi) / (l \cos \theta))$. С другой стороны, угол между касательным вектором интерполируемой дуги и хордой участка по условию (139) не превышает θ . Таким образом, получаем оценку угловой погрешности в виде

$$\gamma \leq \theta + \arctg(w(\Phi) / (l \cos \theta)). \quad (153)$$

Локализованная конечная ε -сеть. Построение такой сети для множества точек интерполируемой кривой возможно, если для каж-

дого значения x местной абсциссы имеется оценка уклонения по-местной ординате точек интерполирующей и интерполируемой кривых: $|f(x) - g(x)| < h$.

Выберем по хорде участка точки x_j , отстоящие одна от другой на расстояние, не большее h , и для этих абсцисс найдем точки B_j на интерполирующей кривой. На интерполируемой кривой этим абсциссам соответствуют точки A_j . Рассмотрим на интерполируемой кривой произвольную точку $A(x)$, и пусть x_j есть ближайшая к x_A точка дробления участка. По свойству треугольника имеем

$$\rho(A, B_j) \leq \rho(A, A_j) + \rho(A_j, B_j). \quad (154)$$

Используя оценку погрешности интерполяции, можно заключить, что второе слагаемое не превышает h .

Угол наклона прямолинейного отрезка AA_j к хорде участка не может быть больше максимального угла между касательной интерполируемой кривой и хордой участка. По условию (139) угол между касательной и хордой ограничен значением θ в случае плоской кривой. Для пространственной кривой такое же ограничение вытекает из условия (139), если учесть третье свойство раскрутки. Таким образом, и для плоской и для пространственной кривых первое слагаемое в (154) оценивается величиной $h/2\cos\theta$. Если $\theta \leq \pi/3$, то $\cos\theta \geq 1/2$ и справедливо неравенство

$$\rho(A, B_j) \leq h + h = 2h = \varepsilon,$$

т. е. точки B_j образуют ε -сеть для участка интерполируемой кривой при $\varepsilon = 2h$. А поскольку по построению для каждой точки B_j найдется точка интерполируемой кривой, отстоящая от B_j не далее, чем на ε , то совокупность точек B_j лежит в ε -окрестности интерполируемой кривой. Следовательно, построенная ε -сеть является локализованной. Повторив это построение на всех участках, получим локализованную конечную ε -сеть для всей интерполируемой кривой.

Оценочные неравенства связывают между собой величины l и ε . Величина l характеризует густоту дискретного задания кривой, а величина ε — допустимое отклонение приближенных точек от кривой. Оставляя в оценках знаки равенства и считая ε заданным, получим уравнения для определения густоты точек задания, обеспечивающей нужную точность.

Все более загромождая дискретное задание кривой, неограниченно уменьшая величину l — максимальное расстояние между соседними узлами, можно неограниченно уменьшать погрешность интерполяции. Например, для (136), разлагая радикалы в ряды по степеням $lk_{\max}/2$ и $lk_{\min}/2$, находим

$$\begin{aligned} |g(x) - f(x)| &\leq \frac{2}{k_{\max}} \left[\frac{1}{2} \left(\frac{lk_{\max}}{2} \right)^2 + \frac{1}{2 \cdot 4} \left(\frac{lk_{\max}}{2} \right)^4 + \dots \right] - \\ &- \frac{2}{k_{\min}} \left[\frac{1}{2} \left(\frac{lk_{\min}}{2} \right)^2 + \frac{1}{2 \cdot 4} \left(\frac{lk_{\min}}{2} \right)^4 + \dots \right] = \\ &= \frac{l^3}{8} (k_{\max} - k_{\min}) + O(l^5). \end{aligned}$$

Таким образом, интерполяция сходится к кривой с порядком $O(l^2)$. Для случая непрерывной кривизны разность $k_{\max} - k_{\min}$ на каждом участке сама убывает не медленнее, чем l , поэтому порядок сходимости в этом случае не ниже третьего.

Сходимость аппроксимации кривой с касательной. Для исследования сходимости рассмотрим неравенство (153), оценивающее угловую погрешность. Отношение длины дуги кривой к длине l стягивающей ее хорды стремится к единице при стягивании дуги в точку. Угол θ поворота касательной на стягивающейся дуге как интеграл от кривизны по длине дуги убывает не медленнее, чем $O(l)$. Величина ω во втором слагаемом правой части (153) имеет вид правой части неравенства (152), преобразуя которую, получим

$$\omega(\Phi) = \sqrt{(H_1 - H_2)^2 + 2H_1H_2(1 - \cos\Phi)}. \quad (155)$$

Здесь величина Φ выражается через максимально допустимый диапазон скручивания участка Ψ согласно формуле (143). Вследствие ограниченности $\cos\theta$ величина Φ асимптотически при l , стремящемся к нулю, ведет себя точно так же, как максимальное скручивание Ψ . Для дважды кусочно-дифференцируемой кривой максимальное скручивание ограничено, но может и не стремиться к нулю. Для дважды непрерывно дифференцируемой кривой величина Ψ как интеграл от кручения по длине дуги стремится к нулю не медленнее, чем $O(l)$.

Величины H_1 и H_2 имеют тот же смысл, что и величины в выражении (149) и могут быть представлены в виде ряда

$$H = \frac{2}{k} \left[\frac{1}{2} \left(\frac{kl}{2} \right)^2 + \frac{1}{2 \cdot 4} \left(\frac{kl}{2} \right)^4 + \dots \right]$$

по степеням l . Ясно, что каждое из H стремится к нулю не медленнее l^2 , а поскольку величина $(1 - \cos\Phi)$ по меньшей мере ограничена, то второе слагаемое под знаком радикала в (155) есть малая порядка $O(l^4)$.

Первое слагаемое под знаком радикала в (155) представляется в виде

$$(H_1 - H_2)^2 = [l^2 (k'_{\max} - k'_{\min})/8]^2 + \dots$$

Величины k'_{\min} и k'_{\max} согласно (147) и (148) ограничены, поэтому $(H_1 - H_2)^2$ имеет порядок не ниже четвертого. Величина $\omega(\Phi)$ стремится к нулю не медленнее, чем l^2 , а величина $\arctg(\omega/l \cos\theta)$ — не медленнее, чем l .

Таким образом, порядок стремления к нулю угловой погрешности не ниже первого. Такой же результат для плоского случая с очевидностью вытекает из оценки (138).

Конечная аппроксимация. Построив для каждого ε_n из убывающей числовой последовательности $\{\varepsilon_n\}$ свою локализованную конечную ε_n -сеть кривой, мы получим последовательность $\{L_n\}$ конечных множеств, все теснее «прижимающихся» к множеству точек кривой, которое обозначим D .

Рассмотрим сходящуюся последовательность $\{b_{pq}\}$ точек, взятых не более чем по одной из множеств L_n , и пусть b есть предел

этой последовательности. Поскольку каждое L_n лежит в ε_n -окрестности множества D , то в D для каждого $n = p_q$ найдется точка d_n такая, что $\rho(d_n, b_n) < \varepsilon_n$. Пользуясь свойством треугольника, запишем $\rho(d_n, b) \leq \rho(d_n, b_n) + \rho(b_n, b) = \varepsilon_n + \varepsilon_n = 2\varepsilon_n$.

Отсюда ясно, что при неограниченном убывании ε_n величина $\rho(d_n, b)$ стремится к нулю, т. е. точка b — это предел точек d_n , принадлежащих множеству D . Поскольку множество точек кривой есть замкнутое множество, точка b и, вообще, предельные точки всех сходящихся последовательностей $\{b_{p_q}\}$ принадлежит кривой.

С другой стороны, поскольку D лежит в ε_n -окрестности множества L_n , при каждом n для любой точки d кривой найдется точка b_n из L_n такая, что $\rho(d, b_n) < \varepsilon_n$. Следовательно, любую точку кривой можно рассматривать как предел сходящейся последовательности $\{b_n\}$ точек, взятых ровно по одной от каждого L_n . Множество пределов всех сходящихся последовательностей $\{b_{p_q}\}$ включает множество пределов всех сходящихся последовательностей $\{b_n\}$. Согласно изложенному выше указанные множества совпадают. Это единое множество называется *топологическим пределом последовательности точечных множеств* $\{L_n\}$ и обозначается $\lim_{n \rightarrow \infty} L_n$. Таким

образом, множество точек кривой является топологическим пределом последовательности своих локализованных ε -сетей при неограниченном убывании величины ε .

Наконец, следует рассмотреть отношение полученных здесь оценок (136) и (149) к понятиям близости геометрических объектов (гл. II. п. 2). Мера близости μ для кривых определена как нижняя грань множества максимумов расстояний между точками кривых при любых допустимых законах соответствия. Наши оценки получены для случая, когда соответствующими считаются точки, расположенные в одной плоскости, ортогональной к хорде участка. Итак, оценена величина, которая не меньше меры близости μ , а следовательно, не меньше χ , так как $\chi \leq \mu$.

В некоторых случаях значение μ может достигаться именно на принятом соответствии точек кривых. Например, когда плоские интерполируемая и интерполирующая дуги выпуклы «вверх», их максимумы имеют одно и то же значение, равное значению местной абсциссы, и при этой же абсциссе достигается максимальное расстояние по ординате между точками дуг. Если подобное обстоятельство имеет место на всех участках, то μ достигается на рассматриваемой параметризации для обеих кривых в целом.

Описанный процесс приближенной замены кривой другим геометрическим объектом — конечной совокупностью точек — есть *конечная аппроксимация континуального геометрического объекта*. Этот процесс может быть распространен на более сложный континуальный геометрический объект, каковым является кривая с касательной. Точки конечной локализованной ε -сети принадлежат интерполирующей кривой. Если вместе с точкой брать вектор, касательный к интерполирующей кривой в этой точке, то получится

конечная совокупность пар точка—вектор. Имея в виду оценку (153) угловой погрешности γ , можем считать, что полученная совокупность пар образует *конечную* (ε, γ) -сетью для кривой с касательной.

3. КАРКАСНАЯ АППРОКСИМАЦИЯ ПОВЕРХНОСТИ

Рассмотрим процедуру приближенного восстановления точек поверхности и нормалей в них, которая опирается только на интерполяцию кривых. Выберем некоторую ячейку каркаса, ограничивающую на поверхности клетку внутреннего диаметра d . Главные кривизны во всех точках клетки не превышают по абсолютной величине константу K . Нормаль к поверхности N известна в каждом узле каркаса точно, поскольку каркасные кривые пересекаются без касания. Если диаметр клетки удовлетворяет неравенству

$$Kd \leq \theta \leq \pi/2, \quad (156)$$

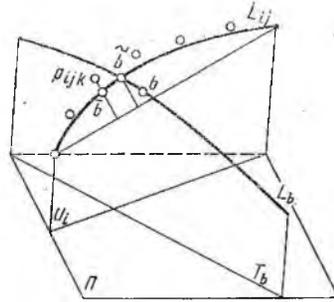
то существуют плоскости Π , называемые в дальнейшем *подстилающими*, такие, что для каждой точки поверхности найдется окрестность, которую можно взаимно однозначно спроецировать на плоскость Π . Например, такой плоскостью является плоскость, ортогональная нормали в одном из углов клетки.

Построение локализованной конечной ε -сети для поверхности. Предположим, что имеется способ интерполирования простой плоской дуги по известному положению ее концов, а погрешность интерполирования оценивается как максимум уклонения по местной ординате и не превышает величины $h_n(\mathcal{K}, l)$, где \mathcal{K} ограничивает кривизну, а l — длину хорды дуги.

Возьмем совокупность отстоящих одна от другой на расстояние $h = h_n(K \cos^{-1} \theta, d)$ плоскостей $\{U\}$, параллельных между собой и перпендикулярных плоскости Π . Если некоторая плоскость U имеет с рассматриваемой клеткой непустое пересечение, то в силу (156) это пересечение является кривой, которая состоит из конечного числа кусков — *компонент связности*. Каждая компонента связности является либо точкой, либо простой гладкой дугой с концами на краю клетки. Будем отмечать индексом i ($i = 1, 2, 3, \dots$) только те из плоскостей U , которые имеют в пересечении с клеткой хотя бы одну простую дугу L . Дуги, образующие сечение клетки плоскостью U_i , будем обозначать L_{ij} ($j = 1, 2, 3, \dots$). Ячейка, образующая край клетки, состоит, по определению, из простых гладких дуг L_{0j} ($j = 1, 2, 3, 4$). В целом имеем совокупность $\{L_{ij}\}$ ($i = 0, 1, 2, \dots$) простых гладких дуг, распределенных по клетке. На дугах, составляющих ячейку, выберем точки p_{0jm} с шагом $4h$ по длине дуги. Для каждой плоской дуги L_{ij} ($i = 1, 2, 3, \dots$) построим интерполирующую кривую, на которой с шагом h по местной абсциссе (вдоль хорды) найдем точки p_{ijm} . Полученное множество точек образует локализованную конечную ε -сеть для клетки.

Действительно. Пусть b — произвольная внутренняя точка клетки, а T_b — плоскость, проходящая через b ортогонально плоскостям U_i и плоскости Π (рис. 27). Точка b принадлежит одной из компонент

Рис. 27. Каркасная аппроксимация поверхности



связности сечения клетки плоскостью T_b . Эта компонента в силу (156) является простой дугой, которую обозначим L_b . В пересечении L_b с дугами L_{ij} найдется точка \tilde{b} , ближайшая к b . Согласно выбору плоскости Π , прямолинейный отрезок, соединяющий любые две точки поверхности, образуют с плоскостью Π угол, не превышающий угла θ .

Поэтому расстояние между точками b и \tilde{b} оценивается неравенством $\rho(b, \tilde{b}) \leq h/2\cos(Kd)$. Если \tilde{b} лежит на краю клетки, т. е. принадлежит некоторой L_{0j} , то ближайшая к \tilde{b} точка p_{0jm} отстоит не дальше, чем на $2h$ и

$$\rho(b, p_{0jm}) \leq \rho(b, \tilde{b}) + \rho(\tilde{b}, p_{0jm}) \leq \frac{h}{2\cos(Kd)} + 2h. \quad (157)$$

Если \tilde{b} принадлежит некоторой L_{ij} ($i \neq 0$), то на кривой, интерполирующей L_{ij} , найдется точка p_{ijm} , значение местной абсциссы x_{ijm} которой является ближайшим к $x_{\tilde{b}}$. Точку на L_{ij} с местной абсциссой x_{ijm} обозначим \tilde{b} . Имеем

$$\rho(b, p_{ijm}) \leq \rho(b, \tilde{b}) + \rho(\tilde{b}, \tilde{b}) + \rho(\tilde{b}, p_{ijm}). \quad (158)$$

Расстояние $\rho(b, \tilde{b})$ не превышает величины $h/2\cos(Kd)$, а расстояние $\rho(\tilde{b}, \tilde{b})$ не больше величины $h/2\cos\varphi$, где φ — максимальный угол наклона дуги L_{ij} к ее хорде. В силу выбора секущих плоскостей U_i дуга L_{ij} имеет кривизну, не большую $K/\cos(Kd)$ (формула Менье). Поэтому на основании оценки (137) можно утверждать, что φ не больше максимального угла наклона дуги окружности с кривизной $K/\cos(Kd)$ и хордой такой же длины, как у L_{ij} :

$$\sin\varphi \leq \frac{Kd}{2\cos(Kd)}, \quad \cos\varphi \geq \sqrt{1 - \left(\frac{Kd}{2\cos(Kd)}\right)^2} \geq \sqrt{1 - \left(\frac{\theta_0}{2\cos\theta}\right)^2}.$$

Последнее неравенство вытекает из условия (156). Ясно, что $\cos\varphi$ будет не меньше 0,5, если $\theta/2\cos\theta \leq \sqrt{3}/2$. Значение угла θ_0 , при котором $\theta_0/2\cos\theta_0 = \sqrt{3}/2$, несколько меньше радиана и не меньше 55° , поэтому, если

$$Kd \leq \theta_0 < 0,96, \quad (159)$$

то $\cos\varphi \geq 0,5$ и $\cos\theta_0 = 0,5$. При условии (159) оценки (157) и (158) можно преобразовать следующим образом:

$$\rho(b, p_{0jm}) \leq 2h + \frac{h}{2\cos\theta} \leq 2h + h = 3h;$$

$$\rho(b, p_{ijm}) \leq \frac{h}{2\cos(Kd)} + \frac{h}{2\cos\varphi} + h \leq h + h + h = 3h.$$

В общем случае для любой точки b , принадлежащей клетке, найдется такая точка из конечного множества $\{p_{ijm} (i = 0, 1, 2, \dots)\}$, что

$$\rho(b, p_{ijm}) \leq 3h. \quad (160)$$

Напомним, что шаг h был взят равным оценке $h_n(\mathcal{K}, l)$ уклонения по местной ординате между интерполируемой и интерполирующей плоскими кривыми на участке. Константа, ограничивающая кривизну интерполируемого плоского сечения клетки, $\mathcal{K} = K/\cos(Kd)$, а константа, ограничивающая длину хорды участка, равна максимальному диаметру d клеток. Поскольку величина h_n не меньше значения хаусдорфова расстояния между кривыми, то интерполирующая кривая лежит в h -окрестности интерполируемого сечения и, следовательно, принадлежит h -окрестности поверхности. Следовательно, точки p_{ijm} , взятые с интерполирующих кривых, тоже лежат в h -окрестности поверхности, т. е. точки p_{ijm} принадлежат поверхности. Итак, все точки p_{ijm} тем более лежат в $3h$ -окрестности клетки. Множество $\{p_{ijm}\}$ точек p_{ijm} , взятых по всем клеткам, лежит в $3h$ -окрестности поверхности в целом, и для каждой точки b поверхности найдется точка из $\{p_{ijm}\}$, расположенная от b на расстоянии, не большем $3h$. Переписав (160) в виде

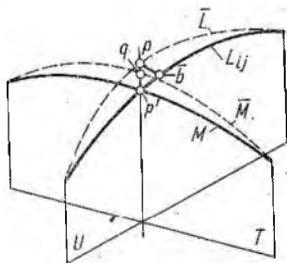
$$\rho(b, p_{ijm}) \leq 3h_n(K/\cos(Kd), d) = \varepsilon \quad (161)$$

и учитывая сказанное выше, мы можем утверждать, что построенное множество точек $\{p_{ijm}\}$ образует локализованную конечную ε -сеть поверхности. Причем $\varepsilon = 3h_n$. В предположении, что интерполяция кривой сходится, т. е. $h_n(\mathcal{K}, l)$ стремится к нулю при неограниченном убывании l , можно ожидать сходимость конечной аппроксимации поверхности при неограниченном загущении каркаса. Действительно, если максимальный диаметр клеток d неограниченно убывает, то значения $K/\cos(Kd)$ остаются ограниченными и, как явствует из (161), величина ε стремится к нулю.

Так как в описанном процессе построения конечной ε -сети используются не интерполирующие поверхности, а только вспомогательные кривые, «загущающие» исходный каркас, для этого способа представляется уместным название *каркасная аппроксимация*. Величина d характеризует густоту исходного каркаса, а величина ε — густоту результирующего каркаса. При известной оценке кривизны поверхности и заданном ε равенство справа в (161) является уравнением для определения максимального диаметра клеток, на которые исходный каркас разбивает поверхность. Так определяется густота исходного каркаса, обеспечивающая заданную точность каркасной аппроксимации поверхности.

Аппроксимация поверхности с нормалью. Зададимся оценкой $\gamma_n(\mathcal{K}, l)$ угловой погрешности интерполяции кривой. В частности, такой оценкой может служить (138). Чтобы построить конечное множество приближенных нормалей на клетке, поступим следующим образом. Возьмем произвольную «внутреннюю» ($i \neq 0$) точку p_{ijm} из множества $\{p_{ijm}\}$. Всякая плоскость, проходящая через эту

Рис. 28. Аппроксимация поверхности с нормалью



точку ортогонально Π -плоскости, сечет клетку по некоторым связным дугам, одна из которых (обозначим ее M) пересекается с дугой L_{ij} в точке p' . Выберем из этих плоскостей плоскость T такую, что хорда дуги M (концы дуги лежат на краю клетки) ортогональна хорде дуги L_{ij} (рис. 28). Интерполирующие дуги для L_{ij} и M обозначим соответственно \bar{L} и \bar{M} .

Точку пересечения интерполирующей дуги \bar{M} с плоскостью U_i обозначим q .

Для точек $p = p_{oim}$, которые распределены по ячейке, окаймляющей рассматриваемую клетку, плоскости T выбирают ортогональными к вектору τ_p , касающемуся каркасной линии в точке p .

Вектор $\tau_{\bar{L}}$, касательный к дуге \bar{L} в точке p , и вектор $\tau_{\bar{M}}$, касательный к дуге \bar{M} в точке q , определяют (с точностью до ориентаций) приближенную нормаль как векторное произведение $v_{ijm} = \tau_{\bar{L}} \times \tau_{\bar{M}}$. В случае $p = p_{ojk}$ вместо $\tau_{\bar{M}}$ берут τ_p .

Вектор v_{ijm} заменяет с определенной угловой погрешностью направляющий вектор истинной нормали в точках некоторого куска поверхности. А именно, этот кусок состоит из точек, для которых p является ближайшей из множества $\{p_{ijm}\}$.

При оценке угловой погрешности необходимо сравнивать направления нормалей, поэтому можно считать все направляющие векторы нормалей исходящими из одной точки. Для множества векторов, имеющих общее начало, вводится функция пары векторов N_1, N_2 , значение которой равно углу ϕ (в пределах от 0 до π) между векторами. Функция $\phi(N_1, N_2)$ обладает всеми метрическими свойствами и может быть названа *угловой метрикой*.

Для того чтобы получить оценку угла между истинной и приближенной нормалью, потребуется следующее.

З а м е ч а н и е: длина участка плоской кривой с максимальной (по абсолютной величине) кривизной k и хордой длиной l такими, что $lk < 2$, не превышает длины дуги окружности с той же кривизной и длиной хорды. Доказательство опускается.

Возьмем произвольную точку b на клетке. Вектор N_b есть нормаль к поверхности в этой точке. Пусть ближайшей к b точкой ϵ -сети оказалась точка p_{oim} , расстояние до которой оценивается неравенством (157). Длина дуги $b\bar{b}$ (см. рис. 27), являющейся частью дуги L_b , в соответствии с замечанием не превышает длины дуги окружности с кривизной $K/\cos(Kd)$ и хордой $b\bar{b}$. Эта дуга в свою очередь по длине не больше полуокружности, имеющей хорду $b\bar{b}$ своим диаметром. Длина последней равна $(\pi/2)r(b, \bar{b})$. Учитывая (157) и тот факт, что изменение (в угловой метрике ϕ) направления нормали поверхности при смещении по этой поверхности вдоль линии длины S не превы-

шает произведения KS , получаем оценку угла между истинными нормальными в точках b и \tilde{b} :

$$\vartheta(N_b, N_{\tilde{b}}) < K \frac{\pi}{2} \rho(b, \tilde{b}) \leq \frac{\pi h K}{4 \cos(Kd)}. \quad (162)$$

Упомянутый факт доказывается следующим образом. Для кривой ac длины S на поверхности при любом целом m существует кривая $a\tilde{c}_m$, составленная из конечного числа m' (необязательно равного m) геодезических дуг длиной $\Delta\sigma_i$, такая, что ее длина

$$\tilde{S}_m = \sum_{i=1}^{m'} \Delta\sigma_i \leq S + 1/m.$$

Главные нормали геодезической линии совпадают с нормальными поверхности, и, следовательно, поворот нормали поверхности при смещении вдоль геодезической линии равен интегралу по длине дуги от кривизны нормального сечения в направлении этой линии. Поскольку кривизны нормальных сечений поверхности ограничены константой K , то угол поворота нормали при смещении по каждой геодезической дуге, входящей в кривую $a\tilde{c}_m$, не превышает величины $K\Delta\sigma_i$. Изменение положения нормали по всей кривой $a\tilde{c}_m$

$$\sum_i K \Delta\sigma_i = K \sum_i \Delta\sigma_i = K\tilde{S} \leq KS + Km.$$

Переходя в правой части к пределу при неограниченном увеличении m , получим требуемое неравенство

$$\vartheta(N_a, N_c) \leq KS. \quad (163)$$

Из доказанного, в частности, вытекает, что нормали на участке поверхности, имеющем внутренний диаметр d , заключены в круговом конусе с углом раствора Kd . Таким образом получает обоснование вывод о существовании подстилающей плоскости Π , который был сделан из условия (156).

Продолжим анализ оценки погрешности аппроксимации нормалью поверхности. По свойству треугольника можно записать

$$\vartheta(N_b, v_{0jm}) \leq \vartheta(N_b, N_{\tilde{b}}) + \vartheta(N_{\tilde{b}}, N_{0jm}) + \vartheta(N_{0jm}, v_{0jm}).$$

Первое слагаемое справа уже оценено в (162). Точки \tilde{b} и p_{0jm} связаны дугой (каркасной линии), длина которой не больше $2h$, поэтому на основании (163) второе слагаемое оценивается величиной $2hK$. Угол между приближенной и истинной нормальными в точке p_{0jm} есть $\gamma = \gamma_a(K/\cos(Kd), d)$. В целом искомая оценка в рассматриваемом случае имеет вид

$$\vartheta(N_b, v_{0jm}) < \{(\pi/4 \cos(Kd) + 2)hK + \gamma\}. \quad (164)$$

Перейдем ко второму из двух возможных случаев, а именно к тому, когда ближайшей к точке b является внутренняя точка $p = p_{ijm}$ ($i \neq 0$). Рассмотрим на аппроксимируемой поверхности путь из точки b в точку p' по дугам $b\tilde{b}$, $\tilde{b}\tilde{b}$, $\tilde{b}p'$ линий L_b и L_{ij} (см.

рис. 27 и 28). В соответствии со свойством треугольника угловой метрики можем записать

$$\vartheta(N_b, N_{p'}) \leq \vartheta(N_b, N_{\bar{b}}) + \vartheta(N_{\bar{b}}, N_{\bar{b}}) + \vartheta(N_{\bar{b}}, N_{p'}) \quad (165)$$

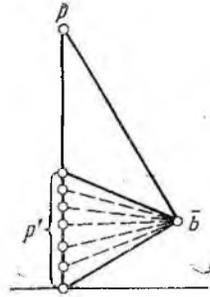
Учитывая замечание и неравенство (163), первые два слагаемых справа можно оценивать величинами $\frac{\pi h k}{4 \cos(Kd)}$ и $\frac{\pi h K}{4 \cos \varphi}$ соответственно.

В треугольнике $\bar{b}pp'$ угол при вершине p не превышает Kd , так как хорда линии L_{ij} не может быть наклонена к подстилающей плоскости на угол, больший Kd . По той же причине отрезок $\bar{b}p'$ составляет с подстилающей плоскостью угол не более Kd , следовательно, максимально возможная длина отрезка $\bar{b}p'$ получается в том случае, когда треугольник $\bar{b}pp'$ прямоугольный с прямым углом при вершине \bar{b} (рис. 29). Поскольку катет $\bar{b}p$ по длине не больше h , то катет $|\bar{b}p'| \leq h \operatorname{tg}(Kd)$. Полуокружность диаметром $h \operatorname{tg}(Kd)$ имеет длину, равную $\frac{\pi h \operatorname{tg}(Kd)}{2}$, а изменение направления нормали на такой длине не превышает величины $\frac{\pi h K \operatorname{tg}(Kd)}{2}$, которая и оценивает третье слагаемое в (165).

Оценка угла между истинной и приближенной нормальными в точках p' и p соответственно. Поскольку нормальный вектор получается как результат векторного умножения касательных векторов, то расхождение в направлениях нормалей можно считать следствием расхождения касательных интерполируемой и интерполирующей дуг. Угол между касательным вектором к дуге L_{ij} в точке p' (см. рис. 28) и касательным вектором к дуге \bar{L} не превышает $\gamma = \gamma_n(K/\cos(Kd), d)$. Точно так же оценивается и угол между касательными векторами к дугам M и \bar{M} в точках p' и q соответственно. Максимальное расхождение нормалей получается, если угловому расхождению между касательными придать максимальное значение γ . Считая, что оси x и y декартовой системы координат (x, y, z) параллельны соответственно хордам дуг L_{ij} и M , в качестве касательного вектора τ_L возьмем вектор с компонентами $(\cos(\gamma/2), 0, -\sin(\gamma/2))$, в качестве τ_M — вектор с компонентами $(0, \cos(\gamma/2), \sin(\gamma/2))$ в качестве $\tau_{\bar{L}}$ — вектор $(\cos(\gamma/2), 0, \sin(\gamma/2))$ и в качестве $\tau_{\bar{M}}$ — вектор $(0, \cos(\gamma/2), -\sin(\gamma/2))$.

Векторное произведение $\tau_L \times \tau_M = N$ будет иметь компоненты $\sin(\gamma/2)\cos(\gamma/2), -\sin(\gamma/2)\cos(\gamma/2), \cos^2(\gamma/2)$, векторное произведение $\tau_{\bar{L}} \times \tau_{\bar{M}} = \nu$ — компоненты $-\sin(\gamma/2), \cos(\gamma/2), \sin(\gamma/2) \times \cos(\gamma/2), \cos^2(\gamma/2)$. Исходя из геометрического смысла модуля векторного произведения $N \times \nu$ получим оценку

$$\sin \vartheta(N_{p'}, \nu_p) \leq \frac{2 \sin(\gamma/2) \cos(\gamma/2)}{1 + \sin^2(\gamma/2)} \sqrt{2} < (\sin \gamma) \sqrt{2}.$$



Поскольку синус не может быть больше своего угла, то

$$\sin \vartheta (N_{p'}, v_p) < (\gamma \sqrt{2}).$$

С учетом (165) получаем искомую оценку погрешности аппроксимации нормалей поверхности

$$\vartheta (N_{b'}, v_p) < \frac{\pi h K}{4 \cos (Kd)} + \frac{\pi h K}{4 \cos \varphi} + \frac{\pi}{2} h K \operatorname{tg} (Kd) + \arcsin \gamma \sqrt{2}.$$

При выполнении условия (159) $\cos \varphi \geq 0,5$, и окончательно получим

$$\vartheta (N, v) < \frac{\pi h K}{4 \cos (Kd)} + \frac{\pi h K}{2} + \frac{\pi}{2} h K \operatorname{tg} (Kd) + \arcsin \gamma \sqrt{2}. \quad (166)$$

Следует отметить, что, хотя секущая плоскость T при выводе оценки выбиралась из условия ортогональности хорд сечений L_{ij} и M , практически вполне достаточно брать ее просто ортогональной плоскости U_i .

Так как выбор клетки и точки p был произвольным, полученные оценки справедливы для всех клеток поверхности и множество пар (p_{ijm}, v_{ijm}) , взятое по всем клеткам, является *конечной аппроксимацией поверхности с нормалью*. На основании оценок (164) и (166) можно положительно решить вопрос о сходимости конечной аппроксимации нормалей. Диаметры клеток поверхности ограничены сверху величиной d , которая входит в правые части оценок (164) и (166) явно — через множители $1/\cos (Kd)$; $\operatorname{tg} (Kd)$ — и неявно — через функции $h = h_{\pi} (K/\cos (Kd), d)$;

$$\gamma = \gamma_{\pi} (K/\cos (Kd), d).$$

При загущении исходного каркаса, когда величина d неограниченно убывает, множитель $1/\cos (Kd)$ остается ограниченным по величине, а множитель $\operatorname{tg} (Kd)$ стремится к нулю. Естественно предположить, что используемая интерполяция кривой с касательной является сходящейся и величины h и γ тоже стремятся к нулю. Например, для окружностной интерполяции это следует из оценок (136) и (138), где роль d играет величина l . Тогда правые части оценок (164) и (166) так же, как и правая часть оценки (161), стремятся к нулю, оставаясь при этом положительными, что означает сходимость конечной аппроксимации поверхности с нормалью.

Об отыскании компоненты связности в плоском сечении поверхности. Как явствует из сказанного выше, процедура каркасной аппроксимации поверхности основана на приближенной замене плоского сечения клетки. Эту замену предлагается осуществлять интерполированием на основе дискретно-точечной информации, которую предоставляет след каркаса на секущей плоскости. Неявно предполагалось, что среди всех точек пересечения каркаса с плоскостью можно идентифицировать и упорядочить по длине дуги те, которые принадлежат одному связному куску (компоненте) сечения. Хотя в большинстве практических ситуаций подобный поиск не вызывает затруднений, это, вообще говоря, не совсем тривиальная задача.

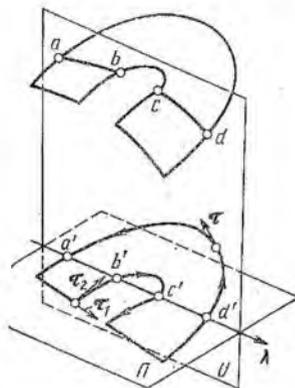


Рис. 30. Сечение подковообразной поверхности плоскостью

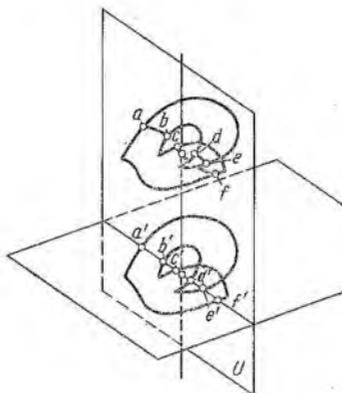


Рис. 31. Накрывающая поверхность

В простейшем случае, когда клетка (вместе с окаймляющей ее ячейкой) взаимно-однозначно проецируется в некоторую подстилающую плоскость Π , задача сводится к исследованию пересечений простого контура — ориентированной проекции ячейки — с прямой, в которую проецируется ортогональная подстилающей плоскости секущая плоскость U . Даже в этом случае сечение поверхности (будь то клетка или многоклеточная область) может иметь несколько компонент связности. Например, вертикальная плоскость U сечет подковообразную поверхность по кривой, имеющей две компоненты связности (рис. 30).

Зафиксированная ориентация поверхности определяет (подробно об этом будет сказано ниже) ориентацию проекции края. В плоскости Π получается простой (без самопересечений) кусочно-гладкий контур, в каждой точке которого определен положительный касательный вектор τ , а в угловых точках определены два таких вектора τ_1 и τ_2 . След секущей плоскости ориентируется выбором направляющего вектора λ .

В нашем примере связанные куски ab и cd на сечении выявляются в результате анализа ориентации векторных пар (λ, τ) в точках a', b', c', d' , где прямая с направляющим вектором λ пересекает контур. При тех ориентациях, которые выбраны для контура и для прямой на рис. 30, точки входа a', c' на проекции связных кусков сечения характеризуются левой парой (λ, τ) , а точки выхода b', d' — правой парой. Если прямая пересекается с контуром в угловой точке, то вопрос о том, является ли последняя точкой входа-выхода, очевидным образом решается однозначно из анализа ориентации векторных пар (λ, τ_1) и (λ, τ_2) . Ориентация пары (λ, τ) неколлинеарных векторов определяется относительно ориентации базисного репера системы координат (x, y) . Для этого вычисляют детерминант матрицы, составленной из координат векторов пары:

$$\det \begin{pmatrix} x_\lambda & y_\lambda \\ x_\tau & y_\tau \end{pmatrix},$$

При положительном знаке детерминанта пара (λ, τ) ориентирована так же, как базисная пара (ox, oy) , при отрицательном знаке — противоположно базисной.

Более сложным является случай, когда для данной области поверхности не существует подстилающей плоскости, в которую область проецировалась бы взаимно-однозначно. Неравенство (156) — это только необходимое условие существования такой подстилающей плоскости, оно обеспечивает лишь локальную взаимную однозначность проецирования. То есть при выполнении условия (156) можно разбить поверхность на такие куски, которые проецируются взаимно-однозначно. Однако нетрудно привести пример, когда при выполнении условия (156) взаимная однозначность проецирования не может быть осуществлена в целом.

На рис. 31 изображена поверхность, два куска которой перекрывают друг друга, и их проекции в подстилающей плоскости совпадают. Можно представить поверхности, проецирование которых в подстилающую плоскость приводит к многократному наложению проекций перекрывающихся частей. Подобное непрерывное проецирование, взаимно-однозначное лишь «в малом» или, как говорят математики, — «локально», служит примером отображений, которые называются *накрывающими*. Сама поверхность по отношению к ее проекции в подстилающей плоскости является *накрывающей поверхностью*.

В примере на рис. 31 след секущей плоскости U имеет общие точки a', b', c', d', e', f' с контуром ориентированной проекции края. Как легко убедиться, приведенное выше правило определения точек выхода-входа не действует в случае, когда контур имеет самопересечения. Действительно, пара $(\gamma, \tau)_a$ является левой, пара $(\lambda, \tau)_e$ — правой, однако отрезок $[a', e']$ отнюдь не есть проекция какого-либо связанного куска в сечении поверхности плоскостью U .

Тем не менее и в общем случае, когда контур ориентированной проекции края не является простым, существует способ отыскания связанных кусков в сечении накрывающей поверхности по известной проекции ее края. Не вдаваясь в теоретические подробности, мы приведем в следующем параграфе только общий алгоритм решения задачи.

4. РЕШЕНИЕ ЗАДАЧИ О СВЯЗНОМ СЕЧЕНИИ

Ориентированный контур. Положительная ориентация контура проекции края должна быть обязательно зафиксирована. Это можно сделать, если известна ориентация самой накрывающей поверхности, заданная выбором вектора положительной нормали N . Край куска поверхности ориентируется выбором направляющего вектора из векторов, касательных к кривой края в данной точке. Осуществим этот выбор согласно следующему правилу: направляющий вектор края, вектор, касательный к поверхности в этой же точке и направленный внутрь поверхности, а также вектор положительной нормали образуют правую тройку. Ориентация края естественным образом индуцирует ориентацию его проекции на подстилающей плоскости: направляющий вектор контура проекции есть проекция направляющего вектора края.

Край взаимно-однозначно проецируемой поверхности дает в проекции на подстилающей плоскости простой контур. Если смо-

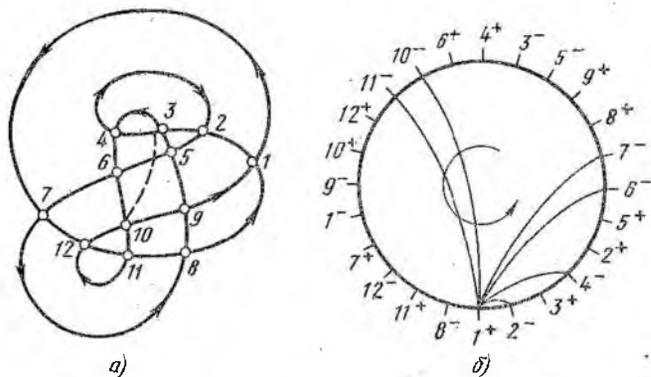


Рис. 32. Контур и диаграмма его самопересечений

треть на подстилающую плоскость в направлении отрицательной нормали поверхности, то обход этого простого контура в положительном направлении совершается против часовой стрелки. Для накрывающей поверхности положительный обход контура-проекции края осуществляется так, что проекция локального участка поверхности остается слева (рис. 32, а). Перенумеровав совершенно произвольно все точки самопересечений, можно составить круговую диаграмму порядка прохождений этих точек при полном обходе контура (рис. 32, б). Каждое самопересечение при полном обходе контура проходится дважды. Одно прохождение осуществляется так, что пересекаемый участок направлен слева направо. Такие прохождения будем отмечать знаком плюс и называть входными. При втором прохождении того же самопересечения пересекаемый участок направлен справа налево. Такие прохождения будем отмечать знаком минус и называть выходными. Число точек на диаграмме всегда четное.

Пусть I и II — два различных самопересечения контура (рис. 33), а I^+ и II^- — такие последовательные прохождения (входные и выходные) этих самопересечений, что между ними нет других прохождений этих же самопересечений. Подобную пару (I^+ , II^-) будем называть парой смежных прохождений или, короче, — смежной парой. В примере на рис. 32 для входного прохождения самопересечения I смежными парами (соединены дугами внутри круга диаграммы) будут пары (1^+ , 2^-), (1^+ , 4^-), (1^+ , 6^-), (1^+ , 7^-), (1^+ , 10^-),

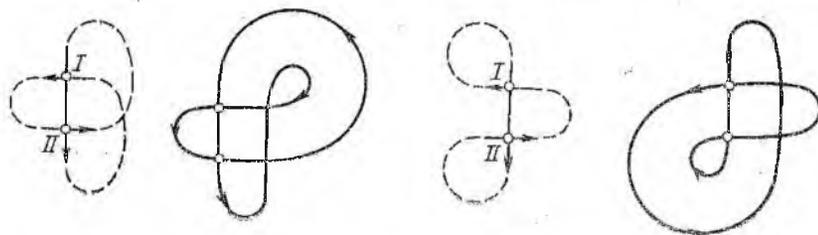


Рис. 33. Примеры контуров со схемами прохождения самопересечений

ТАБЛИЦА 4. Список смежных пар

(+)	(-)
1	2, 4, 6, 7, 10, 11
2	3, 6, 7, 10, 11, 1
3	4, 6, 7
4	10, 11, 9, 1, 8, 2
5	6, 7
6	10, 11, 9, 1, 8, 2, 4
7	8, 2, 4, 6
8	5, 3, 10, 11, 1
9	5, 3, 10, 11
10	9, 1, 8, 2, 4, 6
11	8, 2, 4, 6, 7, 10
12	9, 1

(1⁺, 11⁻). Список всех смежных пар в этом примере содержит 57 пар (табл. 4).

Вспомогательную роль играет процедура *саморазделения контура* на два подконтура по паре смежных прохождений. Возможны два варианта порядка прохождений самопересечений I и II при полном обходе контура; либо I⁺, II⁻, II⁺, I⁻, I⁺, либо I⁺, II⁻, I⁻, II⁺, I⁺. На рис. 33 приведены примеры контуров и даны соответствующие схемы порядка прохождения самопересечений I и II по первому варианту (слева) и по второму (справа).

Саморазделение контура по смежной паре (I⁺, II⁻). В первом варианте один подконтур составляется из последовательно проходимых дуг I⁺II⁻, II⁺I⁻, а другой подконтур — из дуг I⁺II⁻, II⁻II⁺, II⁻I⁺, I⁻I⁺. Во втором варианте один подконтур составляется из дуг I⁻II⁺, II⁻I⁺, а другой, — из дуг I⁺II⁻, II⁻I⁺, I⁺II⁻, II⁺I⁺. Примеры саморазделений приведены на рис. 34. В обоих вариантах саморазделения дуга I—II проходится трижды: два раза в прямом направлении и один раз — в обратном. На рисунке кратные проходы «расщеплены» для наглядности.

В результате саморазделения получаются два «дочерних» контура (рис. 34), каждый из которых не имеет самопересечений в точках I и II. Однако один из дочерних контуров имеет самосовпадение на участке I—II. Суммарное число самопересечений дочерних контуров во всяком случае на два меньше числа самопересечений исходного контура. Каждый дочерний контур тоже может быть подвергнут саморазделению, если он имеет более одного самопересечения.

Всякому контуру, принимаемому за исходный, можно сопоставить дерево саморазделений (дерево — сорт графа, см. гл. V, п. 2) В нем ребра соответствуют саморазделениям, корень — исходному контуру И, а все остальные вершины соответствуют парам дочерних контуров. Абстрактный пример дерева саморазделений приведен на рис. 35. Множество вершин естественным образом стратифицировано по «высоте», а именно; нулевой ярус состоит только из корня,

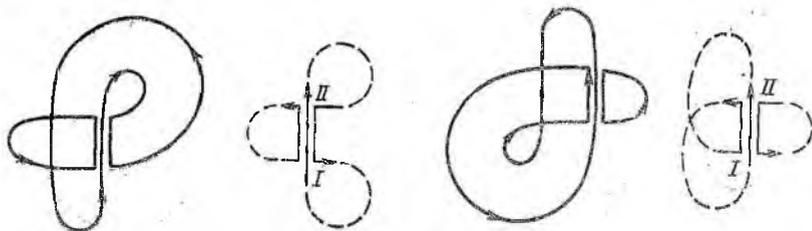


Рис. 34. Саморазделение контура по смежным прохождениям самопересечений I и II

вершины, непосредственно связанные с корнем, образуют первый ярус, вершины, непосредственно связанные с вершинами первого яруса, образуют второй ярус, и т. д. В каждом ненулевом ярусе могут быть так называемые висячие вершины, которые не связаны с вершинами следующего по высоте яруса. Висячие вершины соответствуют таким двум подконтурам, к которым не применялись процедуры саморазделения.

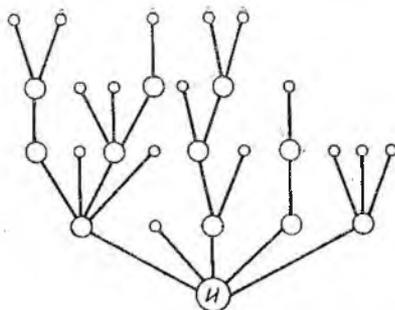


Рис. 35. Отвлеченный пример дерева саморазделений

Ориентированные замкнутые кривые класса Q на плоскости (контуры) разделяются на две группы: одну составляют контуры, которые могут служить проекцией края накрывающей поверхности, в другую группу входят контуры, которые не могут быть такой проекцией. Первые мы будем называть правильными, вторые — неправильными. Положительно ориентированные простые контуры являются правильными, а отрицательно ориентированные простые контуры — неправильными. К неправильным относятся контуры с нечетным числом самопересечений, а также такие контуры с четным числом самопересечений, для которых нельзя указать ни одной пары смежных прохождений. Контуры перечисленных типов с легко устанавливаемым качеством мы будем называть тривиальными, а все прочие — нетривиальными. В полном дереве саморазделений висячие вершины соответствуют тривиальным контурам.

Разделение контура секущей дугой. Эта вспомогательная процедура состоит в следующем. Имеется не принадлежащая контуру простая дуга, которая пересекает контур в двух различных точках I и II таким образом, что при движении по дуге в направлении от I к II точка I проходится как входная I^+ , а точка II — как выходная II^- . При разделении контура данной секущей дугой один подконтур составляется из куска I^+II^- секущей дуги и дуги II^+I^- исходного контура, а второй подконтур — из куска II^-I^+ секущей дуги и дуги I^-II^+ исходного контура. Пример разделения контура секущей дугой приведен на рис. 36, где секущая дуга показана штриховой линией.

Для решения задачи о связном сечении накрывающей поверхности нужно уметь распознавать принадлежность данного нетривиального контура к группе правильных или неправильных контуров. Ответ на вопрос: «Является ли данный контур правильным или нет?» — дает алгоритм распознавания качества контура, который мы изложим в нестрогой словесной форме. Этот алгоритм заключается в последовательном обходе вершин дерева саморазделений, начиная от корня, с целью отыскания висячих вершин, которым соответствуют легко распознаваемые тривиальные контуры. Обход состоит из отдельных переходов между непосредственно связанными вершинами.

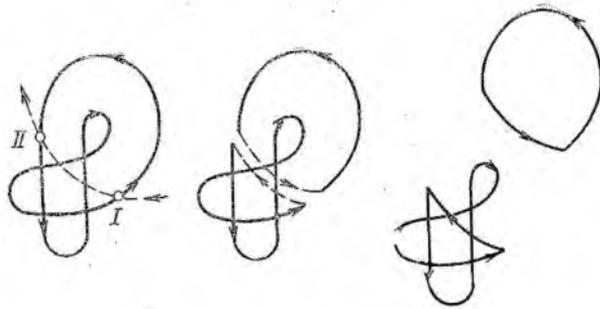


Рис. 36. Три фазы разделения контура секущей дугой

Различают прямые переходы, когда последующая вершина лежит на более высоком ярусе, и обратные переходы, когда предшествующая вершина принадлежит более высокому ярусу.

Алгоритм распознавания качества контура. Этот алгоритм распадается на элементарные действия. Мы сгруппируем эти действия в несколько операторов следующего содержания.

Вспомогательные:

Оператор 1. Строится круговая диаграмма исходного контура и составляется список всех смежных пар для него. Текущий номер яруса полагается нулевым.

Оператор 10. Формируется ответ и осуществляется выход.

Основные:

Оператор 2. Обрабатывается информация, которая относится к корню дерева. Из списка смежных пар выбирается очередная смежная пара, которая передается для обработки оператору 4.

Оператор 3. Обрабатывается информация, относящаяся к обычной вершине дерева. Просматривается список 1 смежных пар. Если список пуст, то контур 1, соответствующий данной вершине, — тривиален. Устанавливается качество контура 1. Если список не пуст и не исчерпан, из него выбирается очередная смежная пара, которая передается для обработки операторами 4 или 5.

Если список 1 исчерпан, просматривается список 2. Если он пуст, то контур 2, соответствующий данной вершине — тривиален. Если список 2 не пуст и не исчерпан, из него выбирается очередная смежная пара, которая передается для обработки операторами 4 или 5.

Оператор 4. Прямой переход к следующей вершине дерева. Осуществляются саморазделение одного из контуров предыдущей вершины по заданной смежной паре. Два дочерних контура соответствуют последующей вершине, находящейся на более высоком ярусе. Для этих контуров составляются соответственно списки 1 и 2 смежных пар прохождений и передаются для обработки оператором 3. Текущий номер яруса увеличивается на единицу.

Оператор 5. Выполняется обратный переход. Списки 1 и 2, относящиеся к предыдущей вершине более высокого яруса, изымаются из области действия оператора 3. Оператору 3 передаются для об-

работки списки, относящиеся к последующей вершине. Текущий номер яруса уменьшается на единицу.

Главные логические:

Оператор 6. Проверяется текущий номер яруса. Если он равен нулю, осуществляется возврат к оператору 2. Если текущий номер яруса отличен от нуля, управление передается оператору 3.

Оператор 7. Проверяются состояния списков 1 и 2. Если один из списков пуст, а другой исчерпан, или оба списка пусты, и хотя бы один тривиальный контур из двух является неправильным, управление передается для выполнения обратного (признак —) хода по дереву. В любом ином случае управление передается на прямой (признак +) ход.

Оператор 8. Если исчерпан список смежных пар (да), то исходный контур является неправильным и осуществляется выход через оператор 10. Иначе (нет) — работает оператор 2.

Оператор 9. Если списки 1 и 2 пусты и оба тривиальных контура являются правильными (ПП), производится выход через оператор 10. Иначе (ПН) — работает оператор 7.

Введенные числовые обозначения операторов позволяют сравнительно просто изобразить укрупненную схему алгоритма распознавания качества нетривиального контура (рис. 37). Из структуры алгоритма явствует, что положительное качество исходного контура может быть обнаружено задолго до окончания полного обхода дерева саморазбиений. По-иному обстоит дело с неправильными контурами. В таких случаях всякой висячей вершине дерева соответствует пара тривиальных контуров, в которой хотя бы один неправильный. Для установления этого факта необходимо просмотреть каждую висячую вершину, т. е. фактически произвести обход всех вершин.

Таким образом, представлен достаточный материал, чтобы изложить метод решения задачи о связном сечении даже в более общей постановке. А именно, пусть накрывающая поверхность сечется цилиндром, направляющая которого есть простая дуга в подстилающей плоскости, а образующие ортогональны этой плоскости. Направляющая цилиндра является секущей дугой по отношению к контуру — ориентированной проекции края — и пересекается с контуром в точках I и II. Требуется определить,

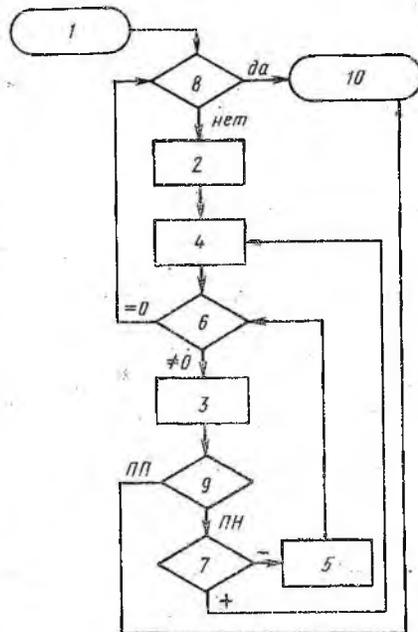


Рис. 37. Блок-схема алгоритма распознавания качества контура

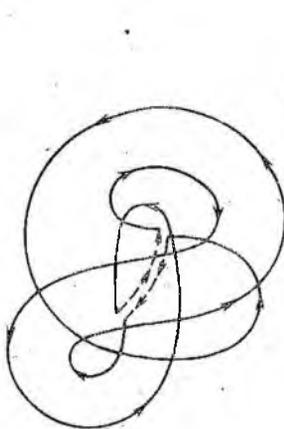


Рис. 38. Разделение основного контура по секущей дуге

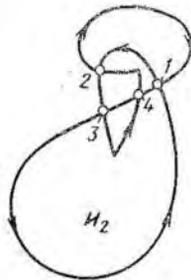
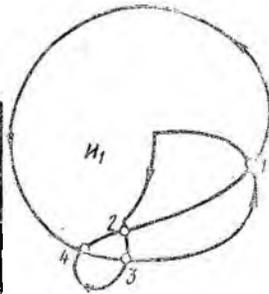
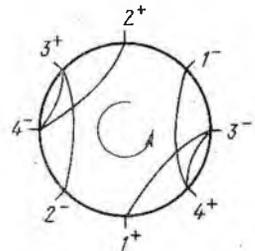
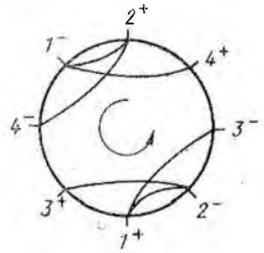


Рис. 39. Результирующие контуры с их круговыми диаграммами



является ли участок $I-II$ секущей дуги проекцией связного сечения поверхности цилиндром.

Алгоритм решения задачи о связном сечении. Решение проводится в три шага.

Шаг 1. Данный контур делится на два подконтур по секущей дуге $I-II$.

Шаг 2. Распознается качество каждого из полученных подконтуров. Если подконтур тривиальный, его качество определяется непосредственно. Если подконтур нетривиальный, он принимается за исходный для алгоритма распознавания качества.

Шаг 3. Если хотя бы один из подконтуров оказался неправильным, ответ задачи отрицателен: секущая дуга не является проекцией связного сечения накрывающей поверхности цилиндром. Если оба подконтуров правильные, ответ задачи положителен.

Ради простоты изложения мы считали, что самопересечения происходят в изолированных точках и в одной точке пересекаются только два участка контура. Допущение кратных самопересечений и самопересечений по целым дугам не вносит ничего принципиально нового в изложенную схему, так как малой деформацией контура всякое самопересечение по дуге можно превратить в точечное (а всякое самосовпадение вообще устранить), не меняя при этом общего порядка прохождения всех самопересечений. В случае пересечения в одной точке нескольких дуг контура пересечение каждой пары дуг рассматривается как самостоятельное самопересечение.

Проиллюстрируем применение алгоритмов на примере контура (будем именовать его основным), изображенного на рис. 32, где секу-

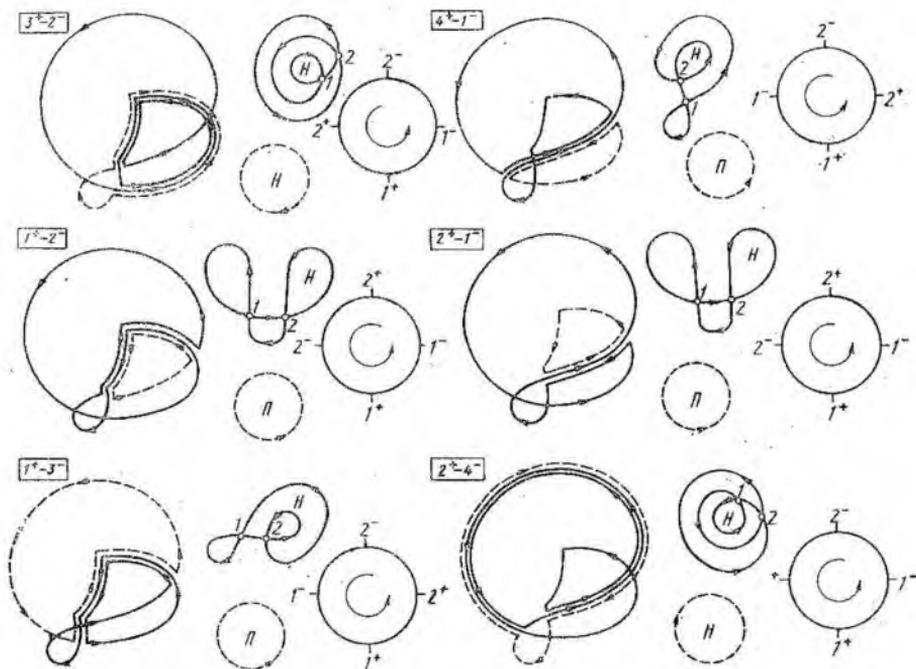


Рис. 40. Саморазделение контура I_1

шая дуга дана штриховой линией. Разделение основного контура по секущей дуге изображено на рис. 38. Первый I_1 и второй I_2 результирующие контуры с их круговыми диаграммами приведены на рис. 39.

Далее на рис. 40 представлены все шесть возможных саморазделений первого контура по смежным парам, указанным в прямоугольных рамках слева на рисунке (для каждого саморазделения).

Один из подконтуров каждого саморазделения дан непрерывной линией, а другой — штриховой. Самосовпадения на подконтурах «малыми деформациями» устранены. Справа от схемы каждого саморазделения изображены наглядные эквиваленты (в смысле порядка обхода) подконтуров и даны круговые диаграммы для подконтуров с самопересечениями. В каждом из шести саморазделений оба под-

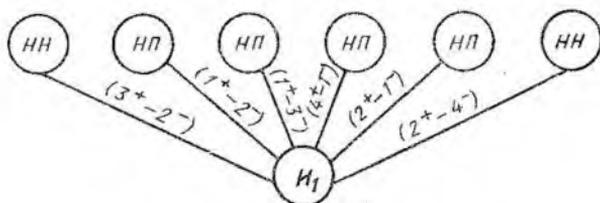


Рис. 41. Дерево саморазделений контура I_1

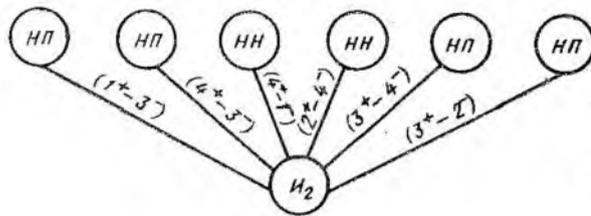


Рис. 42. Дерево саморазделений контура I_2

контура получаются тривиальными: один простой, другой — без смежных пар. Некоторые из подконтуров правильные ($П$), но всегда имеется хотя бы один неправильный ($Н$). Соответствующие деревья саморазделений на рис. 41, 42 имеют уже на первом ярусе все висячие вершины. На ребрах указаны смежные пары, а в вершинах — качество ($П$ или $Н$) получающихся подконтуров.

Таким образом, первый результирующий контур разделения по секущей дуге основного контура оказался неправильным, и в соответствии с алгоритмом секущая дуга квалифицируется как не являющаяся проекцией связного сечения накрывающей поверхности. Интересно, что в случае отрицательного ответа вообще нет необходимости проверять качество второго результирующего контура разделения по дуге, что совершенно обязательно в случае положительного ответа. Получается как бы компенсация за необходимость полного перебора вершин дерева саморазделений при определении неправильности первого контура. Приведем только дерево саморазделений второго контура (рис. 42), оставив саму процедуру в качестве упражнения читателю.

В ходе решения задачи о связном сечении возникают промежуточные подзадачи, геометрические по сути. К их числу можно отнести проецирование края поверхности в плоскость; подсчет числа самопересечений плоского контура, установление порядка и характера их проходов, получение геометрических описаний результирующих подконтуров в процедурах разделения и саморазделения. Все такие подзадачи решаются в принципе на основе сплайн-интерполяции кривых, но постановка на ЭВМ соответствующих алгоритмов — далеко не простое дело. Изложенное здесь может служить лишь руководством для практической реализации.

Однако дело не исчерпывается только этим. По нашему мнению, содержание данного параграфа иллюстрирует на примере задачи о связном сечении, специфику геометрической аппроксимации вообще. Ничего, подобного этой задаче, не может и возникнуть в процессе аппроксимации функций одной или многих переменных.

Завершая главу, следует сказать, что окружностные сплайны, рассматриваемые как функции на отрезке, не образуют подпространства в линейном пространстве непрерывных функций. В этом их недостаток с теоретической точки зрения по сравнению с другими неполиномиальными (например, экспоненциальными) сплайнами.

Глава IV

МЕТОДЫ ЧИСЛЕННОГО РЕШЕНИЯ ЗАДАЧ ИНЖЕНЕРНОЙ ГЕОМЕТРИИ

В предыдущих главах были рассмотрены методы моделирования кривых и поверхностей линейными и нелинейными сплайнами. В данной главе изложены приемы вычисления таких величин, как длина дуги, площадь поверхности, объем тела, методы решения задач об определении расстояния между элементарными геометрическими объектами и об их пересечении, а также геометрическая теория построения разверток и квазиразверток поверхностей на плоскости. Предлагаемые методы в основе своей не зависят от способа математического представления кривых и поверхностей. Но рассмотрены они применительно к V-сплайновым кривым и W-сплайновым поверхностям. В этом случае обеспечивается экономичность вычислительных алгоритмов при автоматизации расчетов с помощью ЭВМ.

1. РАСЧЕТ ХАРАКТЕРИСТИК КРИВОЙ И ПОВЕРХНОСТИ

Вычисление координат точки сплайновой кривой $r = V(t)$ и ее дифференциальных характеристик для заданного значения параметра t . Этот расчет проводится в два шага.

Шаг 1 — отыскание отрезка $[t_i, t_{i+1}]$ разбиения $\Delta: t_1 < t_2 < \dots < t_N$, которому принадлежит t . Забегая вперед, отметим, что при решении различных задач приходится вычислять координаты точек кривой для последовательности значений $t_v, v = 1, 2, \dots$. Это может быть либо возрастающая (убывающая) последовательность, элементы которой, как правило, располагаются на подряд идущих отрезках $[t_i, t_{i+1}]$, либо последовательность, сходящаяся к некоторому \bar{t} . Поэтому поиск отрезка, которому принадлежит значение t_{v+1} , целесообразно вести простым перебором, начиная с отрезка, содержащего t_v . Поиск местоположения t_i можно начать с любого отрезка, например первого. Кроме того, если t_{v+1} совпадает с узлом t_i сетки Δ , то окончательно выбирают отрезок $[t_i, t_{i+1}]$, если $t_{v+1} > t_v$, или отрезок $[t_{i-1}, t_i]$, если $t_{v+1} < t_v$.

Шаг 2 — вычисление значения V-сплайна $V(t)$ и его производных. Для нахождения таких дифференциальных характеристик, как орты естественного трехгранника, кривизна, кручение и т. п., необходимо знать значения производных V-сплайна до третьего порядка.

Пусть требуется вычислить совокупность производных $(V^{(p_1)}(t), V^{(p_2)}(t), \dots)$, которая может содержать до четырех векторов. Используем следующий прием подсчета сразу всех элементов этой совокупности.

ТАБЛИЦА 5. Вычисление элементов матрицы A_p

φ_k	p			
	0	1	2	3
$\tilde{\varphi}_2^{(p)}$	$u^2(3-2u)$	$6u(1-u)/\Delta t_i$	$6(1-2u)/\Delta t_i^2$	$-12/\Delta t_i^3$
$\tilde{\varphi}_1^{(p)}$	$1-\tilde{\varphi}_2$	$-\tilde{\varphi}_2'$	$-\tilde{\varphi}_2''$	$-\tilde{\varphi}_2'''$
$\tilde{\varphi}_4^{(p)}$	$u^2(1-u)\Delta t_i$	$u(3u-2)$	$(6u-2)/\Delta t_i$	$6/\Delta t_i^2$
$\tilde{\varphi}_3^{(p)}$	$(u-\tilde{\varphi}_4)\Delta t_i - \tilde{\varphi}_4$	$(1-\tilde{\varphi}_4) - \tilde{\varphi}_4'\Delta t_i$	$\tilde{\varphi}_4''\Delta t_i - \tilde{\varphi}_4'''$	$\tilde{\varphi}_4''''$

Введем матрицы:

$$A_p = \begin{bmatrix} \tilde{\varphi}_1^{(p_1)}(u) & \tilde{\varphi}_2^{(p_1)}(u) & \tilde{\varphi}_3^{(p_1)}(u) & \tilde{\varphi}_4^{(p_1)}(u) \\ \tilde{\varphi}_1^{(p_2)}(u) & \tilde{\varphi}_2^{(p_2)}(u) & \tilde{\varphi}_3^{(p_2)}(u) & \tilde{\varphi}_4^{(p_2)}(u) \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}; \quad \mathbf{B}_i = \begin{bmatrix} V_i \\ V_{i+1} \\ V_i' \\ V_{i+1}' \end{bmatrix} \quad (167)$$

Тогда $(\mathbf{V}^{(p_1)}(t), \mathbf{V}^{(p_2)}(t), \dots) = A_p \mathbf{B}_i$.

Значения $\tilde{\varphi}_k^{(p)}(u)$ есть производные функции $\varphi_k(u)$ по переменной t , где $u = (t - t_i)/\Delta t_i$. Их рекомендуется вычислять в соответствии с табл. 5.

Вычисление координат точки сплайновой поверхности $\Gamma = W(t, s)$ для заданных значений $t, s \in \Omega$. Это вычисление также выполняется в два шага.

Шаг 1. Находят клетку Ω_{ij} разбиения $\Delta = \Delta_t \times \Delta_s$ области Ω , которой принадлежат значения t, s . С этой целью к каждой из переменных применяют тот же прием, что и в случае V-сплайна.

Шаг 2. Вычисляют значения W-сплайна и любых его производных, используя выражение

$$(W^{(p_1, q_1)}(t, s), W^{(p_2, q_2)}(t, s), \dots) = A_p F_{ij} A_q^T,$$

где $p_k, q_l \in \{0, 1, 2, 3\}$. Матрицы A_p и A_q имеют вид (167). Их элементы рассчитывают для значений u и v соответственно. Матрица F_{ij} имеет тот же вид, что и в (118).

Вычисление длины дуги сплайновой кривой с заданной точностью. Вопрос о приближенном вычислении длин дуг кривых уже был рассмотрен в гл. II, п. 5. Напомним, что если узлы линейной интерполяции кривой взяты с шагом δt по параметру $t \in [a, b]$ кривой в соответствии с соотношением (122), то длина вписанной ломаной будет отличаться от длины дуги кривой на величину не более ε .

Рассмотрим V-сплайн $V(t)$, определенный на сетке Δ . Длину дуги сплайновой кривой будем находить как сумму длин дуг σ_i ее звеньев. Чтобы обеспечить точность ε вычисления всей длины σ , достаточно подсчитать длину каждой из σ_i с точностью $\varepsilon(N-1)^{-1}$. Из

соотношения (122) находим допустимый шаг δt_i при интерполяции i -го звена V -сплайна сплайном первой степени;

$$\delta t_i = \frac{3\varepsilon}{(N-1)(t_{i+1}-t_i) \|V''(t)\|_{[t_i, t_{i+1}]}}.$$

Здесь $\|V''(t)\|_{[t_i, t_{i+1}]} = \max(|V''(t_i)|, |V''(t_{i+1})|)$.

Площадь плоской области. Площадь области G , ограниченной замкнутой сплайновой кривой $r = V(t)$, может быть представлена интегралом

$$S = \frac{1}{2} \left| \int_a^b (V(t) - r_0) \times V'(t) dt \right|,$$

где r_0 — произвольная точка в плоскости, в которой расположена область G . Эту точку выбирают, исходя из удобства вычислений, внутри или вне области. Подынтегральная функция есть кусочно-полиномиальная на сетке Δ вектор-функция (вектор-сплайн четвертой степени с гладкостью на единицу меньше, чем гладкость V -сплайна). Учитывая это и переходя к параметру u , формулу площади представим в виде

$$S = \frac{1}{2} \left| \sum_{i=1}^{N-1} \int_0^1 (V(u) - r_0) \times V'(u) du \right|.$$

Отсюда используя представление V -сплайна (114) и обозначение $\Delta V_i = V_{i+1} - V_i$, получаем

$$S = \frac{1}{2} \left| \sum_{i=1}^{N-1} \left[(V_i - r_0) \times \Delta V_i + \frac{\Delta t_i}{5} (V_i - V_{i+1}) \times \Delta V_i - \frac{\Delta t_i^2}{30} V_i \times V_{i+1} \right] \right|. \quad (168)$$

Таким образом, площадь фигуры явно выражена через коэффициенты сплайна.

Площадь сплайновой поверхности на сетке $\Delta = \Delta_t \times \Delta_s$.
Площадь находят по формуле

$$S = \int_a^b \int_a^a |D^1, 0W(t, s) \times D^0, 1W(t, s)| dt ds.$$

Интеграл, стоящий справа, может быть определен численно с помощью одной из стандартных процедур численного интегрирования, входящих в состав программного обеспечения ЭВМ.

Рассмотрим один из простейших алгоритмов вычисления площади поверхности, основанный на триангуляции. Координатные линии на поверхности, соответствующие линиям сетки Δ , заменяем вписанными ломаными, а пары узлов (t_i, s_j) и (t_{i+1}, s_{j+1}) тоже соединяем отрезками прямых. Таким образом, получается многогранник с тре-

угольными гранями, аппроксимирующий заданную поверхность. Площадь многогранника есть первое приближение площади поверхности. Последующие приближения получаются удвоением на каждом шаге числа узлов каждой из сеток Δ_t и Δ_s . При таком приближении площади обеспечивается точность порядка $O(h/\sin \gamma)$, где γ — минимальный угол треугольников, а h — максимальная длина их сторон [8].

Задача о нахождении площади произвольной области G на сплайновой поверхности сводится к предыдущей. Области G в плоскости параметров t и s соответствует область \tilde{G} , ограниченная замкнутым контуром $\rho(\tilde{s})$. Область \tilde{G} разбиваем на простые подобласти \tilde{G}_k , $k=1, \dots, K$, каждую из которых отображаем на прямоугольник $\tilde{\Omega}_k$. После этого площадь области G подсчитывают по формуле

$$S = \sum_{k=1}^K \iint_{\tilde{\Omega}_k} |W_{\tilde{t}}[t, (\tilde{t}, \tilde{s}), s(\tilde{t}, \tilde{s})] W_{\tilde{s}}[t(\tilde{t}, \tilde{s}), s(\tilde{t}, \tilde{s})]| d\tilde{t} d\tilde{s}.$$

Приведем пример построения упомянутого отображения. Пусть область \tilde{G}_k звездная, т. е. внутри области существует хотя бы одна точка r_0 такая, что любой выходящий из нее луч пересекает границу области ровно в одной точке. Требуемое отображение можно записать в виде $r(\tilde{t}, \tilde{s}) = (1 - \tilde{t}) r_0 + \tilde{t} \rho(\tilde{s})$. Другие приемы отображения можно найти, например, в работе [1].

Объем тела, ограниченного сплайновыми поверхностями $W_k(t, s)$, $t, s \in \Omega_k$, $k=1, \dots, K$. Объем выражается формулой

$$Q = \left| \frac{1}{3} \sum_{k=1}^K \iint_{\Omega_k} (W_k(t, s) - r_0) \times [D^{1,0} W_k(t, s) \times D^{0,1} W_k(t, s)] dt ds \right|,$$

где r_0 — некоторая произвольная точка пространства, выбираемая исходя из удобства вычислений. Подынтегральные выражения представляют собой кусочно-полиномиальные вектор-функции восьмой степени по каждой из переменных. Поэтому каждый интеграл можно записать в явном виде через коэффициенты соответствующего W -сплайна. Однако вследствие громоздкости это, вряд ли, целесообразно делать. Проще воспользоваться стандартными процедурами численного интегрирования, как и при вычислении площади поверхности.

Отметим один частный случай. Пусть $r = W(t, s)$ — сплайновая поверхность такая, что ее координатные линии одного из семейств (например, $s = \text{const}$) суть замкнутые кривые, лежащие во взаимно параллельных плоскостях. Объем тела, ограниченного этой поверхностью и двумя плоскостями, соответствующими значениям $s = s_1$ и $s = s_2$ параметра s , представляется в виде интеграла

$$Q = \int_{s_1}^{s_2} S(\zeta) d\zeta,$$

где $S(\zeta)$ — площадь плоской фигуры, ограниченной кривой $\Gamma(t) = W(t, \zeta)$, выражается явно через коэффициенты W -сплайна по формуле (168). Таким образом, задача об определении объема сводится здесь к однократному интегрированию.

2. КРАТЧАЙШЕЕ РАССТОЯНИЕ МЕЖДУ ЭЛЕМЕНТАРНЫМИ ГЕОМЕТРИЧЕСКИМИ ОБЪЕКТАМИ

В этом и следующем параграфах речь пойдет о взаимном расположении двух геометрических объектов. Одной из характеристик взаиморасположения объектов является кратчайшее расстояние между ними. Под кратчайшим расстоянием $D(C_1, C_2)$ между объектами C_1 и C_2 будем понимать минимум функции $d(C_1, C_2)$ — расстояния в парах точек, из которых одна принадлежит C_1 , а другая — C_2 . Как было отмечено в п. 2 гл. II, эта величина характеризует локальную удаленность объектов.

Элементарные геометрические объекты. К элементарным геометрическим объектам условно отнесем объекты достаточно простой структуры. С каждым из них свяжем некоторые данные, однозначно определяющие сам объект и его местоположение в пространстве относительно какой-либо декартовой системы координат:

- точка в пространстве задается своим радиус-вектором \mathbf{r} ;
- прямая — своей точкой \mathbf{r} и направляющим вектором $\boldsymbol{\tau}$;
- плоскость определяется положением \mathbf{r} одной из ее точек и вектором нормали \mathbf{n} ;
- окружность в пространстве задается радиус-вектором ее центра \mathbf{r} , нормалью \mathbf{n} плоскости, в которой она расположена, и ее радиусом R ;
- сфера — положением \mathbf{r} ее центра и радиусом R ;
- круговой цилиндр определяется своей осью $\{\mathbf{r}, \boldsymbol{\tau}\}$ и радиусом R ;
- круговой конус задается положением \mathbf{r} своей вершины, направлением $\boldsymbol{\tau}$ оси и углом полураствора α ;
- отрезок кривой — гладкая сплайновая кривая $\mathbf{r} = \mathbf{V}(t)$;
- поверхность вращения описывается положением $\{\mathbf{r}, \boldsymbol{\tau}\}$ своей оси и отрезком кривой — образующей;
- сегмент поверхности — гладкая сплайновая поверхность $\mathbf{r} = \mathbf{W}(t, s)$.

Для большинства из этих объектов величина D легко подсчитывается (табл. 6). Если же одним из объектов является кривая или поверхность, то определение кратчайшего расстояния между ними сводится к минимизации некоторой положительной функции в ограниченной области.

Минимизация функции одной переменной. Пусть $f(t)$, $t \in [a, b]$ есть положительная функция, принадлежащая классу $C^2[a, b]$. Требуется найти: а) точку локального минимума функции $f(t)$, если известно ее приближение $t^{(1)}$; б) наименьшее значение $f(t)$ на $[a, b]$; в) все нулевые минимумы $f(t)$.

ТАБЛИЦА 6. Формулы кратчайшего расстояния
для элементарных геометрических объектов

Тип объекта		Данные	$D(C_1, C_2)$
C_1	C_2		
Точка, r_1	Точка Прямая Плоскость Сфера Круговой цилиндр Круговой конус	r_2 $\{r_2, \tau\}$ $\{r_2, n\}$ $\{r_2, R\}$ $\{r_2, \tau_2, R\}$ (r_2, τ_2, α)	$ r_1 - r_2 \equiv \Delta r $ $ \Delta r \times \tau $ $ (\Delta r n) $ $ \Delta r - R $ $ \Delta r \times \tau_2 - R $ $ [\tau_2 \times \Delta r \times \tau_2] \sin \alpha + \tau_2 \cos \alpha \Delta r $
Прямая (r_1, τ_1)	Прямая Сфера Круговой цилиндр	$\{r_2, \tau_2\}$ $\{r_2, R\}$ $\{r_2, \tau_2, R\}$	$ \Delta r, \tau_1, \tau_2 $ при $(\tau_1 \times \tau_2) \neq 0$ $(\Delta r \times \tau_1 - R)_+$ $(\Delta r, \tau_1, \tau_2 - R)_+$ при $(\tau_1 \times \tau_2) \neq 0$
Сфера $(r_1 R)$	Плоскость Круговой цилиндр	$\{r_2, n\}$ $\{r_2, \tau_2, R_2\}$	$(\Delta r n - R)_+$ $(\Delta r \times \tau_2 - R - R_2)_+$

1. Рассмотрим один из возможных алгоритмов определения точки локального минимума функции $f(t)$ на отрезке $[a, b]$. В основе алгоритма лежит метод Ньютона, о котором уже шла речь в п. 7 гл. I. Напомним, что строится последовательность $\{t^{(k)}\}$ приближений $t^{(k+1)} = t^{(k)} + \alpha_k \Delta t_k$, где $\Delta t_k = -f'(t^{(k)})/f''(t^{(k)})$. Значения α_k ($0 < \alpha_k \leq 1$) выбирают максимальными, но вместе с тем обеспечивающими сходимость последовательности $\{t^{(k)}\}$ к точке локального минимума t^* функции $f(t)$ при ограничении $a \leq t^{(k)} \leq b$. Если $f''(t) > 0$ для всех $t \in [a, b]$, то такая последовательность существует и сходится к значению t^* со сверхлинейной или квадратичной скоростью в зависимости от свойств функции $f(t)$. Отдельные варианты метода Ньютона различаются способом выбора величины α_k .

В рассматриваемом алгоритме дополнительно (по сравнению с методом Ньютона) в тех точках, где $f'_k = f''(t^{(k)}) \leq 0$, направление шага Δt_k выбирается обратным к направлению градиента функции $f(t)$. Если при этом $f'_k = 0$, то принимается $\Delta t_k = \Delta t_{k-1}$. В целом алгоритм расчета k -го шага можно представить в виде следующих правил вычислений;

$$\rho_k = \begin{cases} -f'_k/f''_k, & \text{если } f'_k \neq 0, f''_k > 0; \\ -\text{sign}(f'_k) |\Delta t_{k-1}|, & \text{если } f'_k \neq 0, f''_k \leq 0; \\ \Delta t_{k-1}, & \text{если } f'_k = 0, f''_k \leq 0; \end{cases}$$

$$\omega_k = \begin{cases} \rho_k, & \text{если } |\rho_k| \leq |\Delta t_{k-1}|; \\ \text{sign}(\rho_k) |\Delta t_{k-1}|, & \text{если } |\rho_k| > |\Delta t_{k-1}|; \end{cases}$$

$$\Delta t_k = \begin{cases} \omega_k, & \text{если } (t^{(k)} + \omega_k) \in [a, b]; \\ a - t^{(k)}, & \text{если } (t^{(k)} + \omega_k) < a; \\ t^{(k)} - b, & \text{если } (t^{(k)} + \omega_k) > b; \end{cases}$$

$$\alpha_k \begin{cases} = 1, & \text{если } \Delta t_{k-1} \Delta t_k > 0 \text{ или } |\Delta t_k| < |\Delta t_{k-1}|; \\ < 1 & \text{в остальных случаях.} \end{cases}$$

Итерационный процесс заканчивается, как только на очередном шаге удовлетворится условие $|\Delta t_k| \leq \eta$. Константа η устанавливает допустимую погрешность в определении точки минимума, и ее выбирают с учетом свойств минимизируемой функции $f(t)$ и известной погрешности вычислений.

Кроме η и начального приближения $t^{(1)}$ в число исходных данных алгоритма входит значение $\rho = \Delta t_0$. Эта величина задает «разрешающую способность» вычислительного процесса, т. е. способность различать близкие точки локальных минимумов функции $f(t)$. Например, если $\rho \leq \hat{\rho}$, где $\hat{\rho}$ — минимальное расстояние между точками локальных максимумов и минимумов $f(t)$, то алгоритм определит точку локального минимума t^* , и можно утверждать, что между значениями $t^{(1)}$ и t^* нет других локальных минимумов функции $f(t)$. Это свойство алгоритма, как и его сходимость, обеспечивается тем, что $|\Delta t_k| \leq |\Delta t_{k-1}|$, причем $|\Delta t_k| < |\Delta t_{k-1}|$ при $\Delta t_k \Delta t_{k-1} < 0$. Сходимость имеет место и для функций с кусочно-непрерывными первыми и вторыми производными.

2. Наименьшее значение f функции $f(t)$ ищем следующим образом. По заданной величине ρ на отрезке $[a, b]$ строим сетку $\Delta: t_{i+1} = a + i\rho, i = 0, \dots, N-2, t_N = b$ и выделяем такие участки $[t_j, t_{j+1}]$, что $f_j < 0, f_{j+1} \geq 0$. На каждом из них находим локальный минимум $f(t_j^*)$ функции с начальным приближением $t_j^{(1)} = (t_j + t_{j+1})/2$. Окончательно в качестве решения принимаем $f = \min [f(a), f(b), \{f(t_j^*)\}]$. Если $\rho \leq \hat{\rho}$, то f есть глобальный минимум функции $f(t)$.

3. Изложенный прием дает возможность найти точки нулевых минимумов. Для этого достаточно из множества точек $\{t_j^*, a, b\}$ выбрать те, в которых значения функции $f(t)$ не превышают допустимую погрешность $\varepsilon > 0$.

Минимизация функций многих переменных. Пусть $q = \{q_1, q_2, \dots, q_m\}^T$ — вектор-столбец и $F(q), q \in \Omega: \{a_j \leq q_j \leq b_j, j = 1, \dots, m\}$, — положительная дважды непрерывно дифференцируемая по каждой из переменных функция. Требуется найти: 1) точку q^* локального минимума функции $F(q)$ в области Ω , если известно ее приближение $q^{(1)}$; 2) наименьшее значение \bar{F} ; 3) точки нулевых минимумов функции $F(q)$ в области Ω .

Рассмотрим один из алгоритмов решения задачи 1 — определения локального минимума функции $F(q)$, который является обобщением алгоритма минимизации функции одной переменной. Строим последовательность точек: $q^{(k+1)} = q^{(k)} + \delta_k \Delta q^{(k)}, k = 1, 2, \dots$,

где $\delta_h > 0$, $|\Delta q^{(k)}| = |(\Delta q^{(k)})^T \Delta q^{(k)}| = 1$. Вводим градиент функции $F(q)$ в точке $q = q^{(k)}$:

$$F'_k = \left\{ \frac{\partial F}{\partial q_j} \Big|_{q=q^{(k)}}, j = 1, \dots, m \right\},$$

матрицу, составленную из значений вторых производных этой функции,

$$F''_k = \left\{ \frac{\partial^2 F}{\partial q_j \partial q_l} \Big|_{q=q^{(k)}}, j, l = 1, \dots, m \right\},$$

и обратную матрицу $(F''_k)^{-1}$. Тогда, если $F'_k \neq 0$ и $(\Delta q^{(k)})^T F'_k < 0$, то направление шага определяется в виде $\Delta q^{(k)} = -\langle (F''_k)^{-1} F'_k \rangle$, что соответствует направлению шага в методе Ньютона. Если матрица $(F''_k)^{-1}$ не существует или $(\Delta q^{(k)})^T F'_k \geq 0$ при $F'_k \neq 0$, то принимается $\Delta q^{(k)} = -\langle F'_k \rangle$, что соответствует градиентному методу. В случае, когда $F'_k = 0$, выбирают такое направление $\Delta q^{(k)}$, при котором величина $A = (\Delta q^{(k)})^T F''_k \Delta q^{(k)}$ принимает минимальное отрицательное значение, или, если $A > 0$ для всех Δq , то полагают $\Delta q^{(k)} = \Delta q^{(k-1)}$. Полученное направление модифицируется с учетом границ области Ω : $\Delta q_j^{(k)} = 0$, $j = 1, \dots, m$, если $q_j^{(k)} = a_j$, $\Delta q_j^{(k)} \times (b_j - a_j) < 0$ или $q_j^{(k)} = b_j$, $\Delta q_j^{(k)} (b_j - a_j) > 0$.

Шаг δ_h в направлении Δq_h подсчитывают по тем же правилам, что и шаг для функции одной переменной:

$$\rho_h = \begin{cases} -\Delta q^{(k)} F'_k / A, & \text{если } A > 0; \\ \delta_{h-1}, & \text{если } A \leq 0; \end{cases}$$

$$\omega_h = \begin{cases} \rho_h, & \text{если } \rho_h < \delta_{h-1}; \\ \delta_{h-1}, & \text{если } \rho_h \geq \delta_{h-1} \text{ и } (\Delta q^{(k)})^T \Delta q^{(k-1)} \geq 0; \\ e\delta_{h-1}, & \text{если } \rho_h \geq \delta_{h-1} \text{ и } (\Delta q^{(k)})^T \Delta q^{(k-1)} < 0; \end{cases}$$

$$\delta_h = \alpha_h \omega_h; \quad e < 1.$$

Константа α_h ($0 < \alpha_h \leq 1$) выбирается максимально возможной, но такой, чтобы $q^{(k)} + \delta_h \Delta q^{(k)} \in \Omega$. Итерационный процесс заканчивается, как только значение очередного шага δ_h окажется меньше допустимого значения η .

Для приведенного алгоритма остаются в силе все замечания о сходимости и условиях применения, сделанные в случае минимизации функции одной переменной, с той лишь разницей, что $F(q)$ должна иметь теперь непрерывные первые частные производные. Отметим, что здесь в роли параметра, задающего «разрешающую способность» алгоритма, выступает величина $\rho = \delta_0$.

Задачи 2 и 3 могут быть решены с помощью следующего приема. Выбирается множество точек $\{q^{(v)}\}$, $v = 1, \dots, M$, равномерно распределенных в области Ω . Для каждой из них иницируется алгоритм поиска локального минимума $\tilde{q}^{(v)}$ функции $F(q)$ при достаточно большом значении константы η . Из всех $\{\tilde{q}^{(v)}\}$ уточняются

ТАБЛИЦА 7. Функция расстояния между кривой или поверхностью $r_1 = \{V_1(t), W_1(u, v)\}$ и элементарными геометрическими объектами

C_2		$d(r_1, C_2)$
Тип	Данные	
Точка	r	$ r_1 - r \equiv \Delta r(t) $ (или $\Delta r(u, v)$)
Прямая	$\{r, \tau\}$	$ \Delta r(t) \times \tau $
Плоскость	$\{r, n\}$	$ \Delta r(t) n $
Круговой цилиндр	$\{r, \tau, R\}$	$ \Delta r(t) \times \tau - R $
Круговой конус	$\{r, \tau, \alpha\}$	$ [(\tau \times \Delta r(t) \times \tau) \sin \alpha + \tau \cos \alpha] \Delta r(t) $
Сфера	$\{r, R\}$	$ \Delta r(t) - R $
Кривая	$V_1(s)$	$ r_1 - V_1(s) $
Поверхность	$W_1(t, s)$	$ W_1(t, s) - r_1 $

только те, расстояния между которыми превышает ρ . Число M не следует брать больше $\left[\prod_{j=1}^m (b_j - a_j) / 2\rho \right]$.

Кратчайшее расстояние $D(C_1, C_2)$ между двумя геометрическими объектами. Если один из объектов есть кривая или поверхность, то $D(C_1, C_2)$ определяют путем минимизации функции расстояния $d(C_1, C_2)$.

Вид этой функции для конкретных объектов представлен в табл. 7.

Константу η — ограничение на допустимый шаг минимизации — выберем из следующих соображений. Пусть $\varepsilon > 0$ — допустимая абсолютная погрешность в определении кратчайшего расстояния — и $V(t^*)$ — точка на кривой, которая соответствует одному из локальных минимумов функции d . Тогда процесс движения к этому минимуму считаем законченным, если для очередной точки $t^{(k)}$ выполнится условие $|V(t^{(k)}) - V(t^*)| \leq \varepsilon$. Представим $V(t^*)$ в этом неравенстве в виде ряда Тейлора по степеням $\delta t_k = (t^{(k)} - t^*)$, ограничившись первыми двумя членами. В результате получим $|V'(t^{(k)})| |\delta t_k| \leq \varepsilon$ или $|\delta t_k| \leq \varepsilon / |V'(t^{(k)})|$. Таким образом, можно принять, что $\eta = \varepsilon / \|V'(t)\|$. Однако оценка величины $\|V'(t)\|$ не всегда известна, поэтому вместо нее имеет смысл использовать текущее значение $|V'(t^{(k)})|$, т. е. величину η корректировать на каждом шаге процесса минимизации.

Аналогично рассуждая в случае, когда ищется кратчайшее расстояние элементарных геометрических объектов до поверхности, для значения η получим выражение

$$\eta(u^{(k)}, v^{(k)}) = \varepsilon (|W_u(u^{(k)}, v^{(k)})|^2 + |W_v(u^{(k)}, v^{(k)})|^2)^{-1/2}.$$

Учитывая, что звенья V -сплайна и W -сплайна имеют достаточно плавный характер, значение ρ во многих случаях можно принимать равным минимальному шагу сетки ΔV -сплайна или минимальному диаметру областей Ω_{ij} в случае W -сплайна.

3. ПЕРЕСЕЧЕНИЕ ЭЛЕМЕНТАРНЫХ ГЕОМЕТРИЧЕСКИХ ОБЪЕКТОВ

Если два геометрических объекта C_1 и C_2 имеют общие точки, то говорят, что они пересекаются, а множество общих точек этих объектов принято называть их пересечением. Кратчайшее расстояние между пересекающимися объектами равно нулю. Пересечение объектов может содержать конечное число точек, линии и даже куски поверхности. Практический интерес представляют первые два случая из указанных.

Расчет точек пересечения двух геометрических объектов C_1 и C_2 . Он состоит в решении уравнения $D(C_1, C_2) = 0$. Конкретный вид его левой части для некоторых элементарных объектов представлен в табл. 6. В качестве примера приведем решение этого уравнения для случаев, когда их можно выписать в явном виде. Выражение для точки r^* пересечения двух компланарных прямых $P_1 = \{r_1, \tau_1\}$ и $P_2 = \{r_2, \tau_2\}$ есть

$$r^* = r_2 + (|(r_2 - r_1) \times \tau_1| / |\tau_1 \times \tau_2|) \tau_2.$$

Прямая P_1 и плоскость $S = \{r_2, n\}$ пересекаются в точке

$$r^* = r_1 - [(r_1 - r_2), n] \tau_1 / (\tau_1, n).$$

Для прямой P_1 и сферы $C = \{r_2, R\}$

$$r^* = r_2 + (\tau_1 \times \Delta r \times \tau_1) \pm \sqrt{R^2 - (\Delta r \times \tau_1)^2} \tau_1,$$

где $\Delta r = r_1 - r_2$.

Точки пересечения кривой $V(t)$ или поверхности $W(t, s)$ с элементарными геометрическими объектами рассчитывают, как точки нулевых локальных минимумов функции расстояния d . Некоторые из таких функций приведены в табл. 7.

Расчет дуги кривой как пересечения двух поверхностей $W_1(t, s)$, $t, s \in \Omega_1$ и $W_2(u, v)$, $u, v \in \Omega_2$. Он заключается в отыскании последовательности ее точек $\{r_i\}$, $i = 1, \dots, N$, обеспечивающей восстановление самой кривой V-сплайном (эрмитовым или класса C^3) с заданной точностью ϵ . Алгоритм решения этой задачи включает три основных этапа: 1) поиск начальной точки; 2) выбор текущего шага; 3) расчет текущей точки.

1. В качестве начальной точки r_1 искомой кривой $r = r(\sigma)$ (σ — лонгальный параметр) принимается одна из точек пересечения крайних дуг одной поверхности с другой поверхностью. Если такой точки не существует, т. е. если кривая $r(\sigma)$ не выходит на край ни одной из поверхностей, то поступаем следующим образом. Определим любую точку пересечения поверхностей W_1 и W_2 как точку $q = \{t^*, s^*, u^*, v^*\}$ нулевого минимума функции расстояния $d(q) = |W_1(t, s) - W_2(u, v)|^2$. Область параметров, например Ω_1 , делим на две подобласти $\Omega_1^{(1)}$ и $\Omega_1^{(2)}$ линиями $t = t^* = \text{const}$ или $s = s^*$. Из этих двух линий следует выбрать первую, если $|\langle W_t(t^*, s^*) \tau_1 \rangle| > |\langle W_s(t^*, s^*) \tau_1 \rangle|$, где $\tau_1 = \langle n_1 \times n_2 \rangle$ — вектор касательной к искомой кривой в точке r_1 ; n_1, n_2 — векторы нормали к поверхностям

W_1 и W_2 в этой точке. Таким образом приходим к задаче о пересечении одной поверхности с двумя сегментами другой. Для каждого из сегментов известна начальная точка, расположенная на их общей границе. Последняя операция, связанная с начальной точкой, — это выбор константы $e = \pm 1$, определяющей направление вектора касательной $\tau_1 = e \langle n_1 \times n_2 \rangle$ соответственно направлению движения из точки r_1 к следующей точке.

2. Пусть $r_i = r(\sigma_i)$ — очередная точка кривой, τ_i и k_i — вектор касательной и значение кривизны в этой точке. Если $k_1^{(i)}$ и $k_2^{(i)}$ — нормальные кривизны поверхностей W_1 и W_2 в точке r_i в направлении τ_i , то выражение для k_i имеет вид

$$k_i = |k_1^{(i)} n_2^{(i)} + k_2^{(i)} n_1^{(i)}| / |n_1^{(i)} \times n_2^{(i)}|.$$

Шаг $\Delta\sigma_i$ определим из оценки приближения кривых V-сплайном:

$$\Delta\sigma_i = k_0 \sqrt{\varepsilon \|k\|},$$

приняв $\|k\| = \max\{k_{i-1}, k_i\}$.

3. Представим вектор $\Delta\sigma_i \tau_i$ в виде

$$\Delta\sigma_i \tau_i = \Delta t_i W_1^{(1,0)}(t_i, s_i) + \Delta s_i W_1^{(0,1)}(t_i, s_i);$$

$$\Delta\sigma_i \tau_i = \Delta u_i W_2^{(1,0)}(u_i, v_i) + \Delta v_i W_2^{(0,1)}(u_i, v_i).$$

Точке $r_i + \Delta\sigma_i \tau_i$ в области параметров $\Omega = \Omega_1 \times \Omega_2$ соответствует точка $q_{i+1} = q_i + c_i \Delta q_i$, где $\Delta q_i = \{\Delta t_i, \Delta s_i, \Delta u_i, \Delta v_i\}$. Константа c_i , ($0 < c_i \leq 1$) выбирается максимально возможной, но такой, что $q_{i+1} \in \Omega$. Точка r_{i+1} определяется как точка нулевого минимума функции $d(q)$ в окрестности q_{i+1} . Кривая будет считаться полностью определенной, как только для точки r_N окажется выполненным равенство $c_N = 0$.

4. РАЗВЕРТКИ И КВАЗИРАЗВЕРТКИ ПОВЕРХНОСТЕЙ

В инженерной геометрии рассматривается задача нахождения формы плоского листа (металла, кожи, ткани) толщиной d , из которого путем деформаций можно получить кусок поверхности нужной формы. Достаточно точно эта задача формулируется в рамках теории упругопластических деформаций. Мы рассмотрим чисто геометрическую задачу о соответствии области на поверхности некоторой плоской области при условии, что деформации (растяжения и сдвиги) не превышают некоторых заданных значений. Расчеты ведутся либо относительно одной из сторон плоского листа, либо относительно его срединной плоскости. Неоднородность деформаций по толщине листа не учитывается.

В дифференциальной геометрии изучается такое соответствие двух поверхностей, при котором одна налагается на другую путем изгибания без растяжений или сжатий и сдвигов, т. е. с сохранением длин линий и углов между ними. Наглядное представление об этом процессе дает изгибание листа бумаги. Плоскую область, соответствующую области на поверхности, в инженерной геометрии стали

называть *разверткой*. (В дифференциальной геометрии термин «развертка» употребляется в ином смысле).

Развертки существуют не для всяких поверхностей, а только для весьма узкого их класса. В общем случае приходится строить отображение куска поверхности на плоскую область с искажением длин или углов или того и другого одновременно. Если имеется описание поверхности в виде вектор-функции $\mathbf{r} = \mathbf{r}(t, s)$ с границей $s = g(t)$, то его можно рассматривать как задание развертки. Так строятся географические карты в проекции Меркатора, где t и s — углы долготы и широты. Меридианы и параллели переходят в прямые линии $t = \text{const}$ и $s = \text{const}$ соответственно. При этом углы между линиями на сфере и на плоскости сохраняются.

В данном примере с точностью до масштаба сохраняются длины экватора и меридианов, а длины параллелей увеличиваются. Для поверхности общего вида можно добиться той же картины с увеличением или сокращением длин линий второго семейства. Вследствие возможных больших искажений этот способ в технологии в чистом виде не всегда применим. Ниже мы предлагаем методы введения поправок, позволяющие в случае необходимости уменьшить искажение длин линий второго семейства за счет изменения углов. Получающиеся образы обладают некоторыми свойствами разверток и называются *условными развертками* и *квазиразвертками*. При этом искажения длин линий и углов можно оценить, что позволяет выбирать технологически приемлемые варианты.

Развертки. задается параметризация поверхности $\mathbf{r} = \mathbf{r}(t, s)$. Известно, что первая квадратичная форма $I(I)$, характеризующая длины линий на ней (гл. II, п. 3), не изменяется при изгибании. В силу неизменности формы сохраняются и углы между линиями на поверхности.

При изгибании поверхности кривизны в каждой точке меняют свои значения, но произведение главных кривизн — гауссова кривизна — остается неизменной, и это фундаментальное положение внутренней геометрии поверхностей. Следовательно, если поверхность путем изгибания налагается на плоскость, гауссова кривизна которой равна нулю, то и гауссова кривизна поверхности равна нулю, т. е. она является развертывающейся поверхностью. Верно и обратное. Таким образом, это единственный класс поверхностей, допускающих развертки на плоскость.

Свойством инвариантности относительно изгибания поверхностей обладает геодезическая кривизна кривых на поверхности. В частности, при наложении развертывающейся поверхности на плоскость кривизна плоской кривой совпадает с геодезической кривизной ее образа на поверхности.

Геодезические линии как линии нулевой геодезической кривизны отображаются в прямые на плоскости. На этом факте основан графический способ построения разверток по известному эталону поверхности. Он состоит в том, что измеряются внутренние расстояния между узлами каркаса, которые откладываются в виде отрезков прямых на плоскости. При этом для развертывающейся поверхности

можно сохранить углы между отрезками и получить развертку в точном смысле. Для поверхностей общего вида сохранить значения углов можно лишь в окрестности одной точки. При графоаналитических методах, когда нет эталона, обычно вычисляют лишь одну опорную геодезическую линию и длины линий каркаса поверхности. Каркасные линии тоже отображаются в виде прямых, что приводит к погрешностям. Мы рассматриваем чисто аналитический способ построения отображения куска поверхности на плоскость.

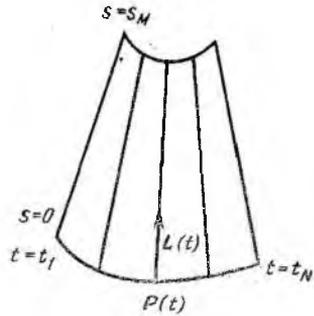


Рис. 43. Развертка

Возьмем уравнение развертывающейся поверхности

$$r(t, s) = \rho(t) + s l(t). \quad (169)$$

Коэффициенты ее первой квадратичной формы суть

$$E = \rho'^2 + 2s\rho'l' + s^2l'^2; \quad F = \rho'l + sl', \quad G = l^2.$$

Коэффициенты второй квадратичной формы $L \neq 0$, $M = l'n = 0$, $N = 0$.

Образ при развертывании поверхности на плоскость (X, Y) будем искать в виде некоторой параметризации

$$R(t, s) = P(t) + sL(t) \quad (170)$$

плоской области, где $P(t)$ — образ $\rho(t)$, а $sL(t)$ — образ $s l(t)$ (рис. 43).

Будем добиваться того, чтобы кривизна линии $P(t)$ была равна геодезической кривизне $\kappa(t)$ линии $\rho(t)$. Если $\beta(t)$ — угол между векторами ρ' и l , то $0 < \beta(t) < \pi$, и орт нормали к поверхности будет $n = (\rho' \times l) / |\rho' \parallel l \sin \beta$. Значит, согласно (113)

$$\kappa(t) = \frac{\rho'' \rho' (\rho' \times l)}{|\rho'|^3 |l| \sin \beta}. \quad (171)$$

Кривизне $\kappa(t)$ припишем знак скалярного произведения $\rho'' l$.

Если взять орты касательной и нормали к линии $P(t)$ в виде

$$T(t) = \cos \alpha(t) e_1 + \sin \alpha(t) e_2; \quad N(t) = -\sin \alpha(t) e_1 + \cos \alpha(t) e_2,$$

то согласно формуле Френе $T'(t) t'(\sigma) = \kappa(t) N(t)$ следует положить $\alpha'(t) = |\rho'(t)| \kappa(t)$, и тогда

$$\alpha(t) = \alpha_1 + \int_{t_1}^t |\rho'(t)| \kappa(t) dt. \quad (172)$$

Уравнение образа кривой $\rho(t)$ при сохранении длин будет

$$P(t) = P_1 + \int_{t_1}^t |\rho'(t)| T(t) dt. \quad (173)$$

Вектор $L(t)$ образующей на плоскости определяется из условия равенства косинусов углов $\beta(t)$ между направляющей и образующими на поверхности и на плоскости и условия сохранения длин образующих

$$T(t)L(t) = \langle \rho'(t) | I(t) \rangle; |L(t)| = |I(t)|. \quad (174)$$

Решение этих уравнений:

$$L(t) = |I(t)| [\cos(\alpha + \beta) e_1 + \sin(\alpha + \beta) e_2]. \quad (175)$$

Из равенства $|P'(t)| = |\rho'(t)|$ следует $P''(t)P'(t) = \rho''(t)\rho'(t)$, и аналогично из $|\bar{L}(t)| = |I(t)|$ вытекает $L'(t)L(t) = I'(t)I(t)$.

Кривизну плоской кривой $P(t)$ можно также записать в виде (171). В силу ее инвариантности и условий (174) получаем

$$P''(t)L(t) = \rho''(t)I(t); P'(t)L'(t) = \rho'(t)I'(t); \quad (176)$$

$$|L'(t)| = |I'(t)|. \quad (177)$$

Как следует из этих равенств, значения коэффициентов первой и второй квадратичных форм при отображении (170) не изменяются. Формула (170) позволяет для каждой точки $r(t, s)$ развертывающейся поверхности находить соответствующую точку $R(t, s)$ на плоскости, а для каждой линии $r = r[t(\omega), s(\omega)]$ — линию $R = R[t(\omega), s(\omega)]$.

Все формулы принимают особенно простой вид, если s — это лонгальный параметр вдоль образующих, а t — лонгальный параметр вдоль направляющей $s = 0$. Тогда $|\rho'| = |I| = 1$, $\rho''\rho' = I''I' = 0$ и $\kappa(t) = \rho''I/\sin\beta$.

Поскольку функция $\kappa(t)$ в формуле (172) стоит под знаком интеграла, то все выводы справедливы для развертывающихся поверхностей с интегрируемой геодезической кривизной линий. Мы рассматриваем поверхности, описываемые частным видом W -сплайнов. Входящие в (169) функции $\rho(t)$ и $I(t)$ являются V -сплайнами на сетке Δ_t . Для построения развертки требуется по какой-либо квадратурной формуле найти значения трех функций $\alpha(t)$ (172), $P_X(t)$ и $P_Y(t)$ (173) в узлах сетки Δ_t . Например,

$$\alpha(t_i) = \alpha_1 + \sum_{k=1}^{i-1} \int_{t_k}^{t_{k+1}} |\rho'(t)| \kappa(t) dt, \quad i = 2, \dots, N.$$

Аналогичные формулы используют для вычисления компонент векторов $P(t_i)$, после чего определяют векторы $L(t_i)$.

По полученным данным на сетке Δ_t находят интерполяционные V -сплайны. Последние строят как эрмитовы или сплайны класса C^2 в зависимости от того, какими V -сплайнами являются $\rho(t)$ и $I(t)$. Величины $P'(t_i)$ вычисляют уже при использовании квадратурной формулы. Величины $L'(t_i)$ определяют дифференцированием $L(t)$ или приближенно по формулам типа (123).

Условные развертки линейчатых поверхностей. Линейчатая поверхность, как и развертывающаяся, представляется формулой (169). Но теперь векторы ρ' , I' , I не компланарны. Вектор $I' = I'_1 + I'_2$

кроме составляющей l'_t в касательной плоскости к поверхности имеет также ненулевую составляющую l'_n , нормальную к ней. Коэффициенты первой квадратичной формы выглядят так же, как коэффициенты для разворачиваемой поверхности. Коэффициенты второй квадратичной формы $L \neq 0$, $N = 0$, как и выше, но $M = l'_n \neq 0$. Гауссова кривизна поверхности согласно (112) $K = -l_n^2 / (EG - F^2)$.

Найдем для линейчатой поверхности выражение $R(t, s)$. Вычисления проводятся по тем же формулам, что и для разворачиваемой поверхности. В данном случае остаются справедливыми формулы (174), (176) и (177), если в них вместо l' поставить l'_t . Но тогда очевидно, что

$$l'^2(t) - L'^2(t) = l_n^2(t). \quad (178)$$

Формулу (170) можно рассматривать как выражение образа некоторой разворачиваемой поверхности, которая аппроксимирует данную линейчатую. Это — так называемая условная развертка.

Из (174), (176) и (178) следует, что коэффициенты первой квадратичной формы поверхности и ее образа связаны формулами

$$E_r = E_R + s^2 l_n^2; \quad F_r = F_R = F; \quad G_r = G_R = G.$$

Значение первой квадратичной формы $I_R(\tau)$ на единичном векторе τ

$$E_R \lambda_1^2 + 2F \lambda_1 \lambda_2 + G \lambda_2^2 = 1. \quad (179)$$

Это есть уравнение окружности в косоугольных (аффинных) координатах λ_1, λ_2 . Для поверхности в тех же координатах она переходит в эллипс. Разность $\sqrt{I_r(\tau)} - \sqrt{I_R(\tau)}$ дает относительное удлинение в направлении вектора τ .

Предположим, что l_n^2 / l'^2 — величина малая по сравнению с единицей, тогда

$$\begin{aligned} \sqrt{I_r(\tau)} - \sqrt{I_R(\tau)} &= \sqrt{1 + s^2 l_n^2 \lambda_1^2} - 1 = \\ &= \frac{1}{2} s^2 l_n^2 \lambda_1^2 + O(l_n^4). \end{aligned}$$

Это значит, что относительные удлинения в любой точке неотрицательны, т. е. имеют место деформации растяжения. В конкретной технологии деформации ограничены, скажем, величиной ε . С точностью до малых более высокого порядка получаем

$$\frac{1}{2} s^2 l_n^2 \lambda_1^2 \leq \varepsilon.$$

Чтобы использовать эту оценку, нужно найти максимальное значение λ_1^2 при условии (179): $\lambda_{1\max}^2 = G(F_R G - F^2)^{-1}$. Следовательно, ограничение на деформации

$$\partial(t, s) = \frac{s^2 l_n^2 l^2}{2(E_R G - F^2)} \leq \varepsilon. \quad (180)$$

Оно определяет пределы применимости формулы (170) в зависимости от значений параметра s .

Величина $\partial(t, s)$ равна нулю при $s = 0$ и монотонно возрастает как в сторону положительных, так и в сторону отрицательных s . Обозначим γ — угол между векторами l и l' . Предельные случаи деформаций:

$$a) s \rightarrow -\infty, \partial(t, s) \rightarrow l_n^*/2 |l'_t| \sin \gamma;$$

$$б) s \rightarrow s^* = |\rho'| \sin \beta / |l'_t| \sin \gamma, \partial(t, s) \rightarrow \infty,$$

так как величина $E_R G - F^2 = |R_t \times R_s|^2 \rightarrow 0$. Последнее означает, что векторы R_t и R_s коллинеарны. Их коллинеарность есть признак огибающей однопараметрического семейства прямых на плоскости, какими являются образующие $L(t)$ образа линейчатой поверхности. Разумеется, наши построения имеют смысл лишь в конечной области $[s_1, s_M]$ при $s_M < s^*$.

Максимальное значение деформаций $\partial(t, s)$ можно уменьшить, используя один из двух следующих приемов. Во-первых, в качестве направляющей следует принимать не произвольную линию $s = \text{const}$, а линию $s = \bar{s} = 0,5(s_1 + s_M)$. Все расчетные формулы при этом не меняются, если в (170) осуществить перенос начала отсчета по переменной s , что равносильно замене вектора ρ на $\tilde{\rho} = \rho + \bar{s}l$. По сравнению со случаем, когда направляющей является крайняя линия $s = s_1$, наибольшее значение деформации уменьшается. В практически важных задачах, когда $\sin \beta \approx \sin \gamma \approx 1$, а $|l'_t| \ll |\rho'|$, уменьшение получается почти в 4 раза.

Во-вторых, вместо формулы (170) применяют зависимости более сложного вида. Рассмотрим, например, формулу

$$R(t, s) = (1 + q)P(t) + s[L(t) + qL_1(t)]. \quad (181)$$

Здесь $P(t)$ и $L(t)$ имеют тот же смысл, что и в (170). Вектор $L_1(t)$ определим формулой

$$L_1(t) = \varphi(t) [-\sin(\alpha + \beta)e_1 + \cos(\alpha + \beta)e_2], \quad (182)$$

т. е. $L_1(t)$ ортогонален $L(t)$; $q > 0$ — малый параметр порядка $O(l_n^*/l'^2)$. Величина $G_R = G_r$ (с точностью до малых $O(q^2)$).

Далее $F_r = F_R = -qP'(t) |L(t) + L_1(t)| + O(q^2)$. Если положить $\varphi(t) = |l| \text{ctg } \beta$, то $F_r = F_R$ с точностью $O(q^2)$. Остается подсчитать

$$E_r - E_R = s^2 l_n^2 - 2q \{P'^2(t) + sP'(t) [L'(t) + L'_1(t)] + s^2 L'(t) L'_1(t)\} + O(q^2).$$

Повторяя рассуждения, проделанные при выводе формулы (180), получаем, что в таком случае деформации будут

$$\partial(t, s) = l^2 (E_r - E_R) / 2 (E_R G - F^2). \quad (183)$$

На линии $s = 0$ получаем отрицательные деформации, т. е. сжатия. При фиксированном q имеются две линии g_1 и g_2 , на которых деформации равны нулю. Вне области, ограниченной этими линиями и образующими $t = t_1$ и $t = t_N$, деформации будут растяжениями.

Путем выбора q можно варьировать зоны сжатия и растяжения. Максимальная величина $|\partial(t, s)|$ меньше, чем в предыдущем случае. При практическом построении условной развертки по формуле (181) необходимо по данным $L_1(t_i)$ и $L'_1(t_i)$ получить еще один интерполяционный V -сплайн.

Квазиразвертки поверхностей общего вида. Предположим снова, что поверхности близки к развертывающимся. Заданную поверхность $r = r(t, s)$ сравним с некоторой линейчатой поверхностью (169). Последнюю проще всего взять в виде поверхности, касающейся заданной поверхности вдоль координатной линии $s = 0$. Тогда $\rho(t) = r(t, 0)$, $l(t) = r_s(t, 0)$. Линиям $t = \text{const}$ соответствуют образующие линейчатой поверхности при тех же значениях t . Точке a на линии $t = \text{const}$ отвечает точка на образующей $l(t)$, лежащая на перпендикуляре, опущенном из a на $l(t)$.

Расхождение между заданной и линейчатой поверхностями характеризуется вектор-функцией $q(t, s) = r(t, s) - \rho(t) - sl(t)$, причем $q(t, 0) = q_s(t, 0) = 0$. Предполагаем, что $|q(t, s)|$ — малая порядка $O(l_n^2/l'^2)$. Будем аппроксимировать $q(t, s)$ функцией $\psi(s)l_1(t)$. Это можно сделать различными способами. Например, разложим $q(t, s)$ в ряд Тейлора по степеням s . Два первых члена разложения равны нулю, третий примем в качестве аппроксимирующей функции $\psi(s)l_1(t) = 0,5s^2q_{ss}(t, 0)$. Если $q_{ss}(t, 0)$ — разрывная функция, то можно взять параболическую интерполяцию по точкам $q(t, s_1)$, $q(t, 0)$, $q(t, s_M)$. В принятом приближении вектор $l_1(t)$, как и вектор $q(t, s)$, будет ортогонален вектору $l(t)$.

Итак, заданная поверхность представлена в форме

$$r(t, s) = \rho(t) + sl(t) + \psi(s)l_1(t).$$

Ее образ на плоскости (X, Y) ищем в виде

$$R(t, s) = P(t) + sL(t) + \psi(s)L_1(t). \quad (184)$$

Здесь вектор-функции $P(t)$, $L(t)$ и $L_1(t)$ определены формулами (173), (175) и (182). В предположении малости $|l_1(t)| = O(l_n^2/l'^2)$ нетрудно видеть, что $G_R = G_r$. Зафиксируем $\varphi(t) = -\text{sign}(\rho' l_1) |l_{1\tau}|$, где $|l_{1\tau}|$ — проекция вектора $l_1(t)$ на касательную плоскость к линейчатой поверхности. В данном случае остаются справедливыми равенства (174), (176) и (178). Кроме того, учитывая, что

$$\rho' l_1 = \rho' l_{1\tau} = \text{sign}(\rho' l_1) |\rho'| |l_{1\tau}| \sin \beta;$$

$$l' l_1 = l'_1 l_{1\tau} + O[|(l'_n/l')|^3] = -\text{sign}(\rho' l_1) |l'_1| |l_{1\tau}| \sin \gamma + \\ + O[(l_n^2/l'^2)^{3/2}],$$

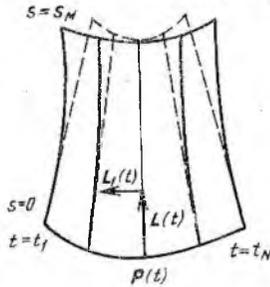
получим формулы (с точностью до малых более высокого порядка)

$$P'(t)L_1(t) = \rho'(t)l_1(t); \quad L'(t)L_1(t) = l'(t)l_1(t).$$

Простые вычисления показывают, что снова $F_R = F_r$. Наконец,

$$E_r - E_R = s^2 l_n^2 + 2\psi(s)[\rho'(t)l_1(t) - P'(t)L_1(t)] + \\ + 2s\psi'(s)[l'(t)l_1(t) - L'(t)L_1(t)].$$

Рис. 44. Квазиразвертка поверхности общего вида



Деформации определяются формулой (183). В общем случае что-либо сказать об их характере нельзя, кроме того, что на линии $s = 0$ они равны нулю (рис. 44). Однако в конкретных ситуациях, как правило, о деформациях можно судить по взаимному расположению заданной и линейчатой поверхностей. Последнее и определяет в основном деформации, тогда как член с $L_1(t)$ в (184) мало влияет на них. Иногда, например, в случае задания поверхности вращения с образующими $t = \text{const}$, он вообще равен нулю, так как $|l_{11}| = 0$. Касательная поверхность будет развертывающейся.

В рассмотренной задаче вместо линейчатой поверхности, касательной к заданной, можно было бы взять секущую поверхность, проходящую через крайние линии $s = s_1$ и $s = s_M$, образующие которой соединяют изопараметрические точки этих линий. Такая замена, естественно, меняет характер деформаций.

Другие возможности решения задачи связаны со способом выбора вектора $L_1(t)$. Мы использовали его, чтобы добиться выполнения равенства $F_r = F_R$, что приводило к деформациям растяжения или сжатия по линиям $s = \text{const}$. Но можно определить $L_1(t)$, например, из условия $E_r = E_R$ на какой-либо линии $s = s^* = \text{const}$. Этот прием обеспечивает сохранение длин двух линий $s = 0$ и $s = s^*$ и близок к графическому способу построения квазиразверток с сохранением длин границ области [17]. Неудобство заключается в том, что для вектора $L_1(t)$ теперь имеем дифференциальное уравнение, интегрирование которого сложнее, чем все остальные расчеты.

Пример. Для линейчатой поверхности (169) при $\varphi(t) = te_1$, $l(t) = 0, 1te_1 + ae_2 + 0,25t^2e_3$ (a — параметр) строятся условные развертки (170) и (181). Пусть S — длина дуги $\gamma(t, a)$, $t \in [-1, 1]$, на поверхности, а S_1, S_2 — длины соответствующих дуг на условных развертках. При $a = 2$ $S = 2,654$, $S_1 = 2,429$. Если в (181) положить $q = 0,05$, то $S_2 = 2,482$. При этом относительное сжатие q дуги $P(t)$ будет примерно равно относительному удлинению линии $R(t, a)$ $\partial = |S - S_2|/S = 0,065$. Для $a = 4$, $q_1 = 0,1$ будем иметь значения $S = 3,573$, $S_1 = 2,826$, $S_2 = 2,921$, $\partial = 0,182$.

Глава V

МАТЕМАТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ МЕТАЛЛОРЕЖУЩИХ СТАНКОВ С ЧИСЛОВЫМ ПРОГРАММНЫМ УПРАВЛЕНИЕМ

Станочник, обрабатывая на станке с ручным управлением серию однотипных деталей, производит, по сути дела, всякий раз уникальное изделие в пределах заданного допуска. Добиться приемлемой идентичности экземпляров тем труднее, чем сложнее геометрия образца. Помимо чисто профессиональных факторов, здесь действует сильный психологический фактор — опасение испортить дорогостоящую деталь. Станки с числовым программным управлением (ЧПУ) свободны от подобных «эмоций» и в этом смысле идеально приспособлены для тиражирования деталей с нестандартными поверхностями. В настоящей главе рассмотрены некоторые вопросы математической подготовки автоматического воспроизведения деталей на программно управляемых станках.

1. РАСЧЕТ ТРАЕКТОРИИ ИНСТРУМЕНТА ПРИ ОБРАБОТКЕ СЛОЖНЫХ ПОВЕРХНОСТЕЙ

Подготовка управляющей программы для станков с ЧПУ выполняется в два этапа. На первом этапе рассчитывают траекторию движения инструмента относительно детали. На втором этапе полученные ранее перемещения инструмента в пространстве координат детали преобразуют в перемещения рабочих органов станка, т. е. определяется траектория в рабочем пространстве координат станка. Затем эта траектория аппроксимируется с заданной точностью функциональной зависимостью, которую реализует конкретное устройство программного управления станком. Чаще всего это кусочно-линейная зависимость. Содержание настоящего параграфа относится к первому из указанных этапов.

Геометрические аспекты формообразования фрезерованием. Движение фрезы в процессе обработки детали складывается из вращения фрезы вокруг ее оси Q и перемещения самой оси в пространстве. Перемещение оси всегда можно представить как движение какой-либо ее точки P и поворот направляющего вектора q оси Q около этой точки. При вращении фрезы вокруг неподвижной оси Q режущая кромка ее зуба образует поверхность вращения F , которую в дальнейшем будем называть рабочей поверхностью фрезы или просто поверхностью фрезы. Для определенности примем, что она гладкая.

Будем исходить из следующей гипотезы: в процессе обработки в каждый момент времени рабочая поверхность фрезы своей внутренней точкой T соприкасается с поверхностью детали так, что точка соприкосновения является единственной общей точкой детали и фрезы. В частном случае возможно соприкосновение вдоль целого отрезка

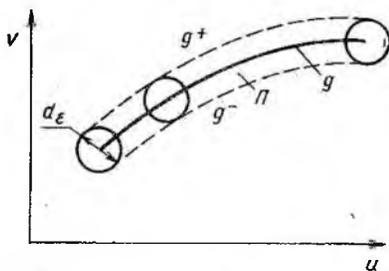


Рис. 45. Полоса воспроизведения

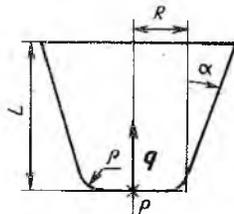


Рис. 46. Параметры фрезы

образующей поверхности F . Эти условия накладывают определенные ограничения как на геометрические параметры применяемого инструмента, так и на его ориентацию в пространстве при контакте с поверхностью детали. На основании этой гипотезы каждой точке $g(u, v)$, $u, v \in \Omega$, обрабатываемой поверхности G можно сопоставить положение фрезы, определив тем самым в области Ω вектор-функции $P = P(u, v)$ и $q = q(u, v)$. Они задают положение точки P (в дальнейшем ее будем называть расчетной точкой фрезы) и направление оси Q фрезы. Установленное таким образом соответствие назовем *связью инструмента с деталью*. Практический интерес представляют такие связи, для которых $P(u, v)$ и $q(u, v)$ обладают достаточно высоким порядком гладкости. Так, предположим, что они кусочно дважды дифференцируемы в Ω .

Рассмотрим некоторое положение фрезы $\{P(u^*, v^*), q(u^*, v^*)\}$, определенное заданной связью с деталью. Ее рабочая поверхность в окрестности точки соприкосновения $g(u^*, v^*)$ аппроксимирует поверхность G детали. Обозначим через $d_\epsilon(u^*, v^*)$ максимально возможный диаметр круга с центром в точке (u^*, v^*) области Ω , в пределах которого погрешность аппроксимации не превышает значения $\epsilon > 0$. Траектория движения фрезы с учетом заданной связи с деталью записывается в виде соотношений: $P(t) = P(u(t), v(t))$; $q(t) = q(u(t), v(t))$, $t \in [0, T]$, которые определяют положение фрезы в пространстве в зависимости от времени t . Функции $u = u(t)$ и $v = v(t)$ задают некоторую линию g в области Ω и, соответственно, кривую $g = g(u(t), v(t))$ на поверхности детали, являющуюся геометрическим местом точек соприкосновения с рабочей поверхностью фрезы. Из каждой точки линии g как из центра проведем окружность диаметром $d_\epsilon(u(t), v(t))$. Огибающие семейства этих окружностей g^- и g^+ ограничивают некоторую полосу Π , содержащую линию g (рис. 45). Из определения величины d_ϵ следует, что движением фрезы по траектории, задаваемой в области Ω линией g , воспроизводится полоса поверхности G с точностью ϵ , соответствующая полосе Π . Такие полосы на поверхности G обычно называют *строками*. Существенным параметром обработки является ширина строки. Полосу Π будем называть *полосой воспроизведения*.

Геометрия фрезы. Она характеризуется параметрами меридианного сечения ее рабочей поверхности: R, ρ, L, α (рис. 46), где R —

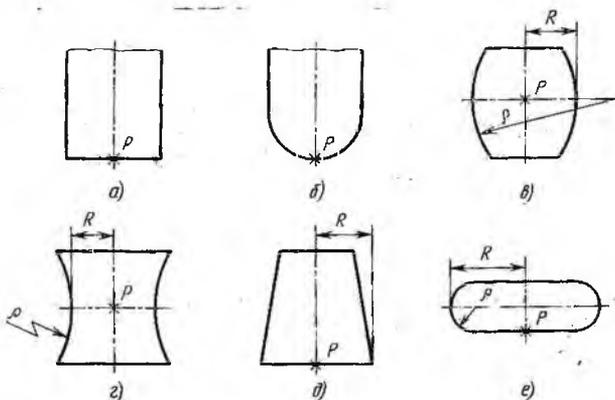


Рис. 47. Зависимость формы фрезы от значений ее параметров:

$a - \rho = 0, \alpha = 0$; $b - \rho = R, \alpha = 0$; $c - \rho > R, \alpha = 0$; $d - |\rho| > R, \rho < 0, \alpha = 0$;
 $e - \rho = 0, \alpha < 0$; $f - L = 2\rho, \alpha = -\pi/2, \rho < R$

основной радиус фрезы; ρ — радиус торовой части; L — размер рабочей зоны фрезы; α — угол конусности. Угол α — положителен, если фреза расширяется в направлении от торца, и отрицателен — в противном случае. Некоторые специфические формы фрез рассматриваемого класса представлены на рис. 47.

Расчет траектории фрезы. По известной поверхности детали $\Gamma = \Gamma(u, v)$, $u, v \in \Omega$, и заданной геометрии фрезы при конкретных условиях связи фрезы с деталью требуется рассчитать перемещения инструмента, обеспечивающие воспроизведение поверхности с указанным допуском ϵ . Иными словами, в области Ω требуется определить такую линию (или совокупность линий), полоса воспроизведения которой полностью покрыла бы Ω .

Для прямоугольных областей с этой целью удобно использовать координатные линии $u = u_i, i = 1, \dots, N$ или $v = v_j, j = 1, \dots, M$, проходящие в одном направлении (рис. 48, а) или попеременно в обоих направлениях (рис. 48, б). Шаг между двумя соседними линиями необходимо выбирать с учетом ширины их воспроизведения.

В случае сложной области Ω ее следует предварительно преобразовать к прямоугольной, подобно тому, как это делалось в п. 1 гл. IV. В результате получаются соотношения $u = u(\tilde{u}, \tilde{v}), v = v(\tilde{u}, \tilde{v})$,

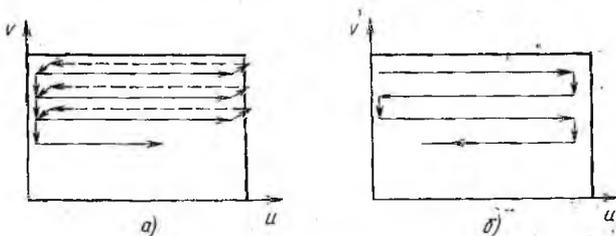


Рис. 48. Схемы обхода области параметров при обработке поверхности

где \bar{u} , \bar{v} — координаты прямоугольной области. Присоединяя эти соотношения к уравнениям связи инструмента с деталью, мы получаем возможность воспользоваться для построения траектории фрезы одной из рассмотренных выше схем.

Иногда перемещение инструмента связывают с линиями пересечения обрабатываемой поверхности с некоторым однопараметрическим семейством поверхностей $A(s)$. В частности, это может быть семейство плоскостей.

Расчет ширины строки. Ширина строки зависит от формы инструмента, характера поверхности в окрестности строки, вида связи инструмента с деталью и величины ε . Эта величина характеризует «поперечную шероховатость» обработанной поверхности. Ее принято называть высотой гребешка.

Рассмотрим алгоритм расчета допустимой ширины полосы воспроизведения d_ε для некоторой точки траектории в предположении, что фреза соприкасается с поверхностью детали своей торовой частью. Пусть $\mathbf{r}^* = \mathbf{r}(u^*, v^*)$ — некоторая точка поверхности детали и $\boldsymbol{\tau}$ — касательный вектор к поверхности в этой точке в направлении движения инструмента. Главные кривизны торовой части фрезы в точке касания с поверхностью детали определяются формулами

$$k_1 = 1/\rho; \quad k_2 = \sin \varphi / [(R - \rho) + \rho \sin \varphi],$$

где $\sin \varphi = |\mathbf{q}(u^*, v^*) \times \mathbf{n}(u^*, v^*)|$, \mathbf{n} — орт нормали к поверхности детали. Нормальную кривизну k_F в направлении $\boldsymbol{\chi}$, перпендикулярном вектору $\boldsymbol{\tau}$, находят по формуле Эйлера:

$$k_F = k_1 \cos^2 \psi + k_2 \sin^2 \psi,$$

где $\sin \psi = \mathbf{q}(u^*, v^*) \cdot \mathbf{n}(u^*, v^*) \cdot \boldsymbol{\tau}(u^*, v^*)$.

Обозначим через k_G нормальную кривизну поверхности детали в точке \mathbf{r}^* в том же направлении $\boldsymbol{\chi}$. Тогда расстояние C между точками двух соприкасающихся окружностей с радиусами $R_1 = |1/k_F|$ и $R_2 = |1/k_G|$ в окрестности точки соприкосновения приближенно выражается формулой $C = e^2 |k_F - k_G|/8$. Здесь e есть расстояние по касательной к окружностям, отсчитываемое от точки их соприкосновения. Эту величину можно взять в качестве оценки ширины строки в точке \mathbf{r}^* . Из условия $C \leq \varepsilon$ вытекает, что $e \leq \sqrt{8\varepsilon / |k_F - k_G|}$.

Представим вектор $e\boldsymbol{\chi}$ в виде ряда $e\boldsymbol{\chi} = \mathbf{r}_u \Delta u^* + \mathbf{r}_v \Delta v^* + \dots$. Ограничиваясь первыми двумя членами, получим

$$|\Delta u^*| = e \frac{|\boldsymbol{\chi} \times \mathbf{r}_v|}{|\mathbf{r}_u \times \mathbf{r}_v|}; \quad |\Delta v^*| = e \frac{|\boldsymbol{\chi} \times \mathbf{r}_u|}{|\mathbf{r}_u \times \mathbf{r}_v|}.$$

Тогда, учитывая, что $d_\varepsilon(u^*, v^*) = ((\Delta u^*)^2 + (\Delta v^*)^2)^{1/2}$, окончательно для ширины полосы воспроизведения в точке \mathbf{r}^* будем иметь следующее выражение:

$$d_\varepsilon(u^*, v^*) \approx 2 \sqrt{\frac{2\varepsilon}{|k_F - k_G|} \frac{(|\boldsymbol{\chi} \times \mathbf{r}_v|^2 + |\boldsymbol{\chi} \times \mathbf{r}_u|^2)^{1/2}}{|\mathbf{r}_u \times \mathbf{r}_v|}}.$$

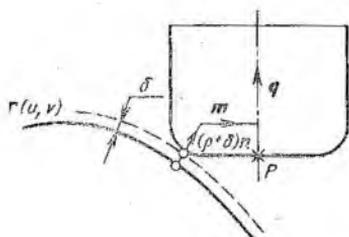


Рис. 49. Контакт фрезы с поверхностью

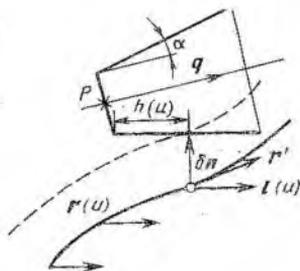


Рис. 50. Контакт фрезы с контуром

Для упрощения расчета траектории ширину полосы воспроизведения на одной строке следует выбирать постоянной и равной минимальному значению d_{ϵ} на множестве вычисляемых точек строки.

Контакт инструмента с деталью. Рассмотрим два наиболее распространенных варианта. Принципиальное различие между ними состоит в том, что в формообразовании участвуют разные части рабочей поверхности фрезы. Это различие определяется их назначением.

1. Фреза своей торовой частью соприкасается с поверхностью, отстоящей от поверхности детали $\mathbf{r} = \mathbf{r}(u, v)$ на расстояние, равное припуску δ . Уравнение связи с деталью в этом случае имеет вид

$$\begin{aligned} \mathbf{P}(u, v) &= \mathbf{r}(u, v) + (\rho + \delta) \mathbf{n}(u, v) + (R - \rho) \mathbf{m}(u, v) - \rho \mathbf{q}(u, v); \\ \mathbf{q} &= \mathbf{q}(u, v), \end{aligned}$$

где $\mathbf{m} = \langle \mathbf{q} \times \mathbf{n} \times \mathbf{q} \rangle$ — орт смещения оси; δ — постоянный припуск на обработку. Геометрический смысл приведенной связи ясен из рис. 49. Если $\mathbf{q}(u, v) = \mathbf{q}_0 = \text{const}$, то фрезерование принято называть *трехкоординатным*. В противном случае говорят о *многокоординатной* обработке.

Вид зависимости $\mathbf{q} = \mathbf{q}(u, v)$ при многокоординатном фрезеровании обычно определяется технологическими условиями обработки и кинематикой станка. Например, это может быть кусочно-постоянная вектор-функция, когда число одновременно управляемых координат не обеспечивает непрерывное изменение направления оси инструмента. С технологической точки зрения наилучшие условия обработки обеспечиваются в случае, когда эта зависимость имеет вид $\mathbf{q}(u, v) = \cos \beta \mathbf{n}(u, v) + \sin \beta \boldsymbol{\tau}(u, v)$, т. е. ось фрезы наклонена в сторону движения под углом $\beta > 0$ (угол опережения) к нормали поверхности. При небольших углах опережения достигается ширина строки, близкая к диаметру фрезы.

2. Траектория фрезы определяется кривой $\mathbf{r} = \mathbf{r}(u)$, $u \in [a, b]$, вектором $\mathbf{l}(u)$ и значением $h(u)$, заданными в каждой точке этой кривой. При движении фреза своей периферийной частью соприкасается с кривой $\mathbf{r} = \mathbf{r}(u) + \delta \mathbf{n}(u)$, где $\mathbf{n}(u) = \langle \mathbf{l}(u) \times \mathbf{r}'(u) \rangle$. Прямолинейная образующая рабочей поверхности фрезы сонаправлена с вектором $\mathbf{l}(u)$. Точка соприкосновения отстоит от плоскости



Рис. 51. Допустимые соотношения между кривизнами k_F и k_G

торца на расстояние $h(u)$, отсчитываемое вдоль образующей (рис. 50). Уравнение связи с деталью есть

$$\mathbf{P}(u) = \mathbf{r}(u) + (\delta + R \cos \alpha) \mathbf{n}(u) - (h(u) + R \sin \alpha) \mathbf{l}(u);$$

$$\mathbf{q} = \mathbf{n}(u) \sin \alpha + \mathbf{l}(u) \cos \alpha.$$

При $\alpha = 0$ и $\delta = -R$ эта зависимость приобретает вид

$$\mathbf{P}(u) = \mathbf{r}(u) - h(u) \mathbf{l}(u); \quad \mathbf{q} = \mathbf{l}(u),$$

т. е. ось инструмента в каждый момент времени проходит через точку кривой $\mathbf{r}(u)$, совпадая по направлению с вектором $\mathbf{l}(u)$. При $h(u) \equiv h_0 = \text{const}$ точка пересечения торца с конической образующей отслеживает кривую $\mathbf{r} = \mathbf{r}(u) + h_0 \mathbf{l}(u)$.

Приведем выражение для $h(u)$ в случае, когда точка \mathbf{P} во время движения фрезы остается в плоскости, определяемой векторным уравнением $(\mathbf{r} - \mathbf{r}_0) \mathbf{N} = 0$;

$$h(u) = [(\mathbf{r}(u) + (\delta + R \cos \alpha) \mathbf{n} + R \sin \alpha) \mathbf{N}] / (\mathbf{N} \mathbf{l}).$$

Определение допустимых параметров фрезы. Принятая гипотеза о характере связи инструмента с деталью накладывает определенные ограничения на форму и размеры фрезы. Им можно дать следующую геометрическую интерпретацию. Рассмотрим произвольное нормальное сечение в точке соприкосновения A . Возможны три варианта допустимых соотношений между нормальными кривизнами k_G и k_F поверхностей G и F в этом сечении (рис. 51). Математически они записываются в виде $k_F < -k_G$. Это условие назовем условием локальной совместимости фрезы с поверхностью детали. Обозначим через K_G и K_F максимальные нормальные кривизны поверхностей G и F . Параметры фрезы следует выбирать такими, чтобы выполнялось соотношение $K_F < -K_G$.

Величина K_G характеризует сплайновую поверхность, аппроксимирующую поверхность детали, в среднем. Поэтому в процессе расчета траектории в каждой ее точке следует контролировать локальную совместимость фрезы с обрабатываемой поверхностью. Нарушение совместимости в точке $\mathbf{r}(u^*, v^*)$ приводит к тому, что поверхность, описываемая вектор-функцией связи $\mathbf{P} = \mathbf{P}(u, v)$ в окрестности значений (u^*, v^*) , имеет самопересечение. Признаком этого обстоятельства служит условие $\mathbf{P}_u \mathbf{r}_u < 0$ или $\mathbf{P}_v \mathbf{r}_v < 0$. Оно позволяет установить лишь факт нарушения локальной совместимости. Геометрически это условие означает, что вследствие петлеобразности поверхности $\mathbf{P}(u, v)$ в окрестности самопересечения ее координатные линии меняют ориентацию по отношению к координатным линиям поверхности $\mathbf{r}(u, v)$.

2. ОПТИМИЗАЦИЯ ТРАЕКТОРИИ ФРЕЗЕРОВАНИЯ СЛОЖНОЙ ОБЛАСТИ

Обработка детали сложной формы на программно управляемом станке может занимать несколько рабочих смен. Только часть этого времени является собственно рабочим временем, когда фреза режет металл, остальное — тратится на вспомогательные операции и на холостые ходы, т. е. перемещения инструмента в воздухе. При фиксированных параметрах резания эффективность тем выше, чем больше доля собственно рабочего времени, и, следовательно, желательно сократить холостые ходы. В случае обработки неодносвязных областей с криволинейными границами решение задачи выбора рациональной траектории далеко не всегда возможно на интуитивном уровне. Здесь обсуждается алгоритмический поиск рациональной траектории.

Постановка задачи и подход к решению. Требуется цилиндрической фрезой в плоской заготовке выфрезеровать полость (рис. 52), которая имеет постоянную глубину и в проекции на плоскость (X, Y) дает неодносвязную область с криволинейной границей. В примере на рис. 52 граница области состоит из одного наружного и двух внутренних контуров. Предполагается, что технические ограничения требуют выполнить одно врезание, выборку крупнозубой фрезой объема полости с припуском Δ ($0 < a \leq \Delta \leq b$) по контуру и, наконец, мелкозубой фрезой осуществить чистовое оконтуривание. Условия вычислений состоят в задании контуров границы конечным числом точек, достаточным для обеспечения погрешности сплайн-интерполяции в пределах заданного значения ε . Предполагается, что $\varepsilon \ll a$.

С помощью сплайн-интерполяции легко строится *эквилидистанта* к точно заданной кривой. В нашем случае таким образом строится граница a -эквилидистантной области, лежащей внутри заданной. Затем в a -эквилидистантную область вписывается многоугольная область таким образом, что каждая замкнутая ломаная ее границы лежит в $(b-a)$ -окрестности a -эквилидистантной области (рис. 53). Эта задача решается на основании учета кривизны граничных контуров a -эквилидистантной области. Полученная неодносвязная многоугольная область аппроксимирует исходную в том смысле, что их

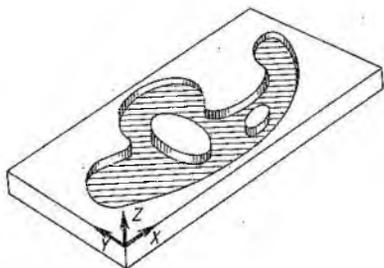


Рис. 52. Фрезерованная полость с криволинейной границей

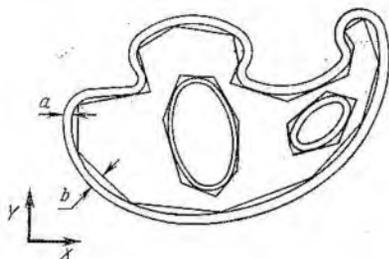


Рис. 53. Многоугольная область, аппроксимирующая криволинейную область

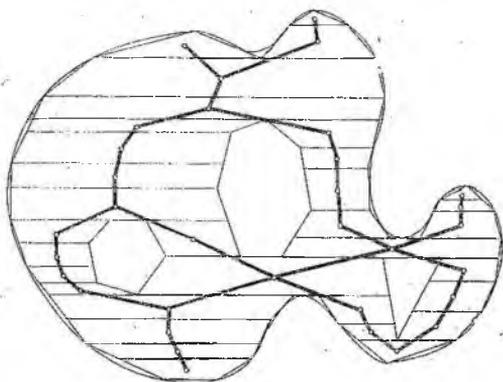


Рис. 54. Разбиение на трапеции

многоугольной области на трапеции (рис. 54). В частности, некоторые из трапеций могут быть вырождены в треугольники.

Отрезок линии раздела, принадлежащий двум трапециям, будем называть *контактом*. По построению любые две трапеции разбиения могут иметь не более одного контакта, и любой контакт лежит на основании (нижнем или верхнем) контактирующей трапеции.

Ясно, что при обработке трапеции оптимальной в указанном выше смысле будет зигзагообразная траектория, ходы которой параллельны основаниям трапеции. Если в критерий оптимальности ввести сумму углов изломов траектории или время торможения на участках реверса, то оптимальной может оказаться иная траектория. Чтобы получить близкую к оптимальной траекторию в целом по области, нужно подобрать рациональную последовательность обхода трапеций разбиения. Если r есть радиус фрезы оконтуривания, то кривая, r -эквидистантная внутрь для границы первоначальной области, завершает полную траекторию обработки.

Поставим в соответствие разбиению *неориентированный связный граф* $G(X, A)$ таким образом, что вершины x_1, x_2, \dots, x_n (множество X) изображают трапеции разбиения, а ребра a_1, a_2, \dots, a_m (множество A) — контакты. Если x_1 и x_2 — концевые вершины ребра a , то говорят, что вершины x_1, x_2 инцидентны ребру a (или что ребро инцидентно вершинам x_1, x_2). Концевые вершины ребра могут совпадать. Ребро с совпадающими вершинами образует *петлю*.

В силу единственности контакта двух трапеций любые две вершины могут быть инцидентны самое большее одному ребру, поэтому $G(X, A)$ является именно графом, а не мультиграфом (рис. 55, а). В *мультиграфе*, хотя бы одна пара вершин инцидентна более чем одному ребру. Трапецию разбиения (и соответствующие вершины графа G) нумеруем сверху вниз и слева направо. Для каждой трапеции разбиения найдем «центр» как точку, координаты которой есть среднее арифметическое соответствующих координат четырех вершин. В случае треугольника вершина, противоположная основанию, берется дважды.

соответствующие граничные контуры лежат в b -окрестности друг друга.

Следующим этапом является подразделение полученной многоугольной области на элементарные фигуры. Для этого из каждой вершины ломаных ее границы внутрь области параллельно фиксированному направлению проводятся прямые до первой встречи с каким-либо участком границы. В результате получается разбиение

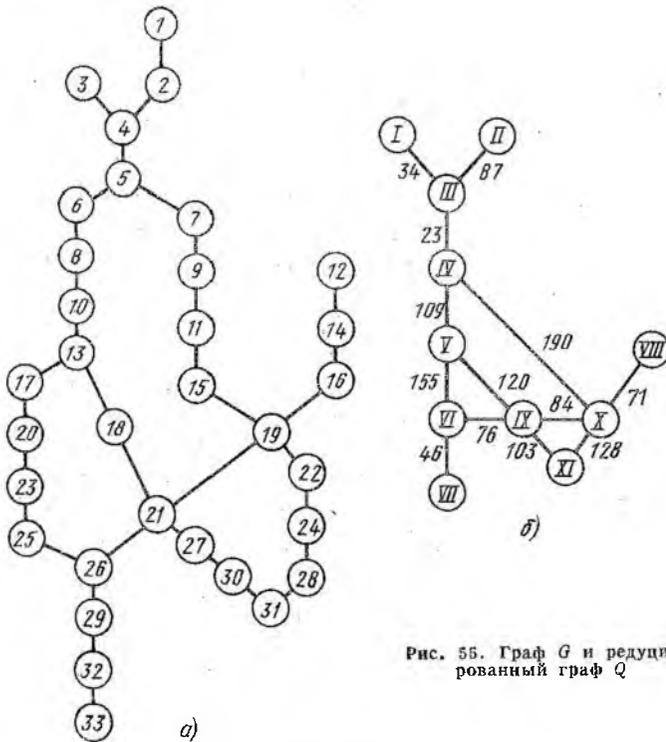


Рис. 55. Граф G и редуцированный граф Q

Некоторые понятия теории графов. Ребра, имеющие общую вершину, называются *смежными*, а число этих ребер — *степенью вершины*. Вершины, являющиеся концами некоторого ребра, называют смежными. Множество вершин, смежных некоторой вершине x_i , обозначают $\Gamma(x_i)$. *Маршрут* есть последовательность попарно смежных ребер $a_1, a_2, a_3, \dots, a_q$, где каждое ребро (кроме, быть может, первого и последнего) связано с предыдущим и последующим обеими своими концевыми вершинами. *Цепь* есть маршрут, в который любое ребро входит не более одного раза. Простая цепь — это цепь, где каждая вершина использована не более одного раза. *Цикл* есть замкнутая цепь. Цикл, исчерпывающий данный мультиграф, называется *эйлеровым циклом*. Связный граф без петель и циклов называется *деревом*.

В теории графов доказывается следующее утверждение: связный неориентированный мультиграф содержит *эйлеров цикл* тогда и только тогда, когда все вершины имеют четные степени.

Ребрам (x_i, x_j) графа $G(X, A)$ припишем положительные веса $c(x_i, x_j) = c_{ij}$, равные расстоянию между центрами контактирующих трапеций. Таким образом $G(X, A)$ станет *графом со взвешенными ребрами*. Вес маршрута (цепи, цикла) определяется как сумма весов ребер, по которым он проходит (с учетом кратности прохождения ребра).

Для цепи из двух ребер (x_{q-1}, x_q) и (x_q, x_{q+1}) преобразование стягивания вершин x_{q-1} и x_q состоит в замене их одной вершиной x и приписывании ребру (x, x_{q+1}) веса, равного сумме весов ребер (x_{q-1}, x_q) и (x_q, x_{q+1}) . Стягиванием смежных вершин степени не большей, чем 2, граф G преобразуется в граф Q (рис. 55, б). В неориентированном графе число вершин нечетной степени обязательно четно. В нашем примере это вершины I—VIII. Редуцированный граф Q содержит топологическую информацию (число фундаментальных циклов) и метрическую информацию (длины ребер) об исходной области с криволинейной границей. С целью получения рациональной последовательности обхода трапеций разбиения будем искать цикл, проходящий через каждое ребро по меньшей мере один раз и такой, что его вес минимален. Это — так называемая *задача китайского почтальона*.

Задача китайского почтальона. Приведенная в [11] методика опирается на решение нескольких вспомогательных задач, которые мы предварительно рассмотрим.

Алгоритм 1. Определение кратчайших цепей между вершиной s и всеми остальными вершинами графа с взвешенными ребрами.

Метод решения известен как алгоритм Дейкстры (по имени первооткрывателя). Вершинам x_i графа присваиваются временные числовые метки $l(x_i)$, определяющие верхнюю границу длины цепи от s к этой вершине, которые в процессе работы алгоритма преобразуются в постоянные — $f(x_i)$. К той вершине, которая только что получила постоянную метку, переходит признак « u » — центральная. Описание алгоритма:

Шаг 1. Дать временные метки $l(x_i) = \infty$ всем вершинам x_i , кроме s . Вершине s дать постоянную метку $f(s) = 0$ и присвоить признак u .

Шаг 2. Для всех $x_i \in \Gamma(u)$, имеющих временные метки, изменить их по правилу

$$l(x_i) = \min[l(x_i), l(u) + c(u, x_i)], \quad (185)$$

где $c(u, x_i)$ — вес ребра (u, x_i) .

Если таковых не оказалось, перейти к следующему шагу.

Шаг 3. Среди всех вершин графа, имеющих временные метки, найти такую x_i^* , для которой $l(x_i^*) = \min[l(x_i)]$.

Шаг 4. Превратить $l(x_i^*)$ в постоянную ($f(x_i^*) = l(x_i^*)$), а саму вершину x_i^* сделать центральной ($u \rightarrow x_i^*$).

Шаг 5. Если в графе осталась хотя бы одна вершина с временной меткой, перейти к шагу 2. Иначе — останов.

Полученная в результате для каждой вершины постоянная метка $f(x_i)$ дает значение веса кратчайшей цепи от вершины s к вершине x_i . Сама кратчайшая цепь отыскивается последовательными переходами от вершины x_q к той вершине $x_{q+1} \in \Gamma(x_q)$, для которой $f(x_{q+1}) = f(x_q) - c(x_{q+1}, x_q)$: как бы обратно к правилу (185).

Для редуцированного графа, представленного на рис. 55, алгоритм Дейкстры, будучи семикратно применен, дает матрицу весов d_{ij} кратчайших цепей между вершинами нечетной степени (табл. 8).

ТАБЛИЦА 8. Матрица весов кратчайших цепей

	I	II	III	IV	V	VI	VII	VIII
I		121	34	57	166	321	367	318
II	121		87	110	219	374	420	371
III	34	87		23	132	287	333	284
IV	57	110	23		109	264	310	261
V	166	219	132	109		155	201	275
VI	321	374	287	264	155		46	231
VII	367	420	333	310	201	46		277
VIII	318	371	284	261	275	231	277	

Когда, например, в качестве s берется вершина V, то после первого выполнения шага 2 другие вершины графа получают согласно правилу (185) следующие временные метки: $l_{IV} = 109$, $l_{VI} = 155$, $l_{IX} = 120$. Остальные вершины сохраняют $l = \infty$. Затем на шаге 4 метка четвертой вершины превращается в постоянную $f_{IV} = 110$, а сама вершина становится центральной для следующего выполнения шага 2 и т. д.

Алгоритм 2. Оптимальное совершенное паросочетание. Пусть D есть полный граф с четным числом $n = 2p$ вершин и взвешенными ребрами, веса d_{ij} — положительные числа, образующие симметрическую матрицу $\|d_{ij}\|$. (Полным называется граф, в котором присутствуют все возможные ребра (x_i, x_j) при $i \neq j$, всего $n(n-1)/2$ ребер). Совокупность из p ребер, исчерпывающая все вершины и такая, что каждая вершина входит ровно в одно ребро, называется совершенным паросочетанием. Задача о минимальном совершенном паросочетании состоит в отыскании совершенного паросочетания, сумма весов ребер которого была бы минимальной.

Для случаев $n = 4$ и $n = 6$ все возможные совершенные паросочетания показаны на рис. 56. Число возможных совершенных паросочетаний в зависимости от числа вершин дается рекуррентным соотношением $M_n = (n-1)M_{n-2}$, откуда непосредственно следует выражение

$$M_n = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (n-1). \quad (186)$$

Например, $M_8 = 1 \cdot 3 \cdot 5 \cdot 7 = 105$; $M_{10} = 1 \cdot 3 \cdot 5 \cdot 7 \cdot 9 = 945$; $M_{12} = 1 \cdot 3 \cdot 5 \cdot 7 \cdot 9 \cdot 11 = 10\,395$ и т. д. С ростом n зависимость опережает экспоненту. Подобное лавинообразное нарастание исключает полный перебор в качестве средства поиска. Существуют алгоритмы, дающие точное решение [11], однако они громоздки в реализации. В качестве компромисса можно предложить следующий алгоритм «усеченного перебора».

Шаг 1. В матрице $\|d_{ij}\|$ ищется наименьший элемент $d_{i_1 j_1}$; соответствующие вершины образуют пару (x_{i_1}, x_{j_1}) , а две строки и два столбца (в силу симметрии матрицы) вычеркиваются. Из оставшихся элементов снова ищется наименьший $d_{i_2 j_2}$, образуется вторая пара (x_{i_2}, x_{j_2}) , вычеркиваются строки и столбцы и так далее, пока

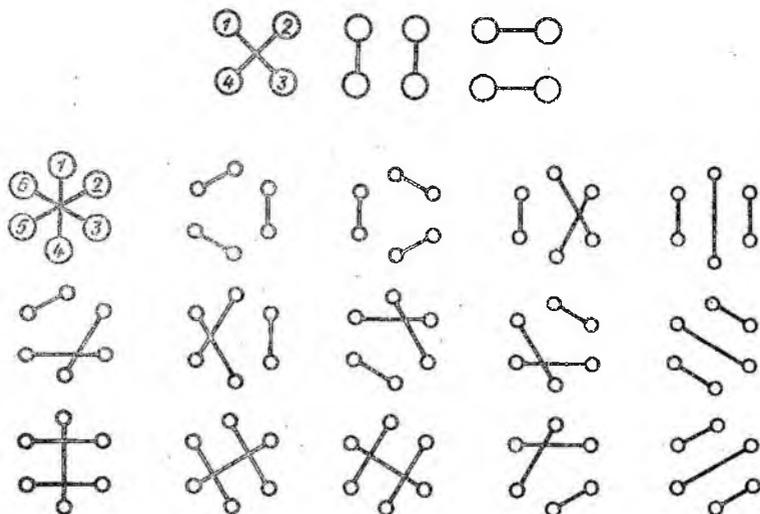


Рис. 56. Совершенные паросочетания для четырех и шести вершин

после p -го раза не получится паросочетание M_n^p , состоящее из пар $(x_{i_1}, x_{j_1}), (x_{i_2}, x_{j_2}), \dots, (x_{i_p}, x_{j_p})$.

Шаг 2. Перебираются сочетания по три элемента из M_n^0 (всего $p(p-1)(p-2)/6$), в каждом из которых участвуют шесть вершин. Для шести вершин согласно рис. 56, существует 15 вариантов совершенных паросочетаний M_6 , из которых перебором находят минимальное. Им заменяют три рассматриваемых пары в паросочетании M_n^0 , преобразуемом таким образом в паросочетание M_n^k . Может оказаться, что исходные три пары уже давали минимальное M_6 , тогда M_n^k совпадает с M_n^0 . Процесс продолжают до тех пор, пока в паросочетании M_n^k на всех $p(p-1)(p-2)/6$ шестерках вершин не будут достигнуты минимумы.

Шаг 3. Если M_n^k и M_n^0 совпадают, выполняется останов. Иначе — полагают M_n^0 совпадающим с M_n^k и возвращаются к шагу 2.

Останов можно осуществлять (в зависимости от запаса времени) на любой стадии выполнения шага 3.

Вообще говоря, алгоритм дает лишь приближенный результат. Этот компромисс представляется оправданным, поскольку рассматриваемая методика изначально содержит элементы эвристики: разбиение области на трапеции, способ «взвешивания» ребер.

Применение алгоритма к матрице, представленной в табл. 8, дает паросочетание (I—II), (III—IV), (V—VIII), (VI—VII) с общим весом 465.

Алгоритм 3. Отыскание эйлерова цикла. Если эйлеров цикл в мультиграфе существует, то он может быть найден с помощью алгоритма, предложенного Флери.

Обход начинать с любой вершины. Вычеркивать пройденное ребро, причем могут возникать изолированные вершины. Не проходить по ребру, если его удаление разбивает мультиграф на две связанные компоненты, каждая из которых содержит невычеркнутые ребра. Таково описание этого простого алгоритма.

Алгоритм решения задачи китайского почтальона. Задача решается следующим образом.

Шаг 1. Используя алгоритм 1, образовать для множества вершин нечетной степени графа Q матрицу $\|d_{ij}\|$, где d_{ij} — наименьший из весов цепей, идущих из вершины x_i в вершину x_j , при всевозможных i, j .

Шаг 2. С помощью алгоритма 2 найти оптимальное паросочетание M для полного графа D , имеющего $\|d_{ij}\|$ в качестве матрицы весов ребер.

Шаг 3. Для каждой пары сочетающихся вершин x_α и x_β с помощью алгоритма 1 определить цепь $\mu_{\alpha\beta}$, соответствующую весу $d_{\alpha\beta}$. Ребра, составляющие все цепи из M , добавить в качестве искусственных к графу Q , получив при этом мультиграф \tilde{Q} , в котором гарантировано существование эйлерова цикла. Если для цепей $\mu_{\alpha\beta}$ и $\mu_{\gamma\delta}$ обнаружится общее ребро, эти цепи заменяют на цепи $\mu_{\alpha\gamma}$ и $\mu_{\beta\delta}$ с меньшим общим весом, в которые то ребро уже не входит.

Шаг 4. Пользуясь алгоритмом 3, найти эйлеров цикл в мультиграфе \tilde{Q} .

Применительно к графу, изображенному на рис. 55, б, алгоритм дает мультиграф \tilde{Q} , показанный на рис. 57, где искусственные ребра даны штриховыми линиями. Оптимальный эйлеров цикл для мультиграфа \tilde{Q} определяется неоднозначно. Два таких цикла приведены на рис. 58.

Спецификой рассматриваемой задачи является то, что обработка на станке ведется по трелециям разбиения, а не по их контактам. В исходном графе G нет необходимости проходить все ребра. Достаточно пройти только все вершины. Возвращаясь от редуцированного

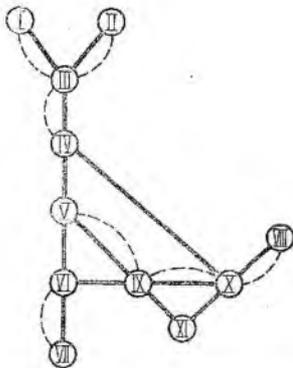


Рис. 57. Мультиграф \tilde{Q}

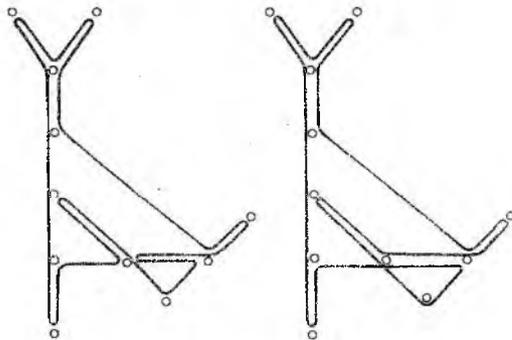


Рис. 58. Оптимальные эйлеровы циклы для мультиграфа \tilde{Q}

графа Q к исходному графу G , мы получим из эйлеровых циклов оптимальные маршруты, на основании которых можно еще выбрать рациональный маршрут, содержащий все вершины. При этом полезно руководствоваться следующими правилами:

начинать с вершины степени единица (если такая имеется), наиболее удаленной от ближайшего перекрестка;

отмечать пройденные вершины, а из данной вершины следовать прежде всего в ту, куда ведет однократно проходимое ребро (если оно есть);

исключать из последовательности обхода те циклы, которые целиком проходят по уже отмеченным вершинам,

заканчивать обход, как только помечена последняя вершина.

Полученная в соответствии с этими правилами рациональная последовательность обхода вершин графа G (рис. 55, a) имеет вид 1, 2, 4, 3, 4, 5, 6, 8, 10, 13, 17, 20, 23, 25, 26, 29, 32, 33, 32, 29, 26, 21, 18, 21, 27, 30, 31, 28, 24, 22, 19, 16, 14, 12, 14, 16, 19, 15, 11, 9, 7. Отношение числа элементов подобной последовательности к числу вершин графа (в нашем примере оно равно 1,24) может служить показателем качества полученного решения. Топология графа G инвариантна относительно изменений разбиения области за счет выбора различных аппроксимирующих многоугольников и направлений разбивающих прямых. Показатель качества получаемых при этом траекторий будет изменяться слабо.

Данная методика может быть распространена на поиск рациональной траектории фрезерования сложных областей на поверхностях, заданных параметрически. В этом случае методика применяется в плоскости параметров с последующим отображением результирующей плоской траектории на поверхность.

Другое возможное направление использования методики связано с фрезерованием, допускающим переходы фрезы вне заготовки (по воздуху) между известными точками сложной области. Формальное отличие этого случая от рассмотренного в том, что графы G и Q здесь не обязательно плоские. Поскольку планарность графов нигде, по сути, не использовалась, методика в целом годится и для этого случая. Разумеется, все «воздушные» участки результирующего маршрута снабжаются специальными признаками.

3. ПОСТАНОВКА НЕКОТОРЫХ ОПТИМИЗАЦИОННЫХ ЗАДАЧ ДЛЯ ПРОЦЕССОВ ПРОГРАММНО-УПРАВЛЯЕМОЙ ОБРАБОТКИ МЕТАЛЛОВ РЕЗАНИЕМ

Токарная обработка. В зависимости от геометрических особенностей обрабатываемых деталей и кинематики соответствующих станков токарная обработка подразделяется на точение, растачивание, нарезание, а также на лоботокарную и карусельную обработки. Здесь речь пойдет только о первом виде токарной обработки — точении.

Типичной для этого вида является обработка удлиненной вдоль оси вращения x поверхности, образующая которой имеет однознач-

ную ортогональную проекцию на ось x . При этом обрабатывается внешняя поверхность детали, и с каждым проходом резца радиус любого поперечного сечения уменьшается. Меридианное сечение реальной детали может быть ограничено кусочно-гладким контуром, т. е. образующая может состоять из нескольких гладких участков с изломами в точках стыка. Отдельные прямолинейные участки образующей могут составлять прямой угол с осью вращения, тогда соответствующие им участки поверхности детали будут кольцевыми площадками, ортогональными оси вращения. Подобные площадки обрабатываются специальными резцами, т. е. при переходе на соответствующие участки образующей требуется смена инструмента.

Режимы резания — число оборотов шпинделя и подача инструмента — зависят от расстояния по радиусу от точки резания до оси вращения. Материал заготовки, тип резца, требуемый класс точности и глубина резания входят в эти зависимости как параметры, а предельная мощность станка играет роль естественного ограничителя.

Токарные станки с ЧПУ созданы довольно давно. Автоматизация подготовки управляющих программ для станков токарной группы не сопряжена с какими-либо принципиальными трудностями. Иногда оказывается достаточно интерполировать точно заданную функцию $r = r(x)$, чтобы получить теоретическую образующую детали. В общем случае проблема сводится к аппроксимации плоского кусочно-гладкого контура, что представляет собой несравненно более простую задачу, чем аппроксимация пространственных кривых или поверхностей.

Рассмотрим несколько специфических задач (в порядке усложнения), которые возникают при подготовке программного обеспечения токарных станков с ЧПУ.

Задача 1. Образующая детали представляет собой гладкую кривую, которая может быть описана однозначной зависимостью радиуса от координаты на оси вращения: $r = r(x)$. Углы $\gamma(x)$ наклона образующей к оси вращения малы, т. е. образующая пологая, и радиус $r(x)$ изменяется в узком диапазоне (рис. 59).

Заданы: направление хода обработки, угол ϕ заточки резца в меридианной плоскости, подача на один оборот шпинделя. Предполагается, что припуск на обработку известен и учтен технологом при назначении режимов резания, которые не меняются в ходе точения.

Программному обеспечению, т. е. комплексу программ для ЭВМ, рассчитывающему управляющие программы для станка, остается получение точек на образующей с шагом по x , равным заданной подаче, и пересчет в элементарные радиальные перемещения инструмента. Кроме того, здесь возможен числовой контроль угла наклона образующей, который на «ниспадающих» участках не должен быть больше угла заточки резца: $\gamma(x) < \phi$.

Задача 2. Образующая также состоит из единственной гладкой кривой, которая может быть представлена однозначной зависимостью $r = r(x)$, однако углы $\gamma(x)$ велики, и радиус $r(x)$ меняется в широких пределах (рис. 60).

Заданы: нормальный припуск (толщина слоя по нормали к образующей), направление хода обработки, угол заточки резца ϕ , угол ψ ,

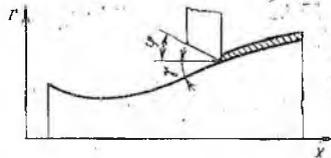


Рис. 59. Простейшая схема точения

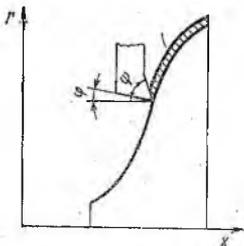


Рис. 60. Схема при широком диапазоне изменения радиуса

задающий положение режущей кромки, данные о материале заготовки и классе точности обработки.

В этом случае программное обеспечение — более сложная задача. Необходимо рассчитывать каждую следующую точку на образующей с учетом информации об уже обработанном участке. Здесь возможен следующий вариант расчета. В зависимости от наклона образующей нормальный припуск пересчитывается на радиальный и рассматривается как глубина резания, исходя из которой с учетом радиуса в данной точке и информации о материале и точности рассчитываются скорость резания и подача на один оборот. Делается шаг по координате x , равный вычисленной подаче, и на образующей выбирается следующая точка. Ее положение сравнивается с положением предыдущей точки, и, если точки оказываются очень удаленными друг от друга, подача соответственно уменьшается. В каждой точке также может осуществляться контроль угла наклона образующей для сравнения его с углом заточки φ , и вычисления текущей глубины резания с учетом угла ψ и припуска на обработку.

Следует заметить, что оборудование с ЧПУ не всегда обеспечивает непрерывное управление угловой скоростью шпинделя, и тогда должно быть задано ее некоторое среднее значение. Кроме того, может быть фиксирован набор некоторых дискретных значений подач, и программное обеспечение должно округлять расчетные подачи до ближайшего или минимального значения из этого набора.

Задача 3. Она объединяет первые две. Образующая состоит из «пологих» и «крутых» гладких кривых, а также содержит кривые с сильно изменяющимся наклоном и прямолинейные участки, перпендикулярные оси вращения (рис. 61). Здесь программное обеспечение должно распознавать качественно отличные от гладких участки образующей и применять к ним соответствующие алгоритмы расчета траектории обработки и режимов резания. Разумеется, что в подобной ситуации выбор направления рабочего хода и смена инструмента также входят в функции программного обеспечения. Если для криволинейного участка образующей выбран фасонный резец, скажем, с эллипсовидной режущей кромкой, расчетные точки траектории инструмента уже не будут лежать на образующей. ЭВМ в этом случае находит точку касания эллипса с криволинейной образующей и вычисляет положение точки привязки инструмента.

Задачи подобной сложности, как правило, имеют многовариантные технологические решения, т. е. возникает проблема выбора оптимального решения в соответствии с заданным критерием качества. Оптимизацию может эффективно провести только соответствующее программное обеспечение.

Задача 4. Задана образующая детали, имеющая сложную конфигурацию и составленная из различных прямолинейных и криволинейных участков, обработать которые нужно по разным классам точности. В отличие от рассмотренных выше задач, здесь задается еще геометрия заготовки.

Рассмотрим следующий условный пример. Заготовка (рис. 62) имеет коническую $A-B$ и фигурную $B-I$ части, причем коническая часть изготовлена из металла повышенной твердости, а фигурная — является отливкой, на которой в зоне контакта с формовочной землей образовалась силикатно-окисная корка. Вся обработка в этом случае представляет собой многоэтапный процесс постепенного приближения к конечной форме, что на рис. 62 схематически изображено семейством штриховых линий. Число итераций (проходов и смен инструмента) тем больше, чем сильнее отличается конечный диаметр (детали) от исходного диаметра (заготовки) и чем выше требуемая точность обработки на данном участке детали.

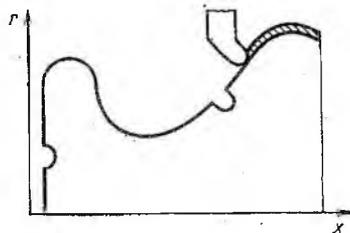


Рис. 61. Образующая с существенно переменным наклоном

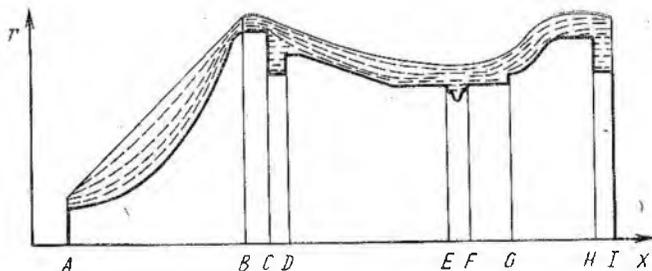


Рис. 62. Последовательные приближения к требуемой конфигурации детали

Технологический процесс представим схематически в виде k -дольного графа. Граф называют k -дольным, если множество его вершин можно разбить самое меньшее на k таких непересекающихся подмножеств (долей), что всякие две смежные вершины входят в различные доли. Пусть участками высокой точности в рассматриваемом примере будут $B-C$ и $H-I$ (см. рис. 62). Предположим, что в распоряжении технолога имеется следующий набор резцов; обдирочный (О), прорезной (П), фасонный (Ф), торцовый левый (ТЛ), торцовый правый (ТП), с эллиптической режущей кромкой (Э), черновой (Ч), черновой твердосплавный (ЧТ), чистовой (ЧИ), доводочный (Д). Один из возможных вариантов технологического процесса представлен в виде 5-дольного графа на рис. 63. Каждая вершина графа изображает технологическую операцию с указанием обрабатываемого участка, используемого инструмента, числа проходов, а также режимов резания и результирующих размеров (на схеме не приведены). Операции выполняются, начиная с верхней доли вершин, последовательно одна доля за другой, а в пределах доли — слева направо. Совокупность ребер графа отражает геометрическую соподчиненность обрабатываемых участков.

Гарантии того, что этот технологический процесс — лучший в смысле производительности, нет. Переставив операции, мы, возможно, сократим количество смен инструмента, а изменив число проходов, — выиграем во времени резания. Программное обеспечение анализирует разные варианты процесса в рамках заданного набора инструментов, исходя из геометрии заготовки. При этом используются сплайны для описания всех кривых переходного семейства, приближающихся к образующей детали.

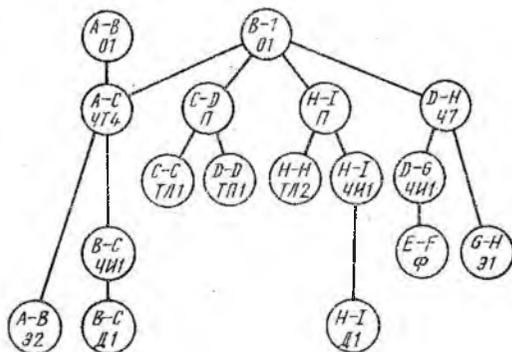


Рис. 63. Представление схемы технологического процесса k -дольным графом

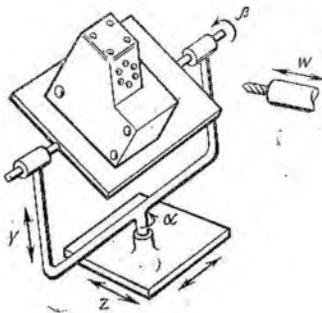


Рис. 64. Кинематическая схема многооперационного станка

Целью поиска может быть, например, минимизация времени обработки. Полное время обработки вычисляется на основании выбираемых режимов резания, длин и числа проходов, а также числа замен инструмента. Естественными ограничениями являются допустимая для каждого инструмента глубина резания, а также технологические возможности станка.

Обрабатывающие центры. Завершая настоящую главу, коснемся коротко программно-управляемой обработки так называемых корпусных деталей, конфигурация которых резко отличается от конфигурации деталей, обычно обрабатываемых на токарных станках. Типичная корпусная деталь (например, передняя бабка токарно-винторезного станка, коробка редуктора, блок цилиндров двигателя и т. п.) — это, как правило, отливка сложной пространственной конфигурации, перфорированная со всех сторон сквозными или глухими отверстиями самых различных типов, размеров и точностей. Обрабатывают такие детали на многооперационных станках (обрабатывающих центрах), снабженных магазином инструментов. Кинематическая схема подобного станка позволяет программировать поступательные перемещения заготовки в трех направлениях (X , Y , Z), ее кантование (α , β), подачу шпинделя (W), а также смену инструмента (рис. 64). Кроме того, могут быть запрограммированы некоторые вспомогательные действия, также как подвод в зону резания масла или охлаждающей эмульсии.

В конце 70-х годов в Институте математики СО АН СССР была создана система ФРАНС, предназначенная для подготовки управляющих программ к станкам типа «обрабатывающий центр». Это была первая в стране система, рассчитывающая как геометрию, так и технологию обработки корпусных деталей. Аббревиатура ФРАНС происходит от названий видов металлорезания: Фрезерование, Растачивание, Развертывание, Нарезание, Сверление. Если к этому добавить зенкерование, зенкование и цекование, получим полный список операций, выполняемых многооперационным станком. Исходной для системы ФРАНС является минимальная информация геометрического (относительные положения центров отверстий, точечные задания контуров полостей, уровни поверхностей заготовки) и технологического (материал заготовки, глубина отверстия или полости, диаметр круглого отверстия, точность, параметры станка) характера. Для упорядочения геометрического описания детали в язык системы было введено понятие грани; все отверстия, которые можно обработать с использованием только поступательных перемещений заготовки без кантования, считаются принадлежащими одной грани [28].

Система анализирует информацию о размерных цепях на каждой грани и осуществляет привязку положений точек задания к общей системе координат. Затем определяются обрабатывающие операции для каждого отверстия или полости и последовательность этих операций. Для выполнения каждой операции назначается инструмент и вычисляются рациональные режимы резания. Полученная технологическая информация записывается на магнитную ленту и в дальнейшем используется постпроцессорной частью системы для формирования управляющей программы к конкретному станку.

Обрабатывающий центр может вести фрезерование сложной области с криволинейной границей, и тогда актуальной окажется задача, рассмотренная в предыдущем параграфе. Однако типичной для обрабатывающего центра является именно обработка круглых отверстий с нерегулярным расположением на детали. Здесь возникает проблема выбора рациональной стратегии обработки множества отверстий, относящаяся к области теории расписаний. Эту проблему можно ставить в ограниченном и полном вариантах.

Ограниченная постановка задачи в следующем: в пределах одной грани найти последовательность обхода рабочих позиций (центров требуемых отверстий) такую, чтобы общее время позиционирования было минимальным. Формально эту задачу можно рассматривать как поиск во взвешенном графе G (t_{ij} — длины ребер во временной мере) маршрута минимальной длины, содержащего все вершины. Это так называемая *задача коммивояжера*.

Алгоритм «иди в ближайшую, еще не пройденную вершину» мы изложим в форме последовательно выполняемых шагов.

Шаг 1. В графе G произвольно выбирается одна из вершин, которой присваивается статус центральной — $ц$. Результирующий маршрут полагается пустым.

Шаг 2. Из непройденных алгоритмом Дейкстры выбирается та вершина, к которой ведет кратчайшая цепь.

Шаг 3. Эта вершина становится центральной — $ц$. Старая вершина $ц_c$ помечается как пройденная, а цепь $ц_c — ц$ добавляется к результирующему маршруту.

Шаг 4. Если еще имеются непройденные вершины, процесс повторяется с шага 2. Иначе — останов.

Наиболее просто алгоритм работает на полном графе, поскольку для любой его вершины $ц$ ближайшая находится в множестве $\Gamma(ц)$, и на каждом шаге цепь $ц_c — ц$ сводится к одному ребру, инцидентному вершинам $ц_c$ и $ц$.

Случай полного графа, очевидно, реализуется на такой грани, где нет крупных выступов и в любую позицию можно переместиться из любой другой по прямолинейной траектории. Когда на грани имеются высокие выступы, их контуры, заданные точечно, могут быть аппроксимированы сплайн-кривыми. Во всяком случае переходы в «закрытую» позицию осуществляются по заданным траекториям, окаймляющим выступы.

Здесь задача коммивояжера тоже может быть поставлена на полном графе. Однако, если в случае «открытого» позиционирова-

ния такая постановка естественна, то при наличии «закрытых» позиций она может привести к неоправданному усложнению задания. Например, если вблизи периметра выступа имеются рабочие позиции, то информация об окаймляющей траектории оказывается излишней. В самом деле, переход в «закрытую» позицию вполне осуществим в виде серии прямолинейных переходов от позиции к позиции вдоль периметра выступа. Но в таком случае придется решать разновидность задачи коммивояжера на неполном графе.

Известно, что при определенных условиях алгоритм «иди в ближайшую» является асимптотически точным. Это означает, что маршруты, получаемые с его помощью, тем ближе к минимальному, чем больше число вершин в графе G .

Полный вариант проблемы выбора рациональной стратегии обработки не ограничивается только позиционированием, но учитывает также и технологический аспект. Это объясняется тем, что требуемое качество поверхности отверстия высокой точности воспроизводится не сразу, а последовательными приближениями с помощью все более точных инструментов, снимающих с каждым проходом слой металла все меньшей толщины. Например, типичная последовательность операций для получения точного отверстия среднего диаметра: сверление тонким сверлом, рассверливание толстым сверлом, зенкерование, развертывание. Для отверстий большого диаметра в заготовке огливаются исходные окна или глухие карманы, а затем требуемая цилиндрическая поверхность получается за счет растачивания черновым и чистовым резцами. В некоторых позициях могут воспроизводиться ступенчатые отверстия, тогда количество операций в одной позиции еще более возрастает.

Ясно, что последовательность операций в каждой позиции жестко определена, так как нельзя, например, вести рассверливание до сверления или черновое растачивание после чистового. С другой стороны, в нескольких позициях могут оказаться операции, выполняемые одинаковым инструментом, даже при различных типоразмерах требуемых отверстий. Поскольку для смены инструмента требуется определенное время, возникает вопрос, что лучше — минимизировать число позиционирований, многократно беря из магазина, устанавливая в патрон и возвращая обратно один и тот же инструмент, или, напротив, стараться одним инструментом с одной установки обработать как можно больше отверстий, смирившись с повторением холостых переходов.

При полной постановке задачи допускаются позиции, которые из соображений обеспечения требуемой точности необходимо обрабатывать с одного позиционирования. Такие позиции характеризуются (для данной задачи) только двумя операциями — первой и последней. Наконец, могут быть позиции с жестко заданной последовательностью обхода, например в случае, когда нужно точнее выдержать межцентровое расстояние или обойти какой-то выступ.

Итак, на грани заготовки имеются множество позиций $\{P_h\}$ и множество инструментов $\{I_l\}$ для выполнения технологических

операций в этих позициях. Каждую операцию можно рассматривать как вершину x^{kl} ориентированного графа O . Индексы здесь означают применение l -го инструмента в k -й позиции. Дуги орграфа соответствуют допустимым переходам от операции к операции. Между некоторыми вершинами имеются прямая и обратная дуги одинаковой длины, в совокупности эквивалентные ребру. Длина дуги, исходящей из вершины x^{pq} в вершину x^{kl} , $t_{pq}^{kl} = t_p^k + \tau_q^l$, где t_p^k — временное расстояние (с учетом скорости перемещений) между p -й и k -й позициями, τ_q^l — время замены q -го инструмента на l -й инструмент. Среди вершин орграфа O может быть такая, которая соответствует не одной операции, а целому набору операций, выполняемых в различных позициях с жестко заданной последовательностью. Дуги, инцидентные такой вершине, могут иметь разные длины, если в установленной для набора технологии начальная позиция не совпадает с конечной и первый инструмент отличен от последнего.

Качественное усложнение задачи в полной постановке состоит в том, что орграф O меняется в процессе поиска: после прохождения определенной вершины в орграфе добавляются новые дуги конечной длины, ведущие из одних вершин в некоторые другие. Эта «динамика» является отражением того факта, что технологически последующая операция допускается только после выполнения технологически предшествующей. В количественном отношении задача усложняется, если рассматривается орграф не для одной грани детали, а для всей детали в целом. При этом в длины некоторых дуг войдет слагаемое, учитывающее время кантования.

Поиск кратчайшего пути, содержащего все вершины орграфа O , является задачей теории расписаний, существенно более сложной, чем задача коммивояжера. К сожалению, не известен не только точный, но, хотя бы, и асимптотически точный алгоритм, который решал бы эту задачу с большей эффективностью, чем простой перебор вариантов. Полный перебор неосуществим за приемлемое время в силу необозримо огромного (хотя и конечного) множества возможных путей. Можно рекомендовать лишь идею локального усеченного перебора, когда взятый в качестве нулевого приближения путь улучшается малыми изменениями посредством проб и ошибок.

Рассмотренные в последних двух параграфах оптимизационные задачи стоит решать и тогда, когда нет гарантии получения глобального минимума. Всякое, пусть даже небольшое, улучшение за счет дополнительных вычислительных затрат случайно выбранной стратегии технологического процесса может принести ощутимый выигрыш во времени при воспроизведении серии деталей. Баланс компьютерного и станочного времени — вот объективное мерило полезности предпринятых усилий.

Глава VI

ЧИСЛЕННАЯ ВИЗУАЛИЗАЦИЯ ПРОСТРАНСТВЕННЫХ СЦЕН

Зрительный образ трехмерного мира создается при анализе мозгом пары двумерных изображений, создаваемых глазной оптикой на сетчатке. С помощью фотокамеры двумерные изображения пространственных объектов фиксируются на светочувствительных пластинах. В обоих случаях имеет место центральное проецирование, т. е. отображение трехмерного объекта на некоторую поверхность конусным пучком прямых. В данной главе речь пойдет о моделировании на ЭВМ оптического процесса построения двумерного изображения пространственного объекта.

1. ИЗОБРАЖЕНИЕ ПРОСТЕЙШИХ ПРОСТРАНСТВЕННЫХ ОБЪЕКТОВ

Пространственный объект может быть представлен в памяти ЭВМ некоторой математической моделью, учитывающей с той или иной степенью полноты его геометрические и физические свойства. В этом параграфе мы рассмотрим построение изображений на основании простейшей модели, содержащей только геометрическую информацию о пространственно расположенных точках и кривых.

Некоторые способы проецирования. Пусть в пространственной системе координат (X, Y, Z) ось Z направлена от проекции к объекту, а центр проецирования F лежит на оси Z за объектом (рис. 65, а). На сфере с центром в F возьмем за начало координат точку P , где эту сферу, являющуюся здесь картинной поверхностью, пересекает отрицательная полуось Z , за координату u примем угол между плоскостью Y, Z и вертикальной плоскостью, проходящей через центр и точку P на сфере, а за координату v — угол между плоскостью X, Z и плоскостью, проходящей параллельно оси X через центр и некоторую точку на сфере. Тогда координаты u, v точки-образа при *центральной проецировании* можно выразить через пространственные координаты точки-прообраза с помощью формул.

$$\begin{aligned} u &= \arcsin \left(\frac{X}{\sqrt{\lambda^2 + (Z - Z_F)^2}} \right); \\ v &= \arcsin \left(\frac{Y}{\sqrt{Y^2 + (Z - Z_F)^2}} \right). \end{aligned} \quad (187)$$

Когда центральное проецирование осуществляется на плоскость (X, Y) и координаты u, v в плоскости изображений совпадают соот-

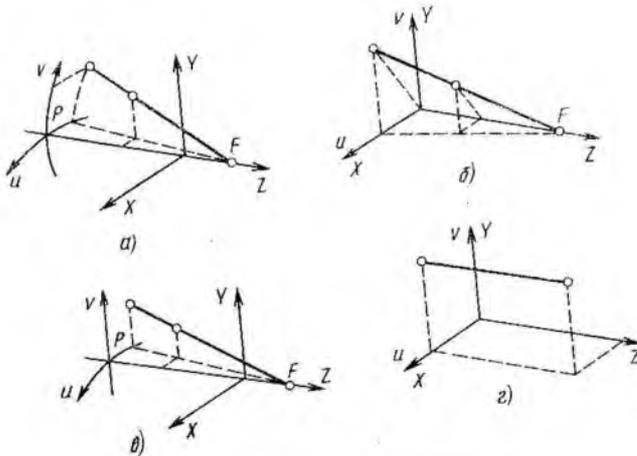


Рис. 65. Различные способы проецирования

ветственно с координатами X, Y (рис. 65, б), то выражения координат точки-образа через координаты точки-прообраза имеют вид

$$u = \left| \frac{Z_F}{Z - Z_F} \right| X; \quad v = \left| \frac{Z_F}{Z - Z_F} \right| Y. \quad (188)$$

При центральном проецировании на поверхность кругового цилиндра, ось которого проходит вертикально через центр F (рис. 65, в), угловая u и линейная v координаты точки на изображении выражаются через пространственные по формулам

$$u = \arcsin \left(\frac{X}{\sqrt{X^2 + (Z - Z_F)^2}} \right); \quad v = \left| \frac{Z_P - Z_F}{Z - Z_F} \right| Y. \quad (189)$$

Если центр F расположен между проекций и объектом, изображение оказывается перевернутым. Дополнительные два отражения в плоскостях $X = 0$ и $Y = 0$ ставят изображение «на ноги», причем формулы (187)—(189) остаются в силе.

При параллельном проецировании проекция пространства на плоскость осуществляется семейством параллельных прямых. Если ось Z задать параллельной проецирующим прямым, а в качестве плоскости изображений взять плоскость $Z = 0$ (рис. 65, г), то координаты проекции будут совпадать с соответствующими координатами точки-прообраза:

$$u = X; \quad v = Y. \quad (190)$$

Правые части формул (187)—(190) записаны с точностью до числового коэффициента, выполняющего роль масштабного множителя.

Простейшие объекты. Когда объект представляет собой совокупность из N пространственно расположенных точек, построение его изображения тривиально: для этого нужно N раз применить формулы типа (187)—(190). Задача усложняется, если объектом является

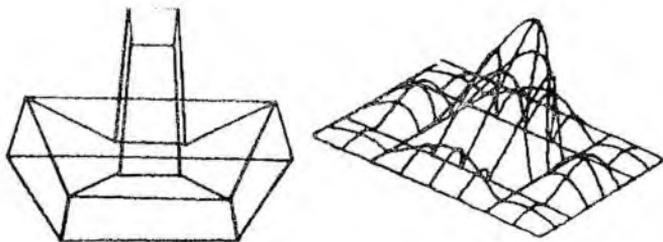


Рис. 66. «Проволочные» конструкции

так называемая «проволочная конструкция» (рис. 66), которая составлена из конечного числа пространственно расположенных линий с кусочно-непрерывной кривизной. В этом случае удобно использовать сплайн-интерполяцию пространственных кривых. На основании интерполяционных зависимостей линии заменяются достаточно густыми конечными последовательностями точек. Для каждой последовательности вычисляются проекции ее точек, и эти проекции связываются прямолинейными отрезками в ломаные линии. Полученные таким путем линии образуют приближенную проекцию «проволочной конструкции». Степень приближения кривой к помощи ломаной можно характеризовать отношением стрелки прогиба h дуги кривой к длине l , стягивающей эту дугу хорды: $\Lambda = h/l$.

Взяв в качестве координатного репера сопровождающий трехгранник кривой в точке \mathbf{r}_0 и используя разложение Тейлора в этой точке нормальных координат радиус-вектора \mathbf{r} текущей точки как функций координаты вдоль касательной, можно показать, что отношение h/l (с точностью до малых высших порядков относительно величины $k_0 l$) совпадает с подобным отношением Λ дуги окружности с кривизной $k = k_0$, где k_0 есть кривизна кривой в точке \mathbf{r}_0 . Для окружности справедливо соотношение $h/l = \Lambda = kl/8 + O(k^3 l^3)$, откуда, пренебрегая малыми высших порядков, получаем

$$l = 8\Lambda/k. \quad (191)$$

Если величину l из (191) выбрать в качестве локального шага вдоль кривой, то с точностью до малых высших порядков будет обеспечено заданное значение Λ отношения h/l .

Кривизна k' проекции, разумеется, не совпадает с кривизной k проецируемой кривой. Для того чтобы вдоль кривой выбрать шаг, обеспечивающий заданное Λ' на проекции, необходимо учитывать взаимное расположение кривой, проецирующего луча и плоскости проекции. Введем обозначения: α, β, γ — углы между проецирующим лучом и соответственно касательной к кривой, главной нормалью кривой и нормалью плоскости проекции; δ — угол между плоскостью P , ортогональной к плоскости проекции и параллельной проецирующему лучу, и плоскостью, параллельной проецирующему лучу и касательной к кривой; φ — угол между плоскостью P и плоскостью, параллельной проецирующему лучу и главной нормали кривой. Проекции l' и h' хорды и стрелки прогиба дуги с точностью

до малых высших порядков относительно величины h/l выражаются формулами

$$l' = \left| \frac{l \sin \alpha}{\cos \gamma} \sqrt{1 - \sin^2 \gamma \sin^2 \delta} \right|; \quad h' = \left| \frac{h \sin \beta}{\cos \gamma} \sqrt{1 - \sin^2 \gamma \sin^2 \varphi} \right|.$$

Чтобы получить заданное отношение $h'/l' = \Lambda'$ на проекции, нужно на кривой выбрать шаг l , исходя из условия

$$\frac{h}{l} = \Lambda = \Lambda' \frac{\sin \alpha}{\sin \beta} \sqrt{\frac{1 - \sin^2 \gamma \sin^2 \delta}{1 - \sin^2 \gamma \sin^2 \varphi}}.$$

Учитывая (191), находим выражение для локального шага по кривой

$$l = \frac{8\Lambda'}{k} \frac{\sin \alpha}{\sin \beta} \sqrt{\frac{1 - \sin^2 \gamma \sin^2 \delta}{1 - \sin^2 \gamma \sin^2 \varphi}}.$$

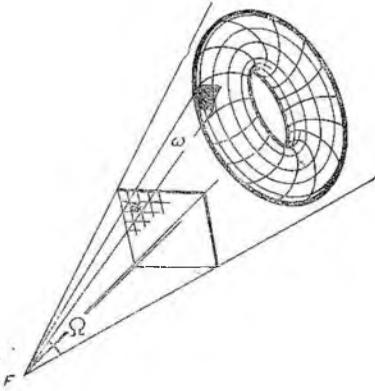
Вычисленный по этой формуле шаг обеспечит (с точностью $o(h/l)$) заданное значение Λ' на проекции. Таким образом, благодаря дополнительным вычислительным затратам достигается сжатие результирующей информации.

2. ИЗОБРАЖЕНИЕ НЕПРОЗРАЧНЫХ ПОВЕРХНОСТЕЙ

Двусторонняя поверхность является геометрической моделью тонкой физической оболочки. В этом параграфе мы рассмотрим два подхода к математическому моделированию оптических свойств (отражательной способности, степени прозрачности, цвета) таких оболочек.

Штриховое изображение. Будем считать, что кусок гладкой поверхности задается каркасом, диаметр ячеек которого не превышает величины D (см. п. 1, 2 гл. II). Одна совокупность каркасных линий — $u = u_i, i = 1, 2, \dots, m$ — зафиксирована в качестве первой, другая — $v = v_j, j = 1, 2, \dots, n$ — в качестве второй. Тем самым в каждом узле каркаса определена упорядоченная пара касательных векторов $(\mathbf{r}_u, \mathbf{r}_v)$. Примем за вектор положительной нормали к каркасу в каждом узле векторное произведение $\mathbf{r}_u \times \mathbf{r}_v$. Таким образом мы получим ориентированный каркас, задающий одну сторону поверхности. Каркас с противоположной ориентацией задает другую сторону той же поверхности. Каркасы обеих сторон поверхности могут быть разными, а могут и совпадать. Замкнутая поверхность составляется из нескольких простых кусков, каждый из которых задается своей парой ориентированных каркасов. Это дает возможность строить изображение замкнутой поверхности не только с внешней точкой наблюдения, но и с точкой наблюдения, расположенной внутри нее. Неориентируемая поверхность ленты Мебиуса задается одним ориентированным каркасом, наложенным сам на себя в два слоя с противоположно направленными нормальными.

По исходному каркасу с использованием сплайн-аппроксимации поверхности строится загущающий каркас соответствующей ориентации, максимальный диаметр ячеек которого $d \ll D$. Узлы исходного и загущающего каркасов используются для создания штрихо-



ции (рис. 67). Этот конус принято называть *пирамидой видимости*, а его сечение — кадром.

Телевизионный способ визуализации основан на разложении изображения в упорядоченную по строкам и столбцам матрицу светящихся пятен, называемых *элементами раstra*. Разобьем пирамиду видимости на элементарные телесные углы ω_{ij} по числу элементов раstra. С каждым ω_{ij} свяжем вещественную переменную ρ_{ij} , которая будет принимать значения расстояния от центра проецирования до ближайшего в данном направлении узла каркаса. Фиксируем малую положительную величину ϵ .

Растровый алгоритм построения штрихового изображения непрозрачной поверхности. Этот алгоритм заключается в следующем.

Шаг 1. Всем ρ_{ij} присваивается значение ρ_{ij}^{\max} , равное максимально возможному в данной ЭВМ числу.

Шаг 2. Для каждой линии каркаса перебираются лежащие на ней узлы. При этом определяется угол ω_{ij} , в который входит конкретный узел M , и вычисляется вектор \overrightarrow{MF} .

Шаг 3. Если $|\overrightarrow{MF}| < \rho_{ij}$, то присваивается значение $\rho_{ij} = |\overrightarrow{MF}|$ и осуществляется переход к шагу 4. В случае, когда $|\overrightarrow{MF}| > \rho_{ij}$, рассматриваемый узел находится от центра проецирования дальше, чем тот из предшествующих узлов, который «породил» данное ρ_{ij} . Поэтому ρ_{ij} сохраняет прежнее значение, рассматриваемый узел помечается как невидимый и осуществляется возврат к шагу 2.

Шаг 4. Вычисляется скалярное произведение $\langle \overrightarrow{MF} \rangle_n$, численно равное косинусу угла ψ между нормалью в узле и направлением на центр проецирования. Если $|\cos \psi| < \epsilon$, узел относится к разряду видимых и выполняем переход к шагу 6, иначе — переход к шагу 5. Таким образом, на гладких участках поверхности выделяются узлы, входящие в очерк объекта.

Шаг 5. Если $\cos \psi < -\epsilon$, узел считается невидимым и производится возврат к шагу 2. Если $\cos \psi > \epsilon$, то выполняется про-

верка: является ли данная линия явной, т. е. входящей в исходный каркас. В случае неявной линии (принадлежащей загущающему каркасу) узел полагается невидимым и производится возврат к шагу 2. В случае явной линии узел полагается видимым.

Шаг 6. Образ видимого узла соединяется прямолинейным отрезком с образом предшествующего видимого узла, если этот узел соседствует с данным. (Видимый узел может оказаться начальным, конечным или промежуточным на видимом участке линии). Возврат к шагу 2.

Если в шаге 2 для некоторой точки не нашлось элементарного угла ω_{ij} , это означает, что изображение точки вышло за пределы кадра. Естественно, такая точка считается невидимой.

По завершении перебора всех узлов на всех линиях полного каркаса получается штриховое изображение поверхности с удаленными невидимыми частями. Очерк гладких кусков поверхности объекта образует изображения тех (не экранированных другими) участков каркасных линий, для точек которых $|\cos \psi| < \varepsilon$. В очерк всего объекта могут входить изображения некоторых ребер на его поверхностях.

Значение ε в каждом конкретном случае выбирают индивидуально. В связи с этим можно высказать следующие общие соображения. Чем больше величина ε , тем длиннее куски каркасных линий, проявляющиеся в очерке. Участок неявной линии между двумя соседними узлами может совсем не проявиться в очерке, если ε меньше, чем угол между нормальными в этих узлах.

Рис. 68 представляет собой кинограмму — хронологическую последовательность мгновенных изображений движущегося объекта. Объект состоит из двух непрозрачных незамкнутых оболочек — параболической и плоской с круглым отверстием. Одна сторона параболической оболочки задана прямоугольным каркасом, а другая — радиальным. Наконец, рис. 69 содержит подборку штриховых изображений, воспринимаемых с различных точек зрения, еще более сложного объекта — некоторого условного самолета. Все изображения были рассчитаны ЭВМ, выведены на экран электронно-лучевой трубки и зафиксированы на киноленте.

Тоновые изображения. Математическую модель совокупности объемных тел фиксированного расположения (объект) с указанием источника освещения называют *трехмерной сценой*. Создание зрительных образов (*визуализация*) трехмерных сцен сводится к *численному синтезу тоновых изображений*. По сравнению со штриховыми тоновые изображения обладают гораздо большей естественностью. Это достигается за счет передачи в изображении отражательных свойств и цвета поверхностей, а также условий освещения объекта. Тоновые изображения в динамике в ряде случаев могут быть использованы вместо смотровых моделей изделий.

Пусть объект освещается конусным пучком света от точечного источника, расположенного в вершине конуса. Световой конус имеет прямоугольное сечение, для обозначения которого будем использовать термин «световое окно». Сам световой конус будем

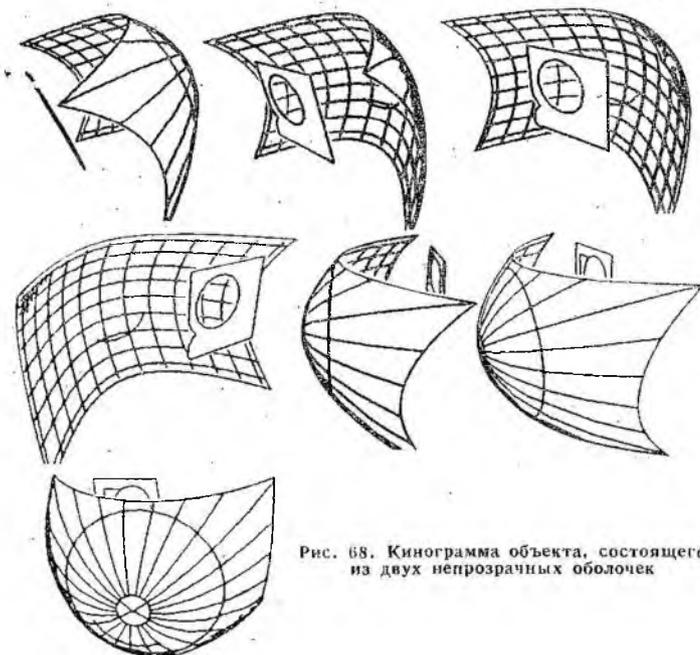


Рис. 68. Кинограмма объекта, состоящего из двух непрозрачных оболочек

называть *пирамидой освещения*. Если на световое окно наложить сетку раstra (как правило, эквивалентную сетке раstra кадра), то пирамида освещения разбивается на элементарные телесные углы ω_{ni} , как и пирамида видимости.

Подготовка данных для растрового алгоритма синтеза тонового изображения состоит в вычислении конечной (ε, γ)-сети поверхности по ее исходному каркасу. Отметим, важное различие между тоновыми и штриховыми изображениями. Если загущенный каркас является в сущности совокупностью конечных последовательностей точек, упорядоченных по линиям на поверхности, то точки конечной ε -сети могут быть распределены около поверхности совершенно произвольно.

В каждой точке ε -сети вычисляется орт приближенной нормали к поверхности. Получается совокупность пар точка — вектор такая,

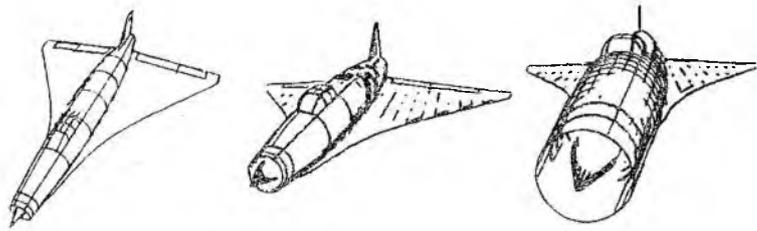
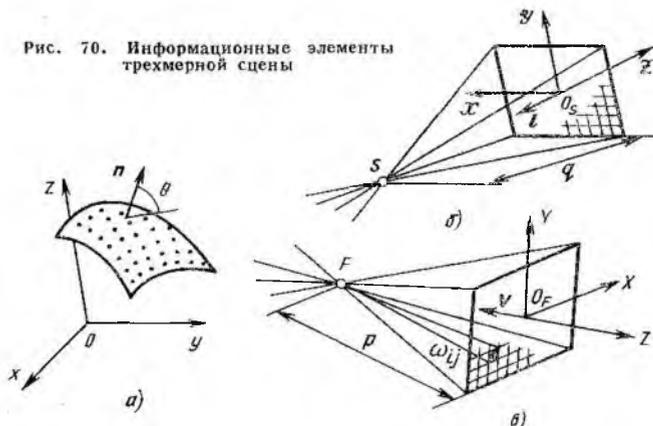


Рис. 69. Изображения сложного объекта в различных ракурсах

Рис. 70. Информационные элементы трехмерной сцены



что для любой точки a поверхности найдется пара, точка которой отстоит от a не далее чем на ϵ , а вектор образует с нормалью n угол, не больший угла γ . Эту совокупность мы и называем (ϵ, γ) -сетью гладкой поверхности. Подобная (ϵ, γ) -сеть заключает в себе геометрическую информацию о поверхности объекта. Координаты x, y, z — точки сети и компоненты x_n, y_n, z_n орта приближенной нормали в ней вычисляются относительно декартовой системы (x, y, z) , связанной с объектом.

Для описания локальных оптических свойств каждой точке (ϵ, γ) -сети приписываются альбедо a_{+1}, a_{-1} и показатель цвета c_{+1}, c_{-1} соответственно внешней (+1) и внутренней (-1) сторон поверхности. Альбедо является неотрицательным числом в пределах от нуля до единицы, которое характеризует отражательную способность участка поверхности, непосредственно прилегающего к данной точке (ϵ, γ) -сети. Показатель цвета есть тройка неотрицательных чисел, сумма которых равна единице. Он задает соотношение интенсивностей основных цветов, смесь которых и дает цвет участка поверхности.

Величины $x, y, z, x_n, y_n, z_n, a_{+1}, a_{-1}, c_{+1}, c_{-1}$ образуют простейшую законченную единицу (запись) исходных данных, которую мы будем называть дискретным элементом поверхности (ДЭП).

Таким образом, исходные данные в целом приобретают простую структуру одномерного массива — ДЭП₁, ДЭП₂, ..., ДЭП_N, — каждая запись которого представляет дискретный элемент поверхности.

Информационное описание трехмерной сцены включает следующие данные: 1) (ϵ, γ) -сети поверхности (рис. 70, а); 2) точечный источник S силы света I с пирамидой освещения, заданной вектором осевой линии e и расстоянием q плоскости светового окна (световой плоскости) от источника (рис. 70, б); 3) центр проецирования F с пирамидой видимости, заданной вектором линии визирования v и расстоянием p от картинной плоскости до точки F (рис. 70, в); 4) растровое поле кадра $2e_r \times 2g_r$ ($2e_r$ строк, $2g_r$ элементов разложения строки) с линейной ценой Δ_r дискретной единицы раз-

мера растрового поля; 5) аналогичные величины e_S, g_S, Δ_S вводятся для растрового поля светового окна; 6) масштабные коэффициенты μ_F и μ_S для кадра и окна соответственно. На рис. 70, а (e, γ)-сеть поверхности показана только точками, а вектор нормали n дан лишь в одной из них.

Помимо системы координат (x, y, z) , связанной с объектом (рис. 70, а), в сцене фигурируют еще системы $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ и (X, Y, Z) , связанные с источником света и центром проецирования соответственно. Координатная система $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ выбрана так, что плоскость $\mathcal{X}O_S\mathcal{Y}$ совпадает со световой плоскостью, начало O_S совпадает с центром окна, оси $O_S\mathcal{X}$ и $O_S\mathcal{Y}$ параллельны сторонам окна, а положительное направление оси $O_S\mathcal{Z}$ выбрано от источника к окну (рис. 70, б), Координатная система (X, Y, Z) выбрана так, что плоскость XO_FY совпадает с картинной плоскостью, начало O_F совпадает с центром кадра, оси O_FX и O_FY параллельны сторонам кадра, а положительное направление оси O_FZ выбрано от точки зрения к кадру (рис. 70, в). Расстоянию p приписывается знак Z -координаты точки F , а расстоянию q — знак \mathcal{Z} -координаты источника света. Связь между системами (X, Y, Z) и (x, y, z) задается матрицей $\|F\|$, а связь между системами $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ и (x, y, z) — матрицей $\|S\|$.

Для размещения информации синтезируемого изображения служит двумерный массив, строки которого соответствуют строкам развертки кадра. Записи этого массива соответствуют элементам раstra и называются дискретными элементами картины (ДЭК). Каждый ДЭК включает: ρ_F — расстояние от точки F до точки поверхности; E — яркость элемента раstra; c — цвет элемента раstra. Для яркостей элементов раstra используют градации от минимальной E_0 до максимальной E_{\max} . Фиксируются некоторые E_Φ и c_Φ — значения яркости и цвета фона.

Для учета условий освещения служит двумерный массив, строки которого соответствуют строкам растрового поля светового окна. Каждая запись этого массива содержит величину ρ_S — расстояние от источника света до точки поверхности, соответствует элементу раstra светового поля и носит название ДЭС — дискретный элемент светности.

В ходе синтеза кадра изображения используются временные массивы. Временный массив является одномерным массивом, каждая запись которого содержит адресно-числовую информацию. Запись временного массива имеет вид

$$A_F, \rho_F, A_S, \rho_S, E, c, \quad (192)$$

где A_F — адрес некоторого ДЭКа; A_S — адрес некоторого ДЭСа. Эти адреса вырабатывают в процессе синтеза изображения, алгоритм которого будет рассмотрен в следующем параграфе.

3. ПАРАЛЛЕЛЬНЫЙ ПРОЦЕСС ЧИСЛЕННОГО СИНТЕЗА ТОНОВЫХ ИЗОБРАЖЕНИЙ

Высокая информативность тоновых изображений известна. Черно-белое телеизображение, например, содержит не менее 10^7 бит информации. Большой объем имеет информация, описывающая поверх-

ности объектов моделируемой сцены. В зарубежной практике численного моделирования трехмерных сцен для описания объектов широко используются многогранные поверхности [16]. В случае простых объектов этот способ оказывается более экономным по сравнению с заданием поверхностей дискретными элементами. Однако, чем «извилистее» поверхность, тем большее число граней приходится задавать для ее описания, тогда как число ДЭПов пропорционально площади поверхности и слабо зависит от кривизны. Кроме того, дискретное задание имеет определенное преимущество перед «многогранным» с точки зрения решения проблемы сокращения времени, затрачиваемого на численный синтез одного кадра изображения, так как обеспечивает возможность параллельных вычислений.

В случае сложных объектов численное моделирование сцен и синтез их изображений потребуют колоссальных вычислительных затрат. Так, для того чтобы вмонтированные в отснятый на 70-миллиметровой пленке мультфильм «численные» кадры были неотличимы от прочего рисованного материала, растровая мозаика должна содержать не менее $27 \cdot 10^6$ элементов. Синтез одного кадра на современной ЭВМ CRAY — 1S занимает 15 мин [33]. Новый двупроцессорный комплекс CRAY — X — MP имеет оперативную память на $4 \cdot 10^6$ 64-разрядных слов, а рабочий цикл центрального процессора составляет 10 нс. Ожидается, что на этом комплексе один кадр будет рассчитываться на 2,5 мин [33].

При использовании визуализации как средства обратной связи в интерактивных системах (например, в системах проектирования), работающих в режиме реального времени, вопрос ускорения численного синтеза изображений трехмерных сцен стоит особенно остро. Дальнейшее сокращение времени счета возможно при условии перехода к параллельным вычислениям на многопроцессорных комплексах. С этой точки зрения весьма существенной оказывается указанная выше взаимонезависимость ДЭПов. Она выражается в отсутствии в каждом дискретном элементе ссылок на другие, что предоставляет идеальные условия для «распараллеливания» процесса синтеза изображения.

Вычисление освещенности локальной площадки, описываемой ДЭПом, и получение ее проекции осуществляются только на основании информации, содержащейся в ДЭПе, общих для всех ДЭПов параметров сцены $\|F\|, \|S\|, l, p, q, \vec{O}_r O, \vec{O}_s O$ (векторы сдвига начал координат) и констант $e, g, \Delta, \mu, E_0, E_{\max}$. Эту часть вычислений мы будем называть процессом пространственного моделирования. Вычисления «пересекаются» на сортировке по глубине, предназначенной для отделения видимых элементов от невидимых и освещенных элементов от затененных. Если при численном синтезе используется разбиение пирамид видимости и освещения на малые телесные углы ω_{ij} , то взаимной сортировке по глубине подвергаются только те дискретные элементы, точки которых «вырезаются» из объекта конусом ω_{ij} .

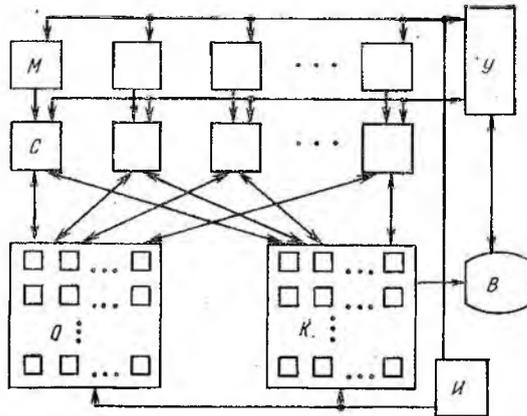


Рис. 71. Схема гибридного вычислительного комплекса

План реализации параллельной версии растрового алгоритма численного синтеза тоновых изображений рассмотрим на примере многопроцессорного вычислительного комплекса. В комплексе (рис. 71) можно условно выделить m идентичных параллельно работающих процессорных цепочек, общую память, управляющий блок $У$, блок исходного состояния $И$ и устройство визуализации $В$.

Каждая цепочка состоит из двух последовательно соединенных специализированных процессоров — моделирующего $М$ и сортировочного $С$. Пусть задание поверхностей объекта содержит N ДЭПов. В память каждого моделирующего процессора вводятся N/m ДЭПов исходного задания. Память каждого сортировочного процессора служит для приема от моделирующего записей временного массива (192). Общая память разделена на две секции — картинную K и осветительную O . Каждая секция составлена из адресуемых модулей, число которых равно числу элементов раstra устройства визуализации. Адреса модулей картинной секции будем обозначать A_P , адреса модулей осветительной секции — A_S . Каждый модуль картинной секции может содержать один ДЭК: каждый модуль осветительной секции — один ДЭС.

Сортировочный процессор каждой цепочки получает доступ к любому модулю секций K или O , выставляя в соответствующем регистре запросов требуемый адрес. Блок $У$ координирует работу всего комплекса, переключает процессоры на тот или иной режим работы, принимает от них сигналы об окончании работы, запускает в работу блоки $И$ и $В$.

Процесс пространственного моделирования реализуется каждым процессором $М$ над имеющимися в его памяти ДЭПами. В режиме полной обработки процессор $М$ запрограммирован на выполнение следующей последовательности операций.

1. Если массив ДЭПов в памяти данного моделирующего процессора исчерпан, обработка прекращается, иначе — выбирается очередной ДЭП.

2. Элементам результирующей записи (временного массива) присваиваются значения: $A_F = \rho_F = A_S = \rho_S = 0$.

3. Вычисляются координаты точки (из ДЭПа) и вектора нормали в системе пирамиды видимости:

$$(X, Y, Z) = \|F\| \begin{pmatrix} x \\ y \\ z \end{pmatrix} + O_F \vec{O}; \quad (X_n, Y_n, Z_n) = \|F\| \begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix},$$

а также координаты центральной проекции точки на картинной плоскости:

$$X' = \frac{\mu_F |\rho|}{|Z + \rho|} X; \quad Y' = \frac{\mu_F |\rho|}{|Z + \rho|} Y.$$

4. Если проекция (X, Y) вышла за пределы кадра, то адрес соответствующего ДЭКа полагается нулевым: $A_F = 0$. Иначе — адрес ДЭКа получается согласно правилу

$$A_F = \left[\text{ent} \left(e_F - \frac{Y'}{\Delta_F} \right) \right] 2g_F + \text{ent} \left(\frac{X'}{\Delta_F} + g_F \right) + 1.$$

5. Вычисляются координаты точки и компоненты вектора нормали в системе пирамиды освещения:

$$(\mathcal{X}, \mathcal{Y}, \mathcal{Z}) = \|S\| \begin{pmatrix} x \\ y \\ z \end{pmatrix} + O_S \vec{O}; \quad (\mathcal{X}_n, \mathcal{Y}_n, \mathcal{Z}_n) = \|S\| \begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix},$$

а также координаты проекции точки на световой плоскости:

$$\mathcal{X}' = \frac{\mu_S |q|}{|\mathcal{Z} + q|} \mathcal{X}; \quad \mathcal{Y}' = \frac{\mu_S |q|}{|\mathcal{Z} + q|} \mathcal{Y}.$$

6. Если проекция $(\mathcal{X}', \mathcal{Y}')$ не вышла за пределы светового окна, то адрес соответствующего ДЭСа получается по правилу

$$A_S = \left[\text{ent} \left(e_S - \frac{\mathcal{Y}'}{\Delta_S} \right) \right] 2g_S + \text{ent} \left(\frac{\mathcal{X}'}{\Delta_S} + g_S \right) + 1.$$

Иначе — полагается $A_S = 0$ и осуществляется переход к п. 10.

7. Вычисляются расстояние¹ от точки до источника света $\rho_S = |\mathcal{X}| + |\mathcal{Y}| + |\mathcal{Z} + q|$ и величина

$$c = - \frac{\mathcal{X}_n \mathcal{X} + \mathcal{Y}_n \mathcal{Y} + \mathcal{Z}_n (\mathcal{Z} + q)}{\rho_S}, \quad (c \approx \cos \theta),$$

а возможная яркость элемента растра и его цвет полагаются равными

$$E_{ij} = \min \left\{ \left| \frac{I c a_{\text{sign } c}}{\rho_S^2} \right|, E_M \right\}; \quad u_{ij} = u_{\text{sign } c}.$$

8. Если $E_{ij} \leq E_0$, полагается $E_{ij} = E_0$.

¹ Неевклидова метрика используется для ускорения счета.

9. Вычисляется расстояние от точки ДЭПа до точки наблюдения

$$\rho_F = |X| + |Y| + |Z + \rho|.$$

10. Сортировочному процессору данной цепочки передается результирующая запись $(A_F, \rho_F, A_S, \rho_S, E_{ij}, \varphi_{ij})$ для временного массива. Производится возврат к п. 1.

Сокращенный вариант процесса пространственного моделирования отличается от изложенного полного варианта концовкой п. 6 и формулировкой п. 8:

6. ... Иначе — производится возврат к п. 1.

8. Если $E_{ij} \leq E_0$, производится возврат к п. 1.

Процесс пространственного моделирования (в обоих вариантах) заканчивается при исчерпании массива ДЭПов.

Процессоры C всех цепочек запрограммированы на выполнение последовательности операций сортировки.

Процесс сортировки применяется к каждой k -й записи временного массива. Значение k счетчика заполнения временного массива увеличивается на единицу, когда процессор C принимает результирующую запись от своего процессора M . Значение k уменьшается на единицу, когда процессор C выбирает очередную запись для обработки процессором сортировки. Полный вариант процесса сортировки состоит в следующем.

1. Осуществляется выборка k -й записи $(A_F, \rho_F, A_S, \varphi_S, E_{ij}, \varphi_{ij})$ из временного массива, и значение k счетчика заполнения уменьшается на единицу.

2. Если $A_F \neq 0$, выполняется переход к п. 4. Иначе —

3. Проверяется A_S . Если $A_S \neq 0$, выполняется переход к п. 8. Иначе —

4. Осуществляется выборка ДЭКа $(\bar{\rho}_F, \bar{E}, \bar{\varphi})$ из адреса A_F .

5. Блокируется доступ к A_F и проверяется A_S . Если $A_S = 0$, полагается $E_{ij} = E_0$ и осуществляется переход к п. 12. Иначе —

6. Сравниваются значения ρ_F и $\bar{\rho}_F$. Если $\rho_F > \bar{\rho}_F$, осуществляется переход к п. 19. Иначе —

7. Сравниваются значения E_{ij} и E_0 . Если $E_{ij} = E_0$, то выполняется переход к п. 19. Иначе —

8. Осуществляется выборка ДЭСа $(\bar{\rho}_S)$ из A_S .

9. Блокируется доступ к адресу A_S .

10. Сравниваются значения ρ_S и $\bar{\rho}_S$. Если $\rho_S \leq \bar{\rho}_S$, полагается $\bar{\rho}_S = \rho_S$ и осуществляется переход к п. 12. Иначе —

11. Полагается $E_{ij} = E_0$ и освобождается доступ к A_S .

12. Проверяется A_F . Если $A_F = 0$, осуществляется переход к п. 17. Иначе —

13. Сравниваются ρ_F и $\bar{\rho}_F$. При $\rho_F \geq \bar{\rho}_F$ производится переход к п. 16. При $\rho_F < \bar{\rho}_F$ выполняется следующий шаг.

14. Полагается $\bar{\rho}_F = \rho_F$, $\bar{\varphi} = \varphi$.

15. Полагается $\bar{E} = E_{ij}$, новый ДЭК возвращается по адресу A_F . Доступ к A_F освобождается.

16. Проверяется A_S . Если $A_S = 0$, осуществляется переход к п. 19. Иначе —

17. Сравниваются ρ_S и $\bar{\rho}_S$. Если $\rho_S \geq \bar{\rho}_S$, выполняется переход к п. 19. Иначе —

18. Новый ДЭС возвращается по адресу A_S , и доступ к A_S освобождается.

19. Проверяется содержимое счетчика k . Если $k \neq 0$, осуществляется возврат к п. 1. Иначе —

20. При наличии сигнала о завершении процесса пространственного моделирования от процессора M процессор C заканчивает работу. Иначе — возврат к п. 19.

Сокращенный вариант процесса сортировки отличается от изложенного полного варианта шагами 4, 8, 10, 11, 15, которые в сокращенном варианте процесса имеют следующий вид.

4. Осуществляется выборка ДЭКа ($\bar{\rho}_F, \bar{E}, \bar{c}$) из адреса A_F и переход к п. 6.

8. Производится выборка ДЭСа ($\bar{\rho}_S$) из адреса A_S и переход к п. 10.

10. Сравниваются ρ_S и $\bar{\rho}_S$. Если $\rho_S \leq \bar{\rho}_S$, выполняется переход к п. 19. Иначе —

11. Полагается $E_{ij} = E_0$, и осуществляется переход к п. 15.

15. Полагается $\bar{E} = E_{ij}$, новый ДЭК возвращается по адресу A_F . Доступ к A_F освобождается, и выполняется переход к п. 19.

Работа вычислительного комплекса в целом происходит следующим образом. Управляющий блок $У$ запускает блок исходного состояния $И$, который передает всем моделирующим процессорам параметры сцены и заполняет модули картинной секции общей памяти исходными ДЭКа́ми — $\rho_F^{\max}, E_\phi, c_\phi$, а модули осветительной секции — исходными ДЭСа́ми — ρ_S^{\max} , где ρ^{\max} — максимальное значение расстояний; E_ϕ, c_ϕ — фоновые значения яркости и цвета. В это время управляющий блок настраивает процессоры M и C всех цепочек на режим полной обработки и, получив от блока $И$ сигнал завершения, запускает в работу все цепочки комплекса.

Каждый моделирующий процессор из своей исходной памяти считывает ДЭП за ДЭПом, применяет к каждому ДЭПу процедуру пространственного моделирования и результаты в виде записей временного массива передает процессору сортировки своей цепочки. Сортировочный процессор применяет к каждой записи временного массива процедуру сортировки, возможно, обращая при этом для чтения или записи к некоторым модулям общей памяти. Когда все ДЭПы, принадлежащие на данную цепочку, исчерпаны и обработаны все записи временного массива, сортировочный процессор выдает управляющему блоку сигнал о завершении работы цепочки.

Получив подобные сигналы от всех цепочек, управляющий блок переключает все процессоры на режим сокращенной обработки и запускает их в работу. Все ДЭПы повторно перебираются и обрабатываются по сокращенному варианту моделирования и сортировки, что необходимо для учета взаимного затенения поверхностей

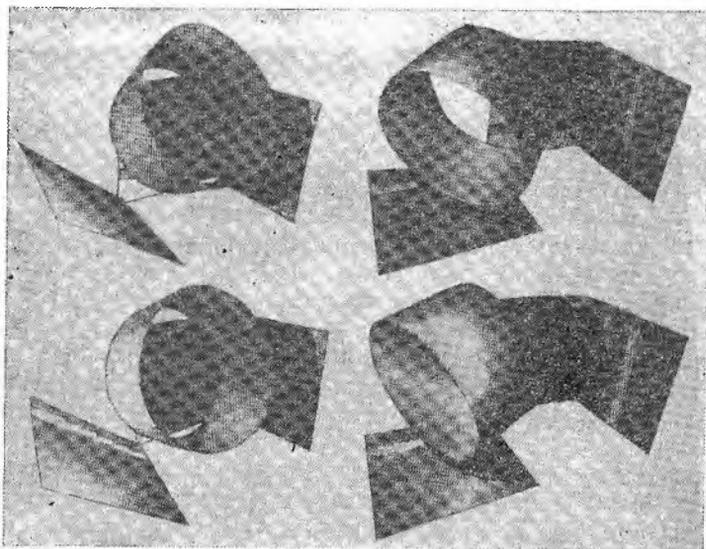


Рис. 72. Серия тоновых изображений динамической сцены

сцены. С завершением работы всех цепочек по сокращенному варианту заканчивается синтез кадра изображения. Управляющий блок запускает визуализатор *B*, который считывает из картинной секции общей памяти информацию о синтезированном изображении. При этом визуализатор преобразует цифровые данные в видеосигнал. Последний отображается на экране телеприемника, он также может быть зафиксирован на магнитном носителе. Визуализатор выдает управляющему блоку сигнал завершения, и комплекс запускается для синтеза следующего кадра, и т. д.

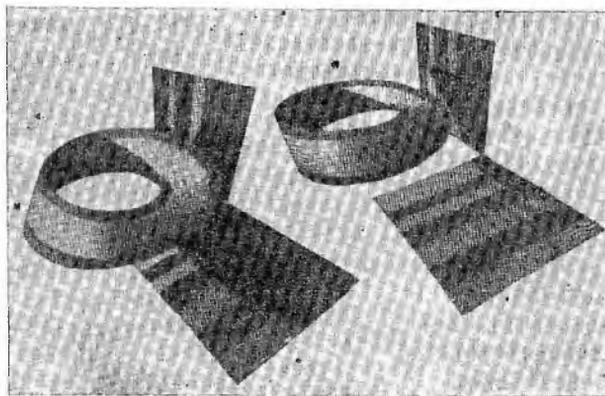
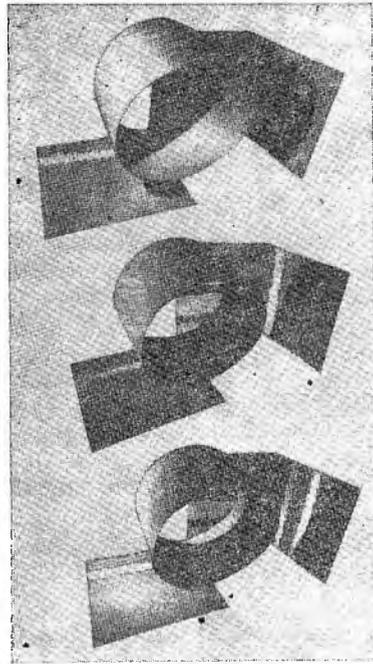


Рис. 73. Тоновые изображения ленты Мебиуса

Рис. 74. Иллюстрация эффекта перемещения источника света

Если параметры сцены целенаправленно изменяются от кадра к кадру, то синтезируется кинограмма динамической эволюции объекта (рис. 72). На рис. 73 приведены другие примеры тоновых изображений, полученных с помощью одной цифровой ЭВМ на основании растрового алгоритма. Рис. 74 интересен тем, что на нем запечатлен результат перемещения в пространстве источника света.

Благодаря тому, что дискретные элементы, задающие поверхности сцены, обрабатываются независимо, процессы моделирования и сортировки могут быть разложены на большое число параллельных ветвей. При условии равномерного распределения ДЭПов по всем вычислительным цепочкам комплекса количество ДЭПов, приходящихся на одну цепочку, уменьшается с ростом их числа. Иначе говоря, с увеличением степени параллельности можно ожидать сокращения времени синтеза одного кадра изображения. Однако оптимистический вывод здесь не так уж очевиден, так как выигрыш по времени счета зависит от организации или, как говорят, от архитектуры вычислительного комплекса.



4. ЗАВИСИМОСТЬ ВРЕМЕНИ ПАРАЛЛЕЛЬНОГО СИНТЕЗА ИЗОБРАЖЕНИЯ ОТ ТИПА МЕЖПРОЦЕССОРНОГО ВЗАИМОДЕЙСТВИЯ

Пусть решается неполная задача синтеза, т. е. учитывается только взаимоэкранирование поверхностей сцены, а их взаимное затенение игнорируется. Предположим, что имеется возможность неограниченно наращивать число процессорных цепочек m вычислительного комплекса. При любом N , характеризующем сложность объекта сцены, можно выбрать $m = N$, так что на каждую цепочку придется ровно один ДЭП. Тогда процесс пространственного моделирования будет завершаться за постоянное время t_m при общих вычислительных затратах (с учетом всех выполненных элементарных операций), пропорциональных сложности объекта, т. е. числу N .

Время сортировки по глубине зависит от непредсказуемого распределения точек по элементарным телесным углам ω_{ij} . Возьмем тривиальный случай: заранее известно, что все N ДЭПов попадают в один элементарный телесный угол. Сортировку по глубине можно

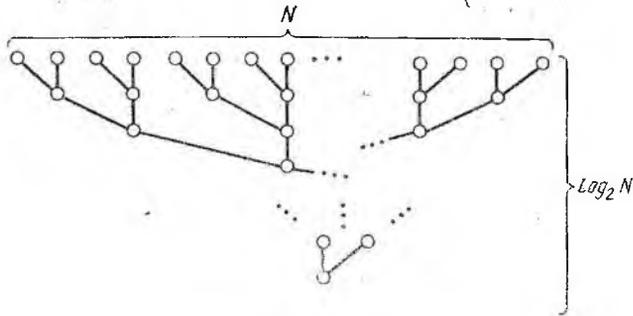


Рис. 75. Каскад сравнений

организовать в форме каскада попарных сравнений (рис. 75). Число ступенек в таком каскаде равно (с точностью до ближайшего целого) $\log_2 N$, и, на первый взгляд, время, затрачиваемое на сортировку, пропорционально логарифму сложности объекта: $t_c = \tau \log_2 N$, где τ — некоторая константа, зависящая от быстродействия вычислительного устройства. Таким образом, общее время синтеза t складывается из затрат на моделирование и сортировку:

$$t = (t_M + \tau \log_2 N). \quad (193)$$

Необходимо отметить, что имеется определенный разрыв между оценками временной сложности собственно алгоритмов и возможностью реализации этих алгоритмов на вычислительных устройствах. Проиллюстрируем это расхождение на примере каскадного алгоритма сортировки по глубине. Пусть сортировочные процессоры в устройстве выстроены в линейку с шагом по расстоянию, равным единице. При этом дискретный линейный размер системы $L = N - 1$. Тогда для перехода ко второй ступени каскада необходимо передать информацию соседу, т. е. на расстояние 1. При переходе к третьей ступени среднее расстояние передачи составит 2, при переходе к четвертой ступени — 4 и т. д. Для последней ступени среднее расстояние передачи достигнет значения $2^{(\log_2 N) - 1}$. Итого: $2^0 + 2^1 + 2^2 + \dots + N/2 = N - 1$.

Передача данных в реальной вычислительной системе осуществляется не мгновенно. Каждый процессор, даже если он расположен от соседа на пренебрежимо малом физическом расстоянии, передавая информацию, т. е. выполняя функцию ретранслятора, тратит на каждое слово конечное время λ . Поэтому при использовании растрового алгоритма для реальной вычислительной системы следует учитывать, что оценка (193) не будет соответствовать действительному положению вещей.

Если каждый акт взаимодействия между соседними процессорами системы происходит совершенно независимо от других процессоров, то коммуникации между любыми (не обязательно близкими) элементами вычислительной системы (ВС), осуществляемые посредством таких актов, называются локальными взаимодействиями. Много-

процессорное устройство, работа которого основана на коммуникациях такого рода, называется асинхронной ВС с локальными взаимодействиями. В подобной ВС «согласование (синхронизация) работы элементов (процессоров) осуществляется посредством локальных операторов, когда каждый элемент системы взаимодействует с соседними элементами по мере готовности их к взаимодействию» [14].

Коммуникации через локальные взаимодействия элементов. При существующей технологии производства сверхбольших интегральных схем (СБИС) наиболее реальным представляется размещение элементов ВС на плоскости в виде упорядоченной структуры, где каждый элемент имеет небольшое число соседей, с которыми он связан непосредственно. Коммуникации между удаленными элементами при такой организации ВС реализуются только с помощью промежуточных элементов, выполняющих по мере необходимости ретрансляционную функцию. Каждый элемент подобной системы характеризуется двумя временными константами: задержкой ретрансляции и вычислительной задержкой. Как правило, задержка ретрансляции много меньше вычислительной.

Кольцевая асинхронная ВС с локальными взаимодействиями, схема которой приведена на рис. 76, ориентирована на реализацию растрового алгоритма. В ней объединены в кольцо модули памяти Π каждого сортировочного процессора, и все сортировочные процессоры C также объединены в кольцо. Здесь каждый моделирующий процессор M , как и в гибридной системе (см. рис. 71), подключен непосредственно к своему сортировочному процессору. При условии, что на один элемент раstra приходится N/R точек, время t синтеза одного кадра изображения на системе, содержащей m моделирующих процессоров, оценивается неравенством

$$t \leq \frac{N}{m} \left[t_M + \frac{N}{R} \left(\tau + \frac{\kappa R}{m} \right) + \kappa (m - 1) \right] = T(m), \quad (194)$$

где τ — время собственно сравнения; κ — время выполнения элементом системы (процессором) ретрансляционной функции; R — число элементов раstra устройства визуализации. Число модулей памяти Π , приходящихся на один сортировочный процессор, равно R/m , и максимальное время доступа к модулю составляет $\kappa R/m$.

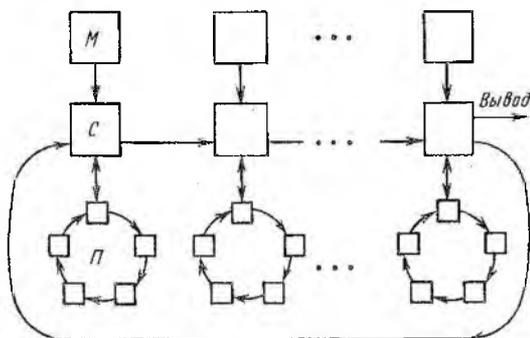


Рис. 76. Схема параллельного синтеза изображений с использованием кольцевых многопроцессорных структур

По завершении синтеза кольцо сортировочных процессоров разрывается в одном месте, и вывод элементов изображения осуществляется через один крайний в образовавшейся «линейке» процессор. Этот процессор сначала разгружает свою память Π , затем переходит в режим ретрансляции и выводит элементы, получаемые от соседа. Последний, разгрузив свое кольцо, тоже переходит на ретрансляцию элементов, получаемых от следующего соседа, и т. д. Разумеется, порядок распределения элементов (i, j) по сортировочным процессорам и по модулям колец должен соответствовать строчной развертке изображения.

Глобальные взаимодействия элементов ВС являются антиподом локальных взаимодействий. Акт глобального взаимодействия начинается с момента посылки информационного слова передающим элементом и завершается в момент получения этим элементом сигнала о приеме слова от принимающего элемента. В течение времени коммуникации все промежуточные элементы выполняют функцию ретрансляции с характерным временем κ . Поэтому время одного акта глобального взаимодействия пропорционально дискретному расстоянию (числу промежуточных элементов) между передающим и принимающим элементами (процессорами). Если в системе на рис. 76 считать взаимодействия глобальными, то оценка времени синтеза примет вид

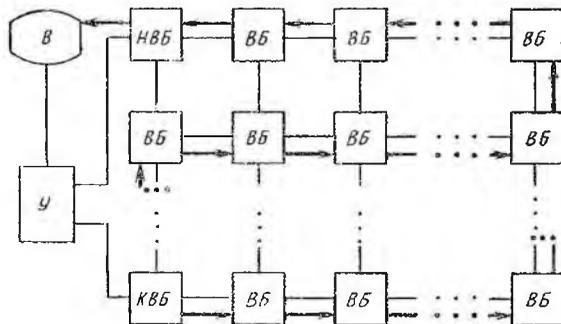
$$t_{\text{глоб}} \leq \frac{N}{m} \left[t_M + \frac{N}{R} \left(\tau + \frac{\kappa R}{m} \right) + m\kappa(m-1) \right] = T'_{\text{глоб}}(m). \quad (195)$$

Число элементарных операций в процессе пространственного моделирования не превышает 10^3 . При быстродействии процессоров 10^8 с^{-1} время $t_M = 10^{-5} \text{ с}$. Число операций при однократном выполнении собственно сравнения не больше 10, поэтому $\tau = 10^{-7} \text{ с}$. Примем время выполнения ретрансляционной функции $\kappa = 10^{-9} \text{ с}$, число элементов раstra $R = 10^6$ и сложность сцены $N = 10^7$. Тогда из (194) получим, что время синтеза на однопроцессорной «системе» ($m = 1$) оценивается значением $T(1) \approx 10^5 \text{ с}$. Частота следования кадров 25 с^{-1} является стандартной для телевидения, поэтому для отображения динамического процесса необходимо синтезировать кадр за время, не большее 0,04 с. В условиях нашего примера значение $T(m) = 0,04 \text{ с}$ достигается при числе процессоров $m \approx 5 \cdot 10^3$. При тех же значениях коэффициентов для глобального варианта из (195) получается, что минимальное значение оценки времени синтеза, равное 8,5 с (что на два порядка превышает стандарт), достигается при числе процессоров $m \approx 280$.

Из всего сказанного следует вывод: решение задачи синтеза тоновых изображений сложных сцен в режиме реального времени возможно только на асинхронных ВС с локальными взаимодействиями элементов.

Вычислительная система. Рассмотрим схему вычислительной системы, ориентированной на применение растрового алгоритма в его максимально параллельной версии. Полный учет этим алгоритмом условий освещения и генерирование теней, отбрасываемых одними поверхностями объекта на другие, позволит достигнуть

Рис. 77. Решетчатая структура асинхронной ВС, предназначенной для параллельного синтеза тоновых изображений



максимальной естественности восприятия объемной сцены через ее плоское изображение.

Асинхронный многопроцессорный вычислительный комплекс с локальными коммуникациями элементов представляет собой прямоугольную решетку из $2e \times 2g$ идентичных вычислительных блоков, каждый из которых соответствует одному элементу раstra ($2e$ — число строк, $2g$ — число элементов разложения строки) (рис. 77). Благодаря технологически выполненному перегибу между e -й и $(e + 1)$ -й строками решетка имеет двуслойную пространственную конфигурацию, и каждая i -я строка налагается на $(2e + 1 - i)$ -ю строку. Таким образом, начальный вычислительный блок $НВБ$ и конечный вычислительный блок $КВБ$ являются соседями. Эти блоки отличаются от вычислительных блоков $ВБ$ только дополнительными связями с управляющим блоком $У$ и визуализатором $В$. Предполагается, что через блок $У$ в комплекс поступают извне исходные данные и директивы операторов, а через визуализатор из комплекса выводятся синтезированные изображения.

Каждый вычислительный блок состоит из нескольких специализированных процессоров. Отметим из них следующие: вводный процессор, процессор пространственного моделирования, картинный процессор, светностный процессор, выводной процессор. Процессоры вводной, картинный и светностный каждого вычислительного блока соединены с одноименными процессорами соседних блоков. Таким образом, фактически имеются три вложенные друг в друга двумерные решетки (на рис. 77 показаны тонкими двунаправленными связями). Выводные процессоры всех вычислительных блоков имеют однонаправленные связи по строкам (жирные стрелки на схеме) с единичными переходами от строки к строке на крайних блоках. В силу этого на двумерные решетки наложена «змейкой» одномерная последовательность связей. Все указанные на схеме соединения отражают направления потоков содержательной и директивной информации, циркулирующей в системе. Само локальное взаимодействие по принципу запрос — ответ между непосредственно связанными процессорами может быть конструктивно реализовано различными способами (см., например, [14]).

Каждый вычислительный блок, будучи элементом решетки, имеет двумерный адрес (i, j) , где i — номер строки; j — номер столбца. В качестве элемента последовательности каждый вычислительный блок имеет одномерный адрес n , являющийся его порядковым номером. Начальный вычислительный блок имеет двумерный адрес $(1, 1)$ и одномерный адрес 1. Конечный вычислительный блок имеет двумерный адрес $(2e, 1)$ и одномерный адрес $1 + 2g(2e - 1)$.

Процессоры комплекса могут функционировать в нескольких рабочих режимах: загрузка, моделирование, затенение, вывод. Вспомогательным режимом является настройка, в ходе которой процессор производит переключения в своей внутренней структуре для выполнения новой функции. Настройка осуществляется по приему соответствующей команды, которая поступает от управляющего блока. По решетке команды распространяются через начальный и конечный вычислительные блоки. Распространение команды по решетке имеет характер двумерной волны: каждый блок крайнего левого столбца ($j = 1$) передает команду своему соседу по столбцу и соседу по строке, а все остальные блоки передают команду только соседу по строке. Для процессоров возможны еще режимы ретрансляции и ожидания. В последнем случае процессор не выполняет никаких операций, а лишь выдает запрос на взаимодействие с соседями или сигнал готовности к взаимодействию.

Синтез изображения на асинхронном комплексе. Этот процесс предполагает предварительную загрузку ДЭПов и равномерное их распределение по вычислительным блокам. Задание моделируемой сцены (используемые системы координат, параметры и константы) то же, что было введено в п. 2 настоящей главы (см. рис. 70). Моделирующие процессоры каждого блока реализуют полный вариант процесса пространственного моделирования, некоторые пункты которого в данной версии растрового алгоритма имеют следующий модифицированный вид.

2. Нет операции.

4. Если проекция (X', Y') вышла за пределы кадра, то двумерный адрес (i, j) полагается нулевым — $(0, 0)$. Иначе — его компоненты вычисляются по формулам

$$i = \text{ent}(e + X'/\Delta); \quad j = \text{ent}(g - Y'/\Delta).$$

6. Если проекция $(\mathcal{X}, \mathcal{Y})$ не вышла за пределы светового окна, то двумерный адрес (k, l) вычисляется по формулам

$$k = \text{ent}(e + \mathcal{X}'/\Delta); \quad l = \text{ent}(g - \mathcal{Y}'/\Delta).$$

Иначе — $k = 0, l = 0$, и результирующая яркость соответствующего элемента раstra полагается минимальной: $E_{ij} = E_0$. Осуществляется переход к п. 10.

10. Полагаются $\rho_{ij} = \rho_r, \rho_{kl} = \rho_s$. Результирующее слово $((i, j), \rho_{ij}, E_{ij}, \rho_{kl})$ передается картинному процессору, другое результи-

рующее слово $((i, j), (k, l), \rho_{ij}, \rho_{kl})$ передается светностному процессору в этом же вычислительном блоке.

Взаимное экранирование и затенение поверхностей сцены учитывается асинхронным вычислительным комплексом с помощью дальних коммуникаций между вычислительными блоками, осуществляемых посредством серии локальных взаимодействий.

Картинный процессор имеет в своей памяти двумерный адрес (I, J) вычислительного блока, к которому этот процессор принадлежит. Точно так же светностный процессор имеет двумерный адрес (K, L) своего вычислительного блока. Естественно, в одном вычислительном блоке имеют место равенства $I = K; J = L$. В режиме моделирования картинный и светностный процессоры выполняют следующее.

Каждый картинный процессор вычислительного блока обрабатывает слова $((i, j), \rho_{ij}, E_{ij}, u_{ij})$, поступающие в него либо от моделирующего процессора либо от соседей по решетке. Обработка заключается в следующем. В случае совпадения (i, j) с (I, J) при $\rho_{ij} < \rho_{IJ}$ картинный процессор заменяет значения содержащихся в нем величин $\rho_{IJ}, E_{IJ}, u_{IJ}$ на $\rho_{ij}, E_{ij}, u_{ij}$, а при $\rho_{ij} \geq \rho_{IJ}$ прежние значения сохраняются. При несовпадении (i, j) с (I, J) процессор передает слово $((i, j), \rho_{ij}, E_{ij}, u_{ij})$ тому из соседей по решетке, двумерный адрес (I', J') которого удовлетворяет неравенству $|i - I'| + |j - J'| < |i - I| + |j - J|$.

Каждый светностный процессор обрабатывает поступающие от моделирующего процессора или соседей по решетке слова вида $((i, j), (k, l), \rho_{ij}, \rho_{kl})$ согласно следующим правилам. Если (k, l) совпадает с (K, L) , то при $\rho_{kl} < \rho_{KL}$ светностный процессор заменяет значения содержащихся в нем величин $\rho_{KL}, (i, j)_{KL}, (\rho_{ij})_{KL}$ на $\rho_{kl}, (i, j), \rho_{ij}$, а при $\rho_{kl} \geq \rho_{KL}$ прежние значения сохраняются. Если (k, l) не совпадает с (K, L) , то процессор передает полученное слово $((i, j), (k, l), \rho_{ij}, \rho_{kl})$ тому из соседних светностных процессоров, двумерный адрес (K', L') которого удовлетворяет неравенству $|k - K'| + |l - L'| < |k - K| + |l - L|$.

Рассматриваемый вычислительный комплекс в целом функционирует следующим образом. Управляющий блок посылает по решетке (через начальный и конечный вычислительные блоки) серию параметров сцены, которые запоминаются моделирующим процессором каждого вычислительного блока.

В каждом вычислительном блоке картинный процессор устанавливает $\rho_{IJ} = \rho^{\max}$, $E_{IJ} = E_{\phi}$, $u_{IJ} = u_{\phi}$ и настраивается на режим моделирования. Светностный процессор устанавливает $\rho_{KL} = \rho^{\max}$ и тоже настраивается на моделирование. Выводной процессор остается в режиме ожидания. Управляющий блок выдает по решетке команду начала синтеза кадра и запускает счет времени. Получив эту команду, моделирующий, картинный и светностный процессоры каждого вычислительного блока начинают работать в режиме моделирования, обрабатывая один за другим ДЭПы, загруженные в данный вычислительный блок. После обработки ДЭПов процессоры переходят в режим ожидания.

Пусть t — это время, за которое необходимо синтезировать кадр изображения. По истечении времени $\vartheta_1 t$ ($\vartheta_1 < 1$) с момента начала синтеза кадра управляющий блок посылает по решетке команду настройки на режим затенения, вслед за которой начинает передавать параметры сцены для синтеза следующего кадра изображения. Моделирующие процессоры по этой команде настраиваются на режим загрузки и принимают новые параметры. Картинные и светностные процессоры настраиваются на режим затенения.

Содержащиеся в картинных процессорах элементы изображения (E_{ij}, u_{ij}) имеют в своем двоичном представлении разряд-метку. Исходные фоновые элементы (E_Φ, u_Φ) содержат в этом разряде единицу, а элементы изображения, получающиеся в ходе моделирования, — нуль.

Процесс затенения предназначен для приписывания единичных меток всем освещенным элементам и только им. Он состоит в следующем. В блоках решетки с адресами (K', L') , где после моделирования оказались минимальные значения $\rho_{K'L'} < \rho^{\max}$, светностные процессоры стремятся передать имеющиеся у них величины (i, j) и ρ_{ij} светностному процессору с адресом $(K, L) = (i, j)$. Эти передачи осуществляются по кратчайшему пути согласно правилу

$$|i - k| + |j - l| < |i - K'| + |j - L'|,$$

где (k, l) — адрес соседа по решетке, принимающего информацию в данном локальном взаимодействии. В случае несовпадения пришедшего адреса (i, j) с адресом (K, L) принимающего процессора величины (i, j) и ρ_{ij} передаются далее. Когда (i, j) совпадает с (K, L) , светностный процессор передает пришедшую величину ρ_{ij} в своем блоке $(I = K, J = L)$ картинному процессору, который сравнивает полученное значение ρ_{ij} с имеющимся у него значением ρ_{IJ} . При $\rho_{ij} \leq \rho_{IJ}$ элемент изображения метится, т. е. разряд-метка устанавливается в единицу. В противном случае элемент (E_{ij}, u_{ij}) остается с прежней меткой. По завершении необходимых коммуникаций картинные и светностные процессоры всех вычислительных блоков переходят в ожидание.

Пока идет затенение, управляющий блок передает серию новых параметров сцены. Признак конца этой серии является командой, по которой моделирующие процессоры всех вычислительных блоков переходят в режим моделирования и приступают к синтезу следующего кадра изображения. По истечении времени $\vartheta_2 t$ ($\vartheta_1 < \vartheta_2 < 1$) с начала синтеза управляющий блок запускает визуализатор на вывод кадра изображения и передает по решетке команду начала синтеза следующего кадра и в момент t начинает счет времени нового кадра. По этой команде картинный процессор вычислительного блока передает содержащийся в нем элемент изображения (E_{IJ}, u_{IJ}) своему выводному процессору, устанавливает $\rho_{IJ} = \rho^{\max}$, $E_{IJ} = E_\Phi$, $u_{IJ} = u_\Phi$ и настраивается на режим моделирования; светностный процессор устанавливает $\rho_{KL} = \rho^{\max}$ и тоже настраивается на моде-

лирование. Каждый выводной процессор, получив от картинного процессора своего блока элемент изображения, настраивается на режим вывода. Элемент изображения, находящийся в выводном процессоре конечного вычислительного блока, получает специальный признак конца развертки изображения.

Параллельно процессу моделирования нового кадра изображения происходит вывод предыдущего кадра, который сводится к ретрансляции элементов изображения (E_{ij} , u_{ij}) по одномерной змеевидной цепочке выводных процессоров в сторону убывания номеров n . Выводной процессор конечного вычислительного блока передает свой элемент, снабженный признаком конца развертки, соседу и переходит в состояние ожидания. Перемещаясь по цепочке, этот концевой элемент переводит один за другим выводные процессоры в ожидание. Визуализатор принимает элементы изображения и с учетом меток преобразует в видеосигнал. Каждому элементу изображения с нулевым разрядом метки сопоставляется минимальная яркость, чем и реализуются тени.

Вывод изображения на визуализатор. Упорядоченный вывод элементов изображения из вычислительного комплекса не сопряжен ни с алгоритмическими, ни с техническими трудностями. Трудоемкость вывода не превосходит трудоемкости процесса моделирования. Во всяком случае, через время Φ_{2l} выводные процессоры всех блоков должны быть готовы к приему элементов изображения следующего кадра.

Отметим любопытную деталь в реализации растрового алгоритма на многопроцессорном комплексе. Мы имеем здесь пример не вполне детерминированного вычислительного процесса, поскольку управляющий блок не получает от процессоров сигналов о завершении ими обработки информации. Все команды, регулирующие необходимое чередование различных этапов вычислительного процесса, управляющий блок выдает строго в соответствии с показаниями собственного таймера. Коэффициенты Φ_1 , Φ_2 и число ДЭПов в задании сцены являются эмпирическими константами конкретного вычислительного устройства. Значения их зависят от быстродействия электронных микросхем и должны обеспечивать стандартный темп генерирования изображений.

Моделирование прозрачных и полупрозрачных поверхностей, называемых для краткости ниже светофильтрами, также возможно на основе растрового алгоритма. Для учета оптических свойств объекта необходима дополнительная информация. В каждом из ДЭПов должен быть задан показатель прозрачности, который принимает ненулевое значение на светофильтрах, а в дискретных элементах картины (ДЭК) должно быть зарезервировано место под величину ρ_{ij} — расстояние от центра проекции до некоторой точки светофильтра. Тогда в процессе синтеза, если в данном угле ω_{ij} точка светофильтра оказывается ближе к центру, чем предыдущая точка непрозрачной поверхности, то старое значение ρ_{ij} в ДЭКе сохраняется, значения E_{ij} и u_{ij} изменяются с учетом оптических свойств светофильтра, а расстояние от центра до точки светофильтра заново

сится на место ρ_ϕ . В дальнейшем для всякой точки непрозрачной поверхности, у которой $\rho_{ij} > \rho > \rho_\phi$, значение ρ_{ij} заменяется на ρ , а новые E_{ij} и ψ_{ij} модифицируются с учетом светофильтра. Только точка непрозрачной поверхности, у которой $\rho < \rho_\phi$, аннулирует светофильтр в данном ω_{ij} .

Аналогичным образом могут быть учтены ситуации, когда светофильтр просто «выступает» из фона, не закрывая никаких поверхностей, или когда перекрываются разные светофильтры. Очевидно, все это усложняет вычислительный процесс. Однако увеличиваются только значения t_M и t , тогда как стоящие в правых частях оценок выражения остаются в силе.

Численному моделированию вполне доступны и такие оптические нюансы, как рассеяние света, блики, текстура поверхности и т. п. Достижимый при учете всех этих признаков эффект информативности визуализации способствует усилению достоверности восприятия синтезированного изображения.

Глава VII

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ГЕОМЕТРИЧЕСКИХ ДАННЫХ

В настоящей главе речь пойдет об автоматизированной системе геометрических расчетов АСТРА, разработанной в Институте математики СО АН СССР. Ее назначение — геометрическое обеспечение процессов автоматизированного проектирования и технологической подготовки производства изделий, имеющих сложные пространственные обводы. К ним относятся объекты таких отраслей, как авиастроение, судостроение, автомобилестроение и т. п.

1. ОБЩАЯ ХАРАКТЕРИСТИКА СИСТЕМЫ АСТРА

На всех этапах создания изделия происходит обработка в той или иной форме различной геометрической информации, относящейся к этому изделию. При проектировании исходные данные об объекте преобразуются в геометрические образы его отдельных агрегатов и синтезируется образ самого изделия — математическая модель объекта. Предварительная оценка проектного решения и этап конструирования включают комплекс расчетов для выявления характеристик (прочностных, гидродинамических и т. п.) изделия и его элементов. Источником информации на этом этапе, а также при подготовке производства, когда проектируется и изготавливается необходимая технологическая оснастка, и, наконец, в процессе контроля изготовленных узлов и деталей является математическая модель объекта. Таким образом, все расчеты, выполняемые в процессе создания объекта, базируются на геометрической информации об изделии, которая как бы объединяет все этапы указанного процесса. Это обстоятельство находит свое отражение в автоматизированных системах проектирования и технологической подготовки производства. Основной компонентой этих систем является подсистема геометрического моделирования. В качестве такой подсистемы может быть использована система АСТРА.

Система АСТРА универсальна, т. е. ее можно адаптировать к различным предметным областям. В связи с этим программное обеспечение системы АСТРА многоуровневое. Первый уровень — базовый — не зависит от конкретной области применения. Он включает системные блоки, набор обслуживающих программ, пакет стандартных модулей. Второй и третий уровни — проблемные — создаются в процессе адаптации системы к конкретной предметной области. Второй уровень объединяют программы, наиболее общие

для данной области. К ним относятся программные средства математического моделирования объектов этой области и программы, реализующие операции над ними. Третий уровень обеспечивает решение задач предметной области в конкретных условиях конструкторского бюро или предприятия. Этот уровень наиболее подвержен изменению.

Разработка программ проблемных уровней начинается после того, как очерчена область автоматизации и определены перспективы ее развития, поставлены задачи, подлежащие решению в рамках системы. Проблемные программы собираются из стандартных модулей базового уровня. В результате адаптации создается проблемно-ориентированная версия системы АСТРА.

Такой подход к созданию системы обеспечивает преемственность и возможность кооперации в развитии ее программного обеспечения.

Основу расчетов, выполняемых с помощью системы, составляет понятие математической модели объекта. Это понятие определяется всякий раз при адаптации системы к конкретной предметной области. При этом устанавливаются структуры данных математической модели и разрабатываются способы и программные средства ее построения. Последние ориентируются на конкретный вид исходной информации об объекте. Это могут быть данные чертежа, результаты решения задачи проектирования, экспериментальные данные и т. п. Основу моделирования объектов составляют стандартные модули моделирования кривых и поверхностей векторсплайнами, содержащиеся в базовом уровне программного обеспечения системы АСТРА. Следует отметить, что разработка средств математического моделирования объектов в процессе адаптации системы является наиболее важным и сложным этапом.

Функциональная схема системы АСТРА дана на рис. 78. Задание пользователя, представленное в виде последовательности инструкций входного языка системы, передается транслятору с помощью терминала, пульта оператора или путем ввода пакета перфокарт. Транслятор проверяет соответствие инструкций синтаксису и семантике входного языка и осуществляет интерпретацию задания с помощью системных и проблемных программ. Передача информации между последовательно выполняемыми инструкциями происходит через рабочие файлы системы и посредством базы геометрических данных. Диагностические сообщения и результат выполнения задания выдаются на печать, устройства взаимодействия с пользователем и устройства отображения графической информации.

Структурная схема системы АСТРА представлена на рис. 79. Программное обеспечение системы имеет восемь компонент. Инициализатор осуществляет подготовку рабочих таблиц системы и рабочих областей памяти, настройку и запуск диспетчера. Диспетчер системы принимает задание и для каждой инструкции в зависимости от ее типа запускает операторный или директивный транслятор или блок расширения входного языка системы. Операторный транслятор обрабатывает инструкции входного языка типа «оператор». При этом компилируется программа, реализующая арифметические

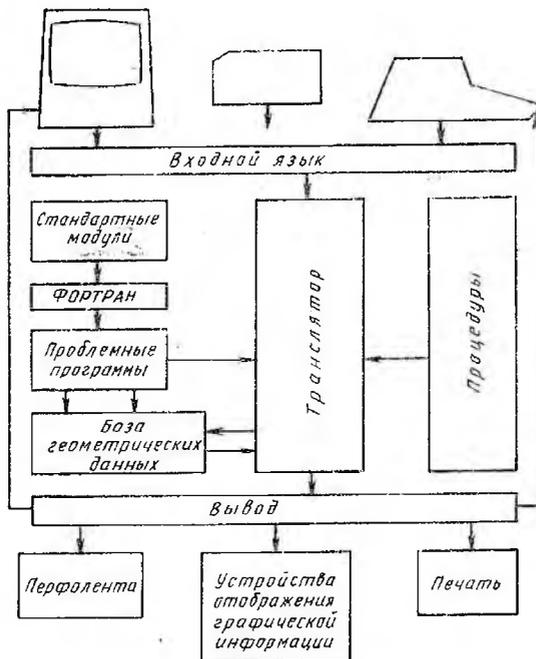


Рис. 78. Функциональная схема системы АСТРА

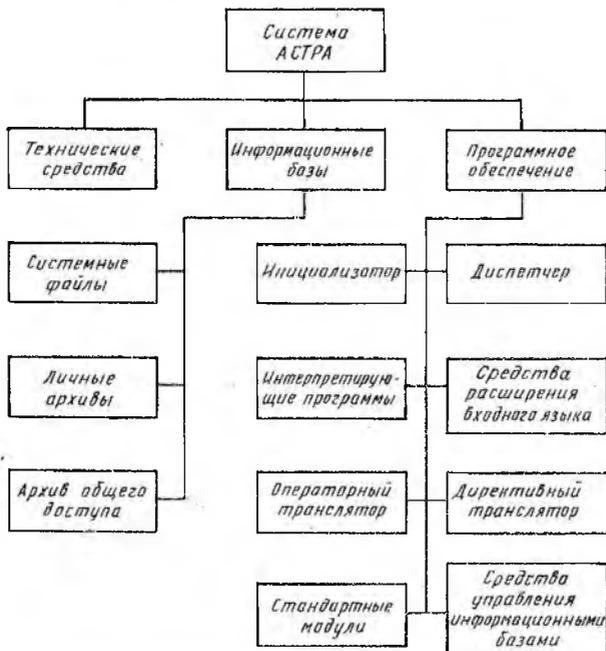


Рис. 79. Структурная схема системы АСТРА

выражения, входящие в состав оператора, и формируется обращение к соответствующей интерпретирующей программе. Директивный транслятор обрабатывает инструкции входного языка типа «директива» и выполняет действия, предписанные директивой. Блок расширения входного языка системы обрабатывает инструкции типа «предложения». При этом преобразуются соответствующие синтаксические таблицы трансляторов и строятся новые семантические программы. Интерпретирующие программы — это системные и проблемные программы, реализующие конкретные операторы входного языка системы. Проблемные интерпретирующие программы собираются из стандартных модулей средствами языка ФОРТРАН под управлением операционной системы ЭВМ. Средства управления информационными базами системы обеспечивают доступ к данным и возможность их реорганизации.

Информационные базы системы АСТРА по своему назначению делятся на три части: системные или рабочие файлы, личные архивы, архивы общего доступа. В системных файлах сохраняется текущая информация, которая создается компонентами системы в процессе прохождения одного задания. Личные архивы создаются по желанию пользователя и предназначены для долговременного хранения информации. Отличие архивов общего доступа от личных состоит только в том, что возможность записывать данные в них существенно ограничена. В архивах сохраняются различные данные (программы и процедуры, представленные на входном языке системы, геометрические данные и т. д.), обработка которых возможна программными средствами системы.

Состояние системы характеризуется набором доступных пользователю объектов, т. е. таких, которые определены во входном языке. Каждому из объектов поставлены в соответствие некоторые данные. В информационной базе эти данные представляются в виде трехмерного массива с размерностями K, L, M . При этом массиву присваивается имя объекта, указанное пользователем, номера его типа P и подтипа Q . Внутренние данные для некоторых элементарных объектов приведены в п. 2 гл. IV. Дополним этот список:

точке на кривой поставлены в соответствие имя кривой и значение t^* ее параметра;

точке на поверхности соответствуют имя поверхности и значения (t^*, s^*) ее параметров в этой точке;

внутренние данные сплайновой кривой и сплайновой поверхности — это их коэффициенты $\{V_i, V_i', t_i\}$ и $\{W_{ij}, W_{ij}^{(1,0)}, W_{ij}^{(0,1)}, W_{ij}^{(1,1)}, t_i, s_j\}$;

кривая на поверхности определяется именем поверхности, которой она принадлежит, и сплайновой кривой $t = t(\omega)$, $s = s(\omega)$, расположенной в области Ω параметров этой поверхности;

отрезок кривой задается именем кривой и значениями t_n, t_k ее параметров, соответствующими начальной и конечной точкам отрезка;

сегмент поверхности определяется именем поверхности и значениями t_n, s_n, t_k, s_k ее параметров, относящимися к двум диагонально расположенным угловым точкам сегмента.

В процессе адаптации системы к конкретной предметной области объекты последней включаются в число доступных объектов системы путем следующей последовательности действий. Для каждого объекта устанавливаются структура его данных и возможные способы его определения на входном языке. Затем для каждого из способов разрабатываются операторы входного языка и соответствующие интерпретирующие программы.

2. ВХОДНОЙ ЯЗЫК СИСТЕМЫ АСТРА¹

Входной язык обеспечивает общение пользователя с системой. Для описания синтаксических конструкций языка принято использовать металингвистические формулы Бэкуса. Поясним их следующим примером:

$$\langle \text{идер} \rangle ::= \langle \text{буква} \rangle | * \langle \text{буква} \rangle | \langle \text{идер} \rangle \langle \text{буква} \rangle$$
$$\langle \text{буква} \rangle ::= A | B | C | D | \Phi$$

Последовательность знаков, заключенная в скобки $\langle \rangle$, представляет собой металингвистическую переменную, значением которой является последовательность символов. Металингвистическая переменная нужна для того, чтобы отличать синтаксическую единицу языка от слов самого языка. Знаки $::=$ и $|$ являются металингвистическими связками, обозначающими «это есть» и «или» соответственно. Любой знак в формуле, не являющийся переменной или связкой, обозначает самого себя.

Для облегчения восприятия в качестве символов, используемых для обозначения металингвистических переменных (т. е. последовательностей знаков, стоящих внутри скобок $\langle \rangle$) выбираются слова, приблизительно описывающие природу соответствующей переменной.

В приведенном примере $\langle \text{идер} \rangle$ это может быть буква или знак $*$, к которому присоединена буква, или, если дано некоторое допустимое значение $\langle \text{идер} \rangle$, то другое значение этой переменной может быть получено присоединением к ней некоторого значения переменной $\langle \text{буква} \rangle$. Буква это есть А или В, или С, или Д, или Φ (т. е. значением металингвистической переменной $\langle \text{буква} \rangle$ может быть один из символов, стоящих справа от связки $::=$).

Основной единицей входного языка является инструкция. В зависимости от назначения различают инструкции трех типов: директивы, операторы, предложения. Директивная часть языка дает пользователю возможность осуществлять подготовку и различные преобразования текста программы (и произвольного символического текста) как с учетом синтаксической структуры, так и без него, получать информацию о состоянии процесса прохождения задачи, запускать исходную программу на трансляцию, счет и обеспечивает другие действия по управлению функционированием системы.

Операторная часть предназначена для написания программ, которые задают тот или иной вычислительный процесс. Любая

¹ Материал параграфа подготовлен с помощью В. М. Гамидова.

Операции отношения используют для изображения отношений между арифметическими выражениями.

⟨логическая операция⟩::= \neg | \wedge | \vee

Логические операции служат для образования логических выражений и имеют названия \neg (операция отрицания), \wedge (логическое умножение), \vee (логическое сложение). Таблица истинности этих операций стандартная.

⟨тип оператора⟩::= \equiv | $=$

Символ \equiv обозначает знак эквивалентности и используется в операторе эквивалентности; равенство используется в операторе присваивания.

⟨разделитель⟩::= $($ | $)$ | $-$ | $.$ | E | $,$ | $;$ | $:$ | $\#$ | $'$

Скобки () применяются в обычном математическом смысле, в них заключаются параметры функций и процедур, арифметические и логические выражения, а также конструкции типовых выражений. Символ — (пробел) служит для построения строк входной программы и удобства ее зрительного восприятия. Десятичную точку и основание степени E применяют для представления чисел. Запятая и точка с запятой используются для разделения параметров. Двоеточие служит для отделения списка целых чисел в конструкциях, а также списка параметров от спецификатора ЗНАЧЕНИЕ. Знак # может служить для обозначения конца оператора. Апостроф применяется для обозначения символьной константы.

⟨спецификатор⟩::=ЗНАЧЕНИЕ

Спецификатор используется при спецификации параметров в описании процедуры.

⟨служебное слово⟩::= \langle основное служебное слово \rangle | \langle модификатор \rangle

⟨основное служебное слово⟩::= \langle тип объекта \rangle | \langle командное слово \rangle |

⟨модальное слово⟩

⟨тип объекта⟩::=ТОЧ | ТОЧК | ТОЧП | ВЕКТОР |

ПЛОСК | КРИВ | КРИВП | ПОВ | ЦИЛ | СФЕРА | КОНУС | ОКРУЖ |

ПРЯМ | ЭЛ | МАТРИЦА | ТРАКТ | МАССИВ | ЦЕЛЫЙ | \langle новый тип \rangle

⟨командное слово⟩::=ИСК | СОХР | УД | УДБД |

ПЕЧ | ВКЛ | ЗАПИСЬ | РИСПО | \langle новое командное слово \rangle

⟨модальное слово⟩::=ФРЕЗА | \langle новое модальное слово \rangle

⟨модификатор⟩::= \langle характер типа \rangle | \langle селектор \rangle | \langle ввод—вывод \rangle

⟨характер типа⟩::=ЭКВ | ЛИН | ЛЕНТ | \langle новый характер типа \rangle

⟨селектор⟩::=КАС | ПЕР | ПАР | ХБ | ХМ | УБ | УМ | ZB | ZM |

\langle новый селектор \rangle

⟨ввод-вывод⟩::=ВК | ВЛ | МЛ | БД | \langle новый ввод-вывод \rangle

Служебные слова. Их применяют при построении операторов входного языка системы. Основные служебные слова используются при описании типовых выражений и могут стоять перед скобкой в операторах, либо при описании процедуры и задают тип ее параметра.

Характер типа определяет подтип объекта. Селектор позволяет делать выбор среди нескольких объектов. Слова группы «ввод-вывод» указывают, с какого внешнего устройства должна быть введена информация об определяемом объекте. Набор служебных слов можно

расширять, исходя из потребностей конкретной предметной области применения системы АСТРА.

(описатель): := СТАРТ | ФИНИШ | КОММТ | ПРОЦЕД | КОНЕЦП

Описатели указывают транслятору на начало и конец программы, начало и конец описания процедуры и на начало комментирующего текста.

Основные конструкции языка. Это числовые и текстовые константы, идентификаторы, функции.

Обозначая буквой m целую часть числа, буквой n — его дробную часть, буквой k — показатель степени, можно записать вещественное число в виде $m.n$ либо $m.nE \pm k$. Целая или дробная часть числа может отсутствовать. Целые константы записываются в виде последовательности десятичных цифр.

Текстовые константы имеют две формы записи: строка символов, заключенная в апострофы, или строка символов, начинающаяся выражением WH , где W — целая константа, указывающая количество символов в строке.

Идентификаторы — это последовательность букв и цифр, начинающаяся с буквы. Идентификаторы используют для обозначения переменных, стандартных и скалярных функций, процедур.

Функции определяют одно числовое значение, которое получается в результате применения заданной совокупности действий, определяемых идентификатором функции, к данной совокупности фактических параметров, заключенных в скобках.

В языке различают два вида функций: стандартные и скалярные. Фактическими параметрами стандартных функций являются арифметические выражения. Предусмотрены следующие стандартные функции: $ABS(E)$ — абсолютное значение арифметического выражения E ; $SIGN(E)$ — знак значения E ; $SIN(E)$, $COS(E)$ — синус и косинус E ; $SQRT(E)$ — корень квадратный из E , $E \geq 0$; $TAN(E)$ — тангенс E , $|E| < \pi/2$; $ARCSIN(E)$, $ARCOS(E)$, $ARCTAN(E)$ — главные значения обратных тригонометрических функций, для первых двух из них $|E| \leq 1$; $LN(E)$, $LG(E)$ — натуральный и десятичный логарифм E , $E > 0$; $EXP(E)$ — экспоненциальная функция e^E ; $ENT(E)$ — целая часть E .

К скалярным функциям относятся $ANGLE(A, B)$ — угол между векторами A и B ; $DIST(A, B)$ — расстояние между объектами A и B ; $SCAL(A, B)$ — скалярное произведение векторов A и B . Эта группа функций ставит в соответствие нескольким геометрическим объектам число.

Вычисление и присваивание значений. Входной язык имеет три вида выражений: арифметические, логические и типовые. Каждому из них соответствуют свои операции и некоторые специфические правила образования. Любое выражение служит для вычисления некоторого значения и представляет собой компактную запись, указывающую, какие величины (объекты), какого типа надо взять и какие операции над ними выполнить, чтобы получить это значение.

Значением арифметического выражения всегда является целое или вещественное число, значением логического выражения —

значение истинности. Значением типового выражения служит числовой или текстовой набор, определяющий объект более сложной природы, тип которого явно указывается в выражении. Это могут быть геометрические объекты (точка, вектор, кривая, поверхность и т. д.), технологические объекты (фреза, станок и др.), объекты типа траектории, документа и т. п.

Простейшее арифметическое выражение состоит из первичного выражения, которым может быть числовая константа, переменная, обращение к стандартной или скалярной функции, выражение, заключенное в скобки. Более сложные арифметические выражения образуются из первичных путем использования арифметических операций.

Простейшее логическое выражение состоит из логического первичного выражения, которым может быть логическая константа, логическая переменная, отношение, логическое выражение, заключенное в скобки. Более сложные выражения образуются из первичных выражений путем использования логических операций. Отношения получаются путем объединения двух арифметических выражений с помощью операций отношения.

Типовое выражение состоит из описания типа объекта. В скобках после описания типа объекта стоит конструкция, задающая способ получения данного объекта:

$\langle \text{типовое выражение} \rangle ::= \langle \text{тип объекта} \rangle \langle \langle \text{конструкция} \rangle \rangle$
 $\langle \text{конструкция} \rangle ::= \langle \text{модификатор} \rangle; \langle \text{список параметров} \rangle | \langle \text{список параметров} \rangle$
 $\langle \text{список параметров} \rangle ::= \langle \text{список ссылок} \rangle | \langle \text{числовые данные} \rangle | \langle \text{литеральные данные} \rangle | \langle \text{числовые данные} \rangle; \langle \text{список ссылок} \rangle | \langle \text{список ссылок} \rangle; \langle \text{числовые данные} \rangle | \langle \text{список ссылок} \rangle; \langle \text{список целых} \rangle | \langle \text{список параметров} \rangle; \langle \text{литеральные данные} \rangle$
 $\langle \text{числовые данные} \rangle ::= \langle \text{список значений} \rangle | \langle \text{числовые данные} \rangle; \langle \text{список значений} \rangle$
 $\langle \text{список значений} \rangle ::= \langle \text{арифметическое выражение} \rangle | \langle \text{список значений} \rangle, \langle \text{арифметическое выражение} \rangle$
 $\langle \text{список ссылок} \rangle ::= \langle \text{идентификатор} \rangle | \langle \text{список ссылок} \rangle; \langle \text{идентификатор} \rangle$

Оператор присваивания вычисляет значение объекта по выражению, заданному в правой части, а затем полученному значению присваивает идентификатор, находящийся в левой части оператора.

$\langle \text{оператор присваивания} \rangle ::= \langle \text{идентификатор} \rangle = \langle \text{выражение} \rangle | \langle \text{идентификатор} \rangle = \langle \text{командное слово} \rangle \langle \langle \text{конструкция} \rangle \rangle$

Процедуры. Под процедурой в языке понимается определенная последовательность специальных образом оформленных операторов, которая многократно используется с различными данными. Язык позволяет обращаться к процедурам (т. е. получать результат выполнения процедуры) из различных мест программы, в которой эти процедуры описаны. Сама последовательность операторов объявляется телом процедуры, изменяющиеся данные этих операторов объявляются формальными параметрами. В тексте программы такую

последовательность операторов можно заменить оператором обращения к процедуре с указанием фактических параметров.

⟨описание процедуры⟩ ::= ПРОЦЕД — ⟨заголовок процедуры⟩
⟨тело процедуры⟩ КОНЕЦП
⟨тело процедуры⟩ ::= ⟨оператор⟩ | ⟨тело процедуры⟩ ⟨оператор⟩
⟨заголовок процедуры⟩ ::= ⟨имя процедуры⟩ (⟨список формальных параметров⟩) | ⟨имя процедуры⟩ (⟨список формальных параметров⟩) ⟨спецификация⟩
⟨список формальных параметров⟩ ::= ⟨формальный параметр⟩ |
⟨список формальных параметров⟩, ⟨формальный параметр⟩
⟨формальный параметр⟩ ::= ⟨идентификатор⟩ | ⟨идентификатор⟩ =
⟨тип объекта⟩
⟨спецификация⟩ ::= ЗНАЧ: ⟨идентификатор⟩ | ⟨спецификация⟩,
⟨идентификатор⟩
⟨оператор обращения к процедуре⟩ ::= ⟨имя процедуры⟩ (⟨список фактических параметров⟩)

Между списком фактических и формальных параметров должно соблюдаться позиционное и количественное соответствие.

Рассмотрим несколько операторов входного языка системы, относящихся к ее базовому уровню. Условимся в списке параметров оператора вместо идентификаторов объектов использовать их тип, записывая его строчными буквами. Заглавными буквами будем обозначать арифметические выражения. Параметр, заключенный в квадратные скобки, при написании оператора можно опускать.

⟨имя⟩ = ТОЧ (X, Y, Z). Определяется точка с координатами (X, Y, Z). Ей присваивается имя ⟨имя⟩.

⟨имя⟩ = ТОЧ (⟨селектор⟩; крив; плоск). Среди точек пересечения кривой и плоскости в соответствии с селектором выбирается одна, которой присваивается имя ⟨имя⟩. Если селектор отсутствует, то в качестве искомой принимается первая найденная точка.

⟨имя⟩ = ТОЧК (⟨селектор⟩; крив; плоск). Отличие этого оператора от предыдущего в том, что результат имеет тип «точка на кривой». Приведем еще несколько операторов определения точки. Смысл их ясен из предыдущего:

⟨имя⟩ = ТОЧ (⟨селектор⟩; крив; сфера)
⟨имя⟩ = ТОЧК (⟨селектор⟩; крив; сфера)
⟨имя⟩ = ТОЧ (⟨селектор⟩; пов; прям)
⟨имя⟩ = ТОЧП (⟨селектор⟩; пов; прям)
⟨имя⟩ = ТОЧ (⟨селектор⟩; пов; крив)
⟨имя⟩ = ТОЧК (⟨селектор⟩; крив; пов)
⟨имя⟩ = ТОЧП (⟨селектор⟩; пов; крив)

Результат четвертого и седьмого операторов имеет тип «точка на поверхности».

⟨имя⟩ = КРИВ (массив; C; [P]). По заданным точкам, содержащимся в исходном массиве, определяется V-сплайн класса C^1 или C^2 в зависимости от параметра C. Если параметр P опущен, то это означает, что параметризация задана. В противном случае параметризация рассчитывается, причем значение параметра в пер-

вой точке устанавливается равным P . В некоторых точках исходного массива могут быть заданы касательные векторы.

$\langle \text{имя} \rangle = \text{КРИВ}$ (пов; плоск). Кривая определяется как линия пересечения поверхности с плоскостью.

$\langle \text{имя} \rangle = \text{КРИВП}$ (пов; плоск). Результат этого оператора имеет тип «кривая на поверхности».

$\langle \text{имя} \rangle = \text{КРИВ}$ (пов₁; пов₂)

$\langle \text{имя} \rangle = \text{КРИВП}$ (пов₁; пов₂)

$\langle \text{имя} \rangle = \text{КРИВП}$ (пов₂; пов₁)

Два последних оператора не являются эквивалентными. Первый из них определяет кривую на поверхности $\langle \text{пов}_1 \rangle$, а второй — на поверхности $\langle \text{пов}_2 \rangle$.

$\langle \text{имя} \rangle = \text{КРИВ}$ (крив; точк₁; точк₂). Определяется отрезок кривой $\langle \text{крив} \rangle$, расположенный между точками $\langle \text{точк}_1 \rangle$ и $\langle \text{точк}_2 \rangle$, которые определены были как точки на этой же кривой.

$\langle \text{имя} \rangle = \text{КРИВАЛ}$ (массив; C ; P). Кривая, определяемая этим оператором, получается путем локальной аппроксимации исходных точек при указанных ограничениях на отклонения приближающего V -сплайна от заданных точек. Параметры C , P описаны ранее

$\langle \text{имя} \rangle = \text{ПОВ}$ (массив; C ; $[P_1, P_2]$). По содержащимся в исходном массиве узлам некоторого каркаса определяется W -сплайн класса $C^{(2, 2)}$ или $C^{2, 2}$ в зависимости от значения C . Параметры P_1 и P_2 имеют тот же смысл, что и параметр P в операторе определения кривой, применительно к каркасным линиям 1-го и 2-го семейств соответственно. В некоторых узлах могут быть заданы касательные векторы к каркасным линиям и векторы нормали.

$\langle \text{имя} \rangle = \text{ПОВ}$ (пов; точп₁; точп₂). Задается сегмент поверхности $\langle \text{пов} \rangle$, определенный в области Ω : $u_1 \leq u \leq u_2$, $v_1 \leq v \leq v_2$, где (u_1, v_1) и (u_2, v_2) — значения параметров точек $\langle \text{точп}_1 \rangle$ и $\langle \text{точп}_2 \rangle$, которые были определены ранее как точки на той же поверхности.

$\langle \text{имя} \rangle = \text{ПОВЛ}$ ($[P]$; крив₁; крив₂). Определяется ограниченный кривыми $\langle \text{крив}_1 \rangle$ и $\langle \text{крив}_2 \rangle$ сегмент линейчатой поверхности. Параметр P задает соответствие между точками этих кривых. Если он отсутствует, то это означает, что соответствие устанавливается по нормированной длине максимальных хорд дуг. Иногда такое соответствие называют процентным. Кроме того, можно задать соответствие по длине дуги кривых или для конечного числа точек — узлов V -сплайнов, описывающих эти кривые.

$\langle \text{имя} \rangle = \text{ПОВЛ}$ (кривп). Задается линейчатая поверхность с направляющей $\langle \text{кривп} \rangle$ типа «кривая на поверхности». В качестве образующей служит нормаль к поверхности, которой принадлежит кривая $\langle \text{кривп} \rangle$.

$\langle \text{имя} \rangle = \text{ПОВВ}$ (прямя; крив). Определяется поверхность вращения с осью вращения $\langle \text{прямя} \rangle$ и образующей $\langle \text{крив} \rangle$.

$\langle \text{имя} \rangle = \text{ПОВАЛ}$ (массив; C ; $[P_1, P_2]$). Оператор определяет поверхность путем локальной аппроксимации содержащихся в исходном массиве узлов некоторого каркаса. Там же расположены значения допустимого отклонения приближающего V -сплайна от заданных узлов. Параметры C , P_1 , P_2 описаны ранее.

РИСПО (пов; плоск; матрица; M; [PR]). Этот оператор назначает расчет сечения поверхности плоскостью с последующим изображением этого сечения. Связь координат рисунка с координатами поверхности задается соответствующей матрицей. Сечение изображается в масштабе M. Одновременно может выдаваться на печать таблица координат точек сечения как в системе координат рисунка, так и в системе координат поверхности в зависимости от параметра PR. Оператор реализован в виде процедуры.

РАТОК (крив; массив), РАТОП (пов; массив). Эти два оператора определяют расстояние от точек, заданных в исходном массиве, до кривой и поверхности соответственно. Результат печатается в виде таблицы.

АНПОВ (пов; N; M). Осуществляет анализ поверхности. В точках поверхности (пов), соответствующих узлам сетки $\Delta = \Delta_u \times \Delta_v$ размера $N \times M$, рассчитываются главные кривизны и гауссова кривизна. Результат печатается в виде таблицы.

И наконец, рассмотрим несколько операторов, предназначенных для описания траектории инструмента при трехкоординатной обработке фрезерованием.

(имя)=ТРАКТ (точ). Назначает точку начального положения инструмента и присваивает идентификатор массиву, в котором будут накапливаться точки траектории и управляющие команды.

ФРЕЗА (R; r; AL; H_x , H_y , H_z). Назначает параметры фрезы. Т (точ, F), Т (X, Y, Z, F). Описывают очередную точку траектории и подачу, с которой будет выполняться перемещение в эту точку.

Т (PR; вект; F).

Очередная точка траектории задается приращением координат либо от предыдущей (PR=0), либо от последующей (PR \neq 0) точки траектории.

КОНТУР ([селектор]; крив; L; DEL; EPS; F). Назначает обработку по контуру, который задается кривой (крив). Ось фрезы перемещается слева (селектор=ЛЕ), справа (селектор=ПР) и по кривой (селектор=НА). Торце фрезы смещен на величину L, DEL — припуск на последующую обработку, EPS — точность воспроизведения контура.

ОБПОВ ([селект.]; пов; L; P; n; DEL; EPS; H; F). Назначает обработку поверхности (пов). Схема обработки задается признаком: селектор=Z1|Z2|C1|C2, где Z — обозначает схему типа «зигзаг»; C — строчную обработку в одном направлении; цифра, стоящая рядом с этими символами, указывает номер семейства координатных линий, вдоль которых происходит движение фрезы. Параметры L и P задают длину пути подхода (выхода) инструмента в конце (начале) строки (для схемы «зигзаг» — по касательной, для обработки строками — по направлению оси шпинделя q). Параметр n = ± 1 задает обрабатываемую сторону поверхности. DEL, EPS, H, F — значения припуска, точности воспроизведения, допустимой шероховатости и рабочей подачи.

НЗЗ (тракт). Назначает вывод управляющей перфоленты для устройства числового программного управления типа Н-33.

3. АВТОМАТИЗАЦИЯ ПРОЕКТИРОВАНИЯ И ПОДГОТОВКИ ПРОИЗВОДСТВА ЛОПАСТЕЙ ГИДРОТУРБИН

Область применения системы АСТРА в данной проблеме включает: геометрическое обеспечение анализа проектных решений; расчет данных, необходимых для оформления конструкторской документации и технологической подготовки производства лопастей гидротурбин; изготовление натуральных и модельных колес на станках с ЧПУ.

Объектами рассматриваемой предметной области являются элементы проточной части гидротурбины: лопасти рабочих колес радиально-осевой, поворотной-лопастной и ковшевой гидротурбины, лопатки направляющего аппарата; поверхности тока; элементы технологической оснастки: штампы, шаблоны.

1. Математическая модель лопасти радиально-осевой и поворотной-лопастной гидротурбин есть совокупность четырех поверхностей: поверхности всасывающей стороны *I* лопасти, поверхности стороны нагнетания *II*, входной поверхности *III* и поверхности всей лопасти в целом, представленных в виде *W*-сплайнов.

Исходная информация о поверхности лопасти радиально-осевой гидротурбины может быть двух видов: а) узлы r_{ij} , $i = 1, \dots, N$; $j = 1, \dots, M$ некоторого каркаса поверхности лопасти, полученные, например, в результате проектных расчетов; б) множество точек плоскопараллельных сечений поверхности лопасти. Этим видам соответствуют два различных способа построения математической модели лопасти и две формы оператора входного языка системы:

$\langle \text{имя} \rangle = \text{ММЛ}$ (массив; R),

$\langle \text{имя} \rangle = \text{ММЛРО}$ (массив; R, N, M).

В результате выполнения этих операторов получают четыре поверхности, указанные выше, которым без участия пользователя присваиваются имена $\langle \text{имя}_1 \rangle$, $\langle \text{имя}_2 \rangle$, $\langle \text{имя}_3 \rangle$, $\langle \text{имя}_4 \rangle$, и список этих имен, идентифицируемый именем $\langle \text{имя} \rangle$. Массивы исходных данных должны быть сформированы по определенному правилу. Величина R задает средний размер «по потоку» входной поверхности. Параметры N и M задают размер сетки *W*-сплайна, описывающего поверхность лопасти в целом.

Программа, реализующая оператор *ММЛ*, выполняет следующие действия. Сначала определяется *W*-сплайн, восстанавливающий

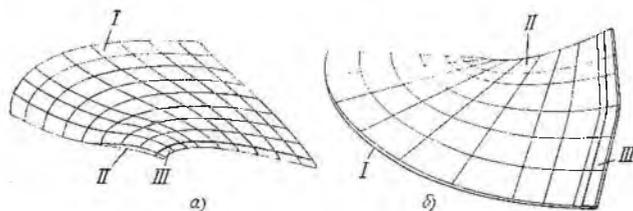


Рис. 80. Лопасти гидротурбин:
а — радиально-осевой; б — поворотной-лопастной

поверхность лопасти по значениям $\{r_{ij}\}$ (см. решение задачи 6 из п. 6 гл. II). Затем по значению R выделяется входная поверхность III и формируются поверхности I и II.

Алгоритм, соответствующий оператору ММЛРО, коротко состоит в следующем. Определяются W -сплайны, восстанавливающие поверхности I и II (см. решение задачи 7, п. 6, гл. II). При этом в качестве поверхностей $G_j(s) = G_j(t_j, s)$, $j = 1, \dots, M$, используются поверхности вращения с образующими

$$r_j(s) = r_b(s)(1 - t_j) + r_n(s)t_j; \quad s \in [0, 1],$$

где $0 = t_1 < t_2 < \dots < t_M = 1$, $r_b(s)$ и $r_n(s)$ — образующие поверхностей верхнего и нижнего обода колеса. Затем по значению R на поверхностях I и II определяется зона входной поверхности. Входная поверхность строится как поверхность, гладко сопрягающаяся с поверхностями I и II и проходящая через линию входной кромки. И наконец, полученные поверхности объединяются в общую поверхность лопасти.

Поверхность лопасти поворотной-лопастной гидротурбины может быть задана либо узлами некоторого ее каркаса, либо набором разверток цилиндрических сечений, представленных в дискретно-точечной форме. В первом случае построение математической модели этой поверхности осуществляется оператором ММЛ. Второму случаю соответствует оператор вида

$$\langle \text{имя} \rangle = \text{ММЛПЛ} (\text{массив}; R, N).$$

В алгоритме, реализующем этот оператор, восстанавливающий W -сплайн находится так же, как и при решении задачи 6 п. 6 гл. II, т. е. цилиндрические сечения лопасти принимаются в качестве одного из семейств каркасных линий. Исходные данные и поверхность, полученную в результате, можно подвергнуть сглаживанию, используя операторы КРИВАЛ и ПОВАЛ. Для оценки полученных математических моделей можно воспользоваться операторами РИСПО и АНПОВ.

2. Геометрическое обеспечение гидродинамических расчетов состоит в подготовке информации, необходимой для решения задач обтекания лопасти, и основывается на ее математической модели. Рассмотрим несколько предназначенных для этой цели операторов входного языка системы, не приводя при этом их формальную запись.

Оператор РЕЖИМ формирует массив данных определенной структуры, который в дальнейшем используется для расчета гидродинамических характеристик лопасти. Массив содержит следующие параметры: диаметр колеса, диаметр потока, расход и число лопастей в колесе; приведенные расход и частоту вращения для нескольких рабочих режимов; значения гидравлического КПД и закрутки потока перед рабочим колесом для множества поверхностей тока и для всех рабочих режимов.

Оператор РАСПР для заданной поверхности тока обеспечивает нахождение линии ее пересечения с поверхностью лопасти и отображение полученного профиля в расчетную плоскость. Эта геометрическая информация совместно с соответствующими данными

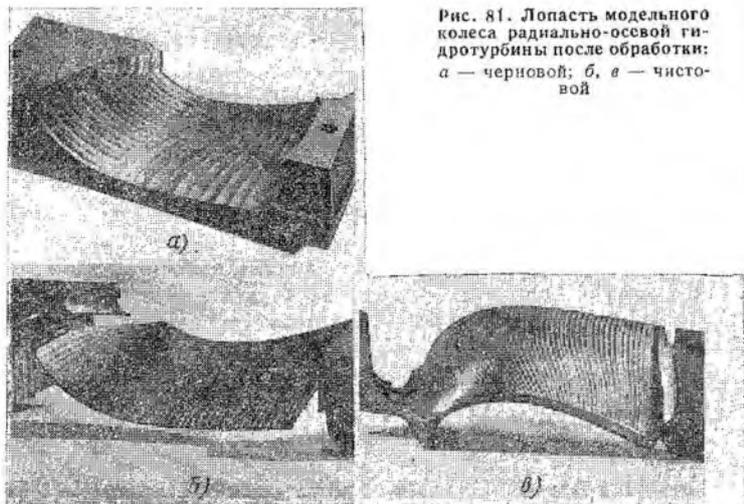


Рис. 81. Лопасть модельного колеса радиально-осевой гидротурбины после обработки: а — черновой; б, в — чистой

для конкретного рабочего режима является исходной для оператора РОБ, осуществляющего решение задачи обтекания решетки профилей. Кроме того, эти данные с помощью оператора ИПРО можно представить в виде изображения профиля на графопостроителе или таблиц, содержащих координаты его точек и точек образующей поверхности тока.

Операторы РСР и РИПОЗ позволяют построить срединную поверхность лопасти и в узлах некоторого ее каркаса рассчитать векторы нормали, толщины лопасти по окружности, коэффициенты стеснения. Эти данные являются исходными для решения прямой осесимметричной задачи обтекания лопасти.

Оператор РФИ для заданных значений R_i , z_i , $i = 1, \dots, N$, цилиндрических координат, естественным образом связанных с системой координат колеса, рассчитывает значения угловой координаты φ_i и векторы нормали \mathbf{n}_i .

3. Расчет данных для оформления рабочего чертежа лопасти обеспечивается проблемными операторами СГЛ, СПЛС, РАСВЕТ, ПОЛО и оператором общего назначения РИСПО.

С помощью оператора РИСПО можно рассчитать произвольные плоские сечения лопасти. В узлах двумерной сетки, заданной в любой из координатных плоскостей с помощью операторов СГЛ, СПЛС находятся значения третьей координаты точек лопасти. В этих точках определяется толщина лопасти как расстояние между точками поверхности нагнетания и поверхностью всасывания. Распределение толщины лопасти по поверхности любой из ее сторон можно получить также с помощью оператора РТЛ.

Оператор РАСВЕТ позволяет определить «расстояние в свету» — расстояние между точками выходной кромки лопасти и поверхностью соседней лопасти. С помощью оператора ПОЛО рассчитывается

положение лопасти при изготовлении ее методом штампования и выдаются необходимые данные для изготовления штампа.

4. Подготовка управляющих перфолент для изготовления лопастей модельных колес на станке с ЧПУ осуществляется с помощью операторов ПОЗА, ТРАКТ, ТСЕТР, НЗЗ. По оператору ПОЗА в соответствии с технологическими требованиями на положение лопасти при ее обработке осуществляется преобразование координат колеса в координаты станка и определяются оптимальные размеры заготовки (размеры минимального параллелепипеда, содержащего лопасть). Затем с помощью операторов ТРАКТ, ТСЕТР, НЗЗ подготавливаются перфоленты с управляющей информацией для всех технологических переходов обработки лопасти: черновой и чистовой обработки поверхностей всасывания и нагнетания, обработки торцовых поверхностей, по которым лопасть соприкасается с поверхностью верхнего и нижнего ободов колеса. На рис. 81 представлены различные этапы обработки лопасти модельного колеса радиально-осевой гидротурбины.

4. АВТОМАТИЗАЦИЯ РАСЧЕТА УПРАВЛЯЮЩИХ ПРОГРАММ ДЛЯ ИЗГОТОВЛЕНИЯ РАБОЧИХ КОЛЕС ЦИРКУЛЯЦИОННЫХ НАСОСОВ НА СТАНКАХ С ЧПУ

Пусть требуется изготовить рабочее колесо циркуляционного насоса (рис. 82) на пятикоординатном станке с ЧПУ. Заготовка представляет собой отливку с некоторым припуском, в общем случае неравномерным по поверхности колеса. Обработке на станке подлежит вся поверхность колеса, за исключением торцовых поверхностей ступицы и лопастей.

Адаптация системы к этой проблеме обусловила необходимость разработки и программной реализации метода построения математической модели колеса, а также программных модулей расчета траектории инструмента при многокоординатной обработке.

Методика расчета управляющих программ для изготовления детали указанного класса, реализованная в рамках системы АСТРА, предполагает три последовательных этапа: построение математической модели колеса, расчет областей обработки, получение управляющих программ для любого технологического перехода.

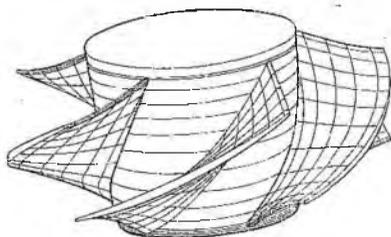


Рис. 82. Рабочее колесо циркуляционного насоса

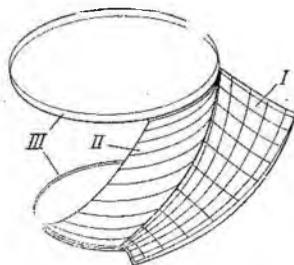


Рис. 83. Области обработки поверхности рабочего колеса циркуляционного насоса

1. Математическая модель колеса состоит из моделей поверхности лопасти и поверхности ступицы, представляющей собой поверхность вращения. Исходными данными для построения математической модели колеса приняты теоретический профиль и набор параметров, задающих закон преобразования его в цилиндрические сечения лопасти, параметры образующей поверхности ступицы. Соответствующий языковый оператор имеет вид

{имя} = ММК(массив исходных данных)

Для анализа полученной математической модели колеса предусмотрен оператор РСЦ, позволяющий рассчитать цилиндрические сечения лопасти и их развертки.

2. Расчет областей обработки проводится с учетом того, что колесо обладает осевой симметрией L -го порядка (L — число лопастей). Поэтому достаточно рассчитать управляющие программы для обработки следующих областей его поверхности (рис. 83): поверхности лопасти I , части поверхности ступицы II , заключенной между соседними лопастями колеса, частей поверхности ступицы III , примыкающих к входным и выходным кромкам лопасти.

Расчет этих областей — областей обработки — назначается оператором ПО. При этом каждая из них параметризуется так, что координатные линии совпадают со строками обработки. Это существенно упрощает расчет траектории инструмента. Одновременно для каждой области обработки рассчитывается функция допустимого угла наклона β оси q фрезы к касательной плоскости обрабатываемой поверхности.

3. Расчет траектории фрезы и получение управляющих перфолент для технологических переходов: подрезка кромок лопасти, черновая и чистовая обработка поверхности лопасти, обработка поверхности ступицы, обработка частей II и III поверхности ступицы назначается операторами ПК, ПОПЛ, ООПЛ, ОПС и ОСТ соответственно. При этом можно варьировать точность воспроизведения элементов детали, шероховатость получаемой поверхности, параметры инструмента, рабочую подачу. Управляющие перфоленты могут быть сегментированы как по заданному максимальному времени обработки, так и по заданному максимальному числу кадров. Для обработки используется фреза с шаровым концом $R = r$. Направление оси инструмента в точке $\mathbf{r}(u, v)$ поверхности определяется зависимостью $\mathbf{q}(u, v) = \langle \boldsymbol{\tau} \times \mathbf{n} \rangle \cos \beta(u, v) + \mathbf{n} \sin \beta(u, v)$, где $\boldsymbol{\tau} = \boldsymbol{\tau}(u, v)$ — касательный вектор к поверхности в направлении строки, \mathbf{n} — вектор нормали к поверхности, $\beta(u, v)$ — угол наклона оси q к касательной плоскости.

5. АВТОМАТИЗАЦИЯ РАСЧЕТНО-КОНСТРУКТОРСКИХ РАБОТ ПРИ ТЕХНОЛОГИЧЕСКОЙ ПОДГОТОВКЕ ПРОИЗВОДСТВА В АВИАСТРОЕНИИ

Назначение системы АСТРА в этой предметной области — геометрическое обеспечение инженерных работ при освоении изделия. Вследствие обширности рассматриваемой области не представляется

возможным на страницах этой книги дать исчерпывающее представление о ней самой и задачах, решаемых с помощью системы. Ограничимся только краткой характеристикой последних.

Подготовка производства основывается на геометрических данных об изделии и, следовательно, так или иначе связана с решением многочисленных геометрических задач. Одной из основных особенностей при этом является большой объем вычислений. В условиях сжатых сроков, отводимых на подготовку производства, этот фактор играет существенную роль, и поэтому на первый план выступает оперативность выполнения расчетов. В системе АСТРА оперативность обеспечивается наличием входного языка высокого уровня, архивов для долговременного хранения данных, а также программных средств управления ими. Высокий уровень автоматизации и оперативности проведения расчетов поддерживается путем расширения входного языка системы за счет разработки новых языковых операторов и типовых процедур.

Объектами данной предметной области являются агрегаты самолета (фюзеляж, крыло, элементы хвостового оперения и механизации крыла, воздухозаборники и т. п.), их узлы и детали, конструктивные и технологические базы, технологическая оснастка.

Математическая модель агрегата. Построение математической модели агрегата является необходимым и основополагающим этапом автоматизированной подготовки производства. Под математической моделью агрегата понимается совокупность его поверхностей, представленных в виде W -сплайнов. Поверхности агрегатов можно отнести к одному из следующих видов: линейчатые (в том числе цилиндрические общего вида), вращения и поверхности общего типа. Из числа последних можно указать некоторые специфические поверхности, характерные только для авиастроения; «нелинейчатое крыло», «зализ», «обтекатель», «аэродинамический гребень» и т. п. Для представления поверхностей в теоретических чертежах обычно используют два способа: кинематический и каркасный (в соответствии с принятой в авиастроении терминологией). Кинематический способ формирования поверхности состоит в определении однопараметрического семейства кривых, образующих требуемую поверхность. При каркасном способе задания поверхности указывается конечное число линий, лежащих на ней. При этом некоторые из них (или каждая) могут быть представлены в дискретно-точечной форме.

Построение математической модели агрегата основано на методах и приемах моделирования кривых и поверхностей V -сплайнами и W -сплайнами. Так, кинематически заданные поверхности интерполируются с требуемой точностью W -сплайном. Поверхности, представленные в дискретно-точечной форме, восстанавливаются интерполяционным или сглаживающим W -сплайном. Сложные по форме поверхности разбиваются на части — отсеки, которые моделируются отдельно от соседних, но с учетом требуемой гладкости их стыковки.

Рассмотрим поверхность типа «зализ». Иногда она задается как огибающая поверхности сферы, перемещающейся в пространстве

и касающейся двух поверхностей. Языковой оператор описания такой поверхности имеет вид
 (имя)=ПОВ (зализ; пов₁; пов₂; R)

Здесь R — радиус сферы. Алгоритм, реализующий этот оператор, коротко состоит в следующем. Ищется линия пересечения поверхностей, R-эквидистантных (отстоящих на величину R по нормали) поверхностям пов₁ и пов₂, которая является геометрическим местом центров сфер радиуса R, касающихся одновременно пов₁ и пов₂. Точки касания образуют граничные линии поверхности зализа. Окончательно поверхность, огибающая семейство сфер, приближается W-сплайном с заданной точностью.

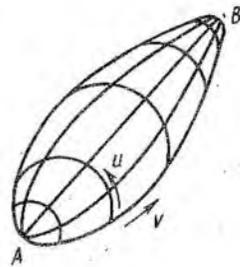


Рис. 84. Поверхность обтекателя

Поверхность типа «обтекатель» или «концевой обтекатель» характеризуется наличием двух особых точек A и B (рис. 84). В эти точки стягиваются крайние координатные линии $v = v_1$, и $v = v_M$. При этом каркасные кривые, соответствующие точкам A и B, задаются значениями:

$$r_{i1} = r_A, \quad \tau_{i1}(u) = \tau_{i2}(u), \quad n_{i1} = n_A, \quad r_{iM} = r_B, \quad \tau_{iM}(u) = \tau_{i, M-1}(u), \\ n_{iM} = n_B, \quad i = 1, \dots, N,$$

где n_A , n_B — орты нормали к поверхности обтекателя в точках A и B.

Полученная математическая модель агрегата после надлежащего анализа отправляется в архив общего пользования и становится основным источником геометрических данных агрегата.

Расчет конструктивных и технологических баз. Конструктивные и технологические базы используются для задания взаимного расположения агрегатов, узлов и деталей. В качестве этих баз используются теоретические поверхности агрегатов, плоскости силового набора, линии стрингеров, кромки обшивок, центры сборочных отверстий, базовые отверстия и т. д. Геометрически они представляют собой точки, прямые, плоскости, кривые и поверхности. Определяются конструктивные и технологические базы по данным чертежей путем решения соответствующих геометрических задач. Например, линии стрингеров, шпангоутов и нервюр — это линии на поверхности, и рассчитываются они как линии пересечения внутренней поверхности обшивки агрегата с плоскостями. В данном случае используется оператор входного языка КРИВП. Конструктивные и технологические базы сохраняются в архиве общего доступа.

Расчет геометрии деталей и технологической оснастки. Характерной особенностью деталей планера самолета является то, что они не имеют полностью заданных в чертеже размеров. Их конфигурация определяется внешним обводом самолета. Кроме того, для них характерны различного вида подсежки, обрезы, отверстия и вырезы, обусловленные спецификой конструкций.

С помощью системы АСТРА определяются все необходимые геометрические параметры деталей. При этом решаются самые разно-

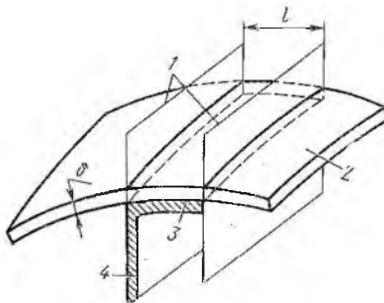


Рис. 85. Фрагмент поверхности борта детали:
 1 — ограничивающие плоскости; 2 — поверхность агрегата; 3 — борт детали; 4 — стенка детали

образные задачи плоской и пространственной геометрии, в том числе осуществляются развертки, квази-развертки и обратные им преобразования — свертки поверхности. Зазоры между элементами соседних деталей подсчитываются как минимальное расстояние между множеством их точек. Местоположение

вырезов под элементы соседних узлов и их форма определяются в общем случае как пересечение геометрических объектов: плоскостей, плоскости и цилиндрической поверхности, двух поверхностей.

Основные элементы деталей каркаса планера самолета по их функциональному назначению можно разделить на две группы: элементы типа «стенка» и элементы типа «борт» («полка», «ребро»). Первые из них определяют деталь в целом и представляют собой плоскость или поверхность. Вторые являются элементами стыковых деталей с обшивкой или с соседними узлами и описываются линейчатой поверхностью или поверхностью общего вида. Чтобы рассчитать поверхность борта детали, можно воспользоваться следующим оператором входного языка системы;

$\langle \text{имя} \rangle = \text{БОРТ} (\text{пов}; \text{плоск}; \delta; l; \varepsilon)$

Этот оператор определяет сегмент, δ -эквилидистантный поверхности $\langle \text{пов} \rangle$, ограниченный плоскостью $\langle \text{плоск} \rangle$ и плоскостью, отстоящей от первой на расстояние l , равное ширине борта детали (рис. 85). Если выделенный сегмент описывается линейчатой поверхностью с точностью ε , она принимается в качестве искомой. В противном случае результат будет представлен в виде W -сплайна.

В процессе подготовки производства проектируется и изготавливается обширная номенклатура технологической оснастки, включающая контрольно-измерительную (эталоны, шаблоны), изготовительную (болванки, обтяжные пуансоны и т. п.) и сборочную (стапели, макеты и т. п.) оснастку. Для определения геометрических параметров элементов оснастки решаются те же задачи, что и при расчете геометрии деталей, и теми же программными средствами. Например, контур шаблона рассчитывается с использованием оператора КРИВ как линия пересечения поверхности агрегата и плоскости шаблона. Рабочая поверхность формообразующей оснастки, например болванки или обтяжного пуансона, представляет собой сегмент поверхности агрегата. В системе АСТРА предусмотрены несколько способов описания таких элементов. Один из них, когда выделяется кусок поверхности, ограниченный четырьмя координатными линиями $u = u_i$ и $v = v_i$, $i = 1, 2$, был приведен при изложении входного языка. Укажем еще один способ, которому соответствует следующий оператор;

$\langle \text{имя} \rangle = \text{СЕГМ} (\text{пов}_1; \text{пов}_2; \delta)$

Этот оператор описывает сегмент, δ -эквидистантный поверхности $\langle \text{пов}_1 \rangle$, ограниченный замкнутой кривой — линией пересечения поверхности $\langle \text{пов}_1 \rangle$ с линейчатой поверхностью $\langle \text{пов}_2 \rangle$. Предполагается, что направляющая поверхности $\langle \text{пов}_2 \rangle$ есть замкнутая кривая. Алгоритм, реализующий этот оператор, состоит в следующем. Ищется пересечение исходных поверхностей $r = W_1(u, v)$ и $r = W_2(t, s)$ в виде линии на поверхности $\langle \text{пов}_1 \rangle$. В плоскости параметров (u, v) ей соответствует замкнутая кривая с уравнением $u = u(T)$, $v = v(T)$, $T \in [0, 1]$. Любое взаимно-однозначное отображение области G , ограниченной этой кривой, на прямоугольник Ω' в плоскости переменных T, S совместно с выражением для W -сплайна $W_j(u, v)$ описывает выделенный сегмент. В рассматриваемом алгоритме для определения этого отображения использован прием, пригодный только для звездообразной области.

Источниками геометрической информации при проведении этих расчетов служат математические модели агрегатов, конструктивные и технологические базы.

Расчет траектории фрезы при изготовлении деталей и оснастки на станках с ЧПУ. Ввиду геометрической и технологической сложности обводообразующей оснастки и самих деталей, выходящих на внешние обводы изделия, а также вследствие высокой точности, предъявляемой к их изготовлению, применение станков с ЧПУ здесь является наиболее эффективным. С помощью системы АСТРА рассчитываются управляющие программы обработки различных элементов поверхностей деталей и оснастки для трех- и многокоординатных станков. Используется обработка по контуру, обработка открытых поверхностей, выборка «колодцев» на поверхностях. Для обработки типовых элементов деталей и оснастки разработан ряд процедур на входном языке системы, что обеспечивает высокую оперативность в получении управляющих программ. Существенным фактором, повышающим оперативность расчета, является то, что основной объем необходимой геометрической информации находится в базе геометрических данных. Это избавляет пользователя от необходимости всякий раз готовить указанную информацию на внешних носителях для ввода в ЭВМ.

СПИСОК ЛИТЕРАТУРЫ

1. Алберг Дж., Нильсон Э., Уолш Дж. Теория сплайнов и ее приложения. М.: Мир, 1972. 318 с.
2. Аппроксимация сплайнами контуров каркаса обувной колодки/Е. М. Добрышман, В. А. Фукин, В. В. Коростылева, А. Ю. Рогожин — Известия вузов. Технология легкой промышленности, 1981, № 2, с. 73—75.
3. Бахвалов Н. С. Численные методы. Т. 1. М.: Наука, 1973. 632 с.
4. Василенко В. А. Сплайны. Теория, алгоритмы, программы. Новосибирск: Наука СО, 1983. 216 с.
5. Волков А. М. Решение практических задач геологии на ЭВМ. М.: Недра, 1980. 224 с.
6. Горелик А. Г. Автоматизация инженерно-графических работ с помощью ЭВМ. Минск: Высшая школа, 1980. 208 с.
7. Гроссман К., Каплан А. А. Нелинейное программирование на основе безусловной минимизации. Новосибирск: Наука СО, 1981. 184 с.
8. Завьялов Ю. С., Квасов Б. И., Мирошниченко В. Л. Методы сплайн-функций. М.: Наука, 1980. 352 с.
9. Исаев В. К. Принцип максимума и П-сплайны. — Докл. АН СССР, 1981, т. 256, № 1, с. 22—25.
10. Кашин Г. М., Пшеничных Г. И., Флеров Ю. А. Методы автоматизированного проектирования самолета. М.: Машиностроение, 1979. 168 с.
11. Кристофидес Н. Теория графов, алгоритмический подход. М.: Мир, 1978. 432 с.
12. Кужутов А. В., Леус В. А., Полещук Н. С. Численный синтез тонового изображения освещенной поверхности и растровая визуализация. — Автометрия, 1978, № 5, с. 59—64.
13. Методы сплайн-функций (Вычислительные системы, 68)/Под ред. Ю. С. Завьялова. Новосибирск: Ин-т математики СО АН СССР, 1976. 120 с.
14. Мишин А. И. Параллельные вычислительные среды с локальными взаимодействиями элементов. — Автоматика и телемеханика, 1982, № 10, с. 147—155.
15. Новгородцев В. А., Мокеева Н. С. Автоматизация изготовления программно-носителей для вышивальных полуавтоматов. М.: Легкая и пищевая промышленность, 1982. 120 с.
16. Ньюмен У., Спруэл Р. Основы интерактивной машинной графики. М.: Мир, 1976. 572 с.
17. Основы технологии судостроения/В. Д. Мацкевич, Э. В. Ганов, В. П. Доброленский и др.; Под ред. В. Д. Мацкевича. Л.: Судостроение, 1980. 352 с.
18. Погорелов А. В. Геометрия. М.: Наука, 1983. 288 с.
19. Применение рекуррентных сплайн-функций для обработки речевых и видеосигналов/А. В. Серединский, В. А. Ванде-Кирков, И. А. Буздалина и др. — Электросвязь, 1982, № 2, с. 60—64.
20. Сплайн-функции в инженерной геометрии. (Вычислительные системы, 86)/Под ред. Ю. С. Завьялова, В. А. Леуса. Новосибирск: Ин-т математики СО АН СССР, 1981. 160 с.
21. Стечкин С. Б., Субботин Ю. Н. Сплайны в вычислительной математике. М.: Наука, 1976. 248 с.

22. **Технология самолетостроения**/А. Л. Абибов, Н. М. Бирюков, В. В. Бойцов и др.; Под ред. А. Л. Абибова. М.: Машиностроение, 1982. 552 с.
23. **Уилкинсон Дж., Райниш К.** Справочник алгоритмов на языке АЛГОЛ. Линейная алгебра. М.: Машиностроение, 1976. 390 с.
24. **Фиакко А., Мак-Кормик Г.** Нелинейное программирование. Методы последовательной безусловной минимизации. М.: Мир, 1972. 240 с.
25. **Фихтенгольц Г. М.** Курс дифференциального и интегрального исчисления. Т. 3. М.: Наука, 1966. 656 с.
26. **Фокс А., Пратт М.** Вычислительная геометрия. Применение в проектировании и на производстве. М.: Мир, 1982. 304 с.
27. **Advances in CAD/CAM**/Ed. T. M. R. Ellis, O. I. Semenkov. Amsterdam—New York—Oxford: North—Holland Publishing Company, 1983. 720 p.
28. **Advances in computer-aided manufacture**/Ed. D. McPher — son. Amsterdam—New York—Oxford: North—Holland Publishing Company, 1977. 467 p.
29. **Boor de C.** A practical guide to splines. New York—Heidelberg—Berlin: Springer Verlag, 1978. 392 p.
30. **Computer aided geometric design**/Ed. R. Barnhill, R. Riesenfeld. New York—San Francisco—London: Academic Press, 1974. 326 p.
31. **Computer languages for numerical control**/Ed. J. Hatvany. Amsterdam: North—Holland Publishing Company, 1973.
32. **Coons S. A.** Surfaces for computer aided design of space forms. Report MAC-TR-41, Project MAC, M. I. T., 1967. 105 p.
33. **Hollywood signs on biggest computer ever.** — Electronics, vol. 55, N. 16, August 1982, p. 56—57.
34. **Roman G. C., Kimura T.** VLSI perspective of real time hidden-surface elimination. — Computer aided-design, vol. 13, N. 13, 1981, p. 99—107.

ОГЛАВЛЕНИЕ

Предисловие	3
Введение	5
Глава I. Элементы теории сплайн-функций	10
1. Аппроксимация функций многочленами	10
2. Аппроксимация многочленами первой, второй и третьей степени	14
3. Аппроксимация функций сплайнами одной и двух переменных	20
4. Сплайны первой степени	24
5. Эрмитовы кубические сплайны	25
6. Кубические сплайны	30
7. Задача сглаживания кубическими сплайнами	40
8. <i>B</i> -сплайны	46
9. Локальная аппроксимация кубическими сплайнами	52
10. Метод прогонки	60
Глава II. Математическое моделирование кривых и поверхностей	63
1. Кривые и поверхности в инженерно-геометрических задачах	63
2. Понятие близости геометрических объектов	68
3. Некоторые сведения из дифференциальной геометрии	76
4. Вектор-сплайны одной и двух переменных	80
5. Аппроксимация кривых <i>V</i> -сплайнами	85
6. Аппроксимация поверхностей <i>W</i> -сплайнами	94
Глава III. Моделирование кривых и поверхностей нелинейными сплайнами	102
1. Окружностная интерполяция плоской кривой	102
2. Окружностная интерполяция пространственной кривой	106
3. Каркасная аппроксимация поверхности	116
4. Решение задачи о связном сечении	124
Глава IV. Методы численного решения задач инженерной геометрии	133
1. Расчет характеристик кривой и поверхности	133
2. Кратчайшее расстояние между элементарными геометрическими объектами	137
3. Пересечение элементарных геометрических объектов	142
4. Развертки и квазиразвертки поверхностей	143
Глава V. Математическое обеспечение металлорежущих станков с числовым программным управлением	151
1. Расчет траектории инструмента при обработке сложных поверхностей	151
2. Оптимизация траектории фрезерования сложной области	157
3. Постановка некоторых оптимизационных задач для процессов программно-управляемой обработки металлов резанием	164

Глава VI. Численная визуализация пространственных сцен	172
1. Изображение простейших пространственных объектов	172
2. Изображение непрозрачных поверхностей	175
3. Параллельный процесс численного синтеза тоновых изображений	180
4. Зависимость времени параллельного синтеза изображения от тила межпроцессорного взаимодействия	187
Глава VII. Автоматизация обработки геометрических данных	197
1. Общая характеристика системы АСТРА	197
2. Входной язык системы АСТРА	201
3. Автоматизация проектирования и подготовки производства лопастей гидротурбин	209
4. Автоматизация расчета управляющих программ для изготовления рабочих колес циркуляционных насосов на станках с ЧПУ	212
5. Автоматизация расчетно-конструкторских работ при технологической подготовке производства в авиастроении	213
Список литературы	218

*Юрий Семенович Завьялов, Владимир Александрович Леус,
Владимир Анатольевич Скороспелов*

СПЛАЙНЫ В ИНЖЕНЕРНОЙ ГЕОМЕТРИИ

Редактор *Т. В. Абизова*

Художественный редактор *С. С. Водчиц*

Переплет художника *Г. Г. Кожанова*

Технический редактор *А. И. Захарова*

Корректоры *О. Е. Мишина и Н. Г. Богомолова*

ИБ № 3781

Сдано в набор 21.06.84. Подписано в печать 28.12.84. Т-24431
Формат 60X90¹/₁₆. Бумага типографская № 2.
Гарнитура литературная. Печать высокая. Усл. печ. л. 14,0.
Усл. кр.-отт. 14,0 Уч.-изд. л. 15,72. Тираж 7900 экз. Заказ № 170
Цена 1 р. 20 к.

Ордена Трудового Красного Знамени издательство «Машиностроение»,
107076, Москва, Стромьинский пер., 4

Ленинградская типография № 6 ордена Трудового Красного Знамени
Ленинградского объединения «Техническая книга» им. Евгении Соколовой
Союзполиграфпрома при Государственном комитете СССР
по делам издательств, полиграфии и книжной торговли,
193144, г. Ленинград, ул. Моңсеенко, 10.

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

С целью получения информации о качестве наших изданий, просим Вас в прилагаемой анкете подчеркнуть позиции, соответствующие Вашей оценке книги
«Слайны в инженерной геометрии»
автор (ы). Ю. С. Завьялов, В. А. Леус, В. А. Ско-
ропелов

1. Необходимость издания:

- значительная
- незначительная

2. Эффективность книги с точки зрения прак-
тического вклада в отрасль:

- высокая
- незначительная

3. Эффективность книги с точки зрения тео-
ретического вклада в отрасль:

- высокая
- незначительная

4. Материал книги соответствует достижениям
науки и техники в данной отрасли:

- в полной мере
- частично
- слабо

5. Книга сохранит свою актуальность:

- 1—2 года
- в течение 5 лет
- длительное время

6. Название книги отвечает содержанию:

- в полной мере
- частично