

В.А. Вознесенский
Т.В. Ляшенко
Б.Л. Огарков

ЧИСЛЕННЫЕ МЕТОДЫ

решения
строительно-
технологических
задач на ЭВМ

Под редакцией профессора
В. А. Вознесенского

Допущено Министерством
высшего и среднего
специального образования УССР
в качестве учебника
для студентов вузов,
обучающихся по специальности
«Производство строительных
изделий и конструкций»

Киев
Головное издательство
издательского объединения
«Выща школа»
1989

ББК 38.3я73
В 64
УДК 691:519(075.8)

Рецензенты: д-р техн. наук, проф. А. В. Ушеров-Маршак (Харьковский инженерно-строительный институт), канд. техн. наук, доц. О. Н. Сикорский (Киевский инженерно-строительный институт).

Редакционная группа литературы по строительству и архитектуре.
Редактор В. В. Тронько

Вознесенский В. А. и др.

В64 Численные методы решения строительно-технологических задач на ЭВМ: Учебник/В. А. Вознесенский, Т. В. Ляшенко, Б. Л. Огарков; Под ред. В. А. Вознесенского.—К.: Выща шк. Головное изд-во, 1989—328 с.: ил.

ISBN 5—11—001439—6.

Рассмотрены особенности применения и методики численных методов решения на ЭВМ задач по анализу и оптимизации структуры и свойств строительных материалов и изделий, а также технологических режимов их производства. Проанализированы основные принципы математического программирования и математической теории эксперимента, методы поиска оптимальных решений в строительно-технологических задачах, включая имитационное моделирование. Рекомендации подтверждены конкретными примерами.

Для студентов вузов, обучающихся по специальности «Производство строительных изделий и конструкций».

В $\frac{3306000000-257}{M211(04)-89}$ 222—89

ББК 38.3я73

ISBN 5—11—001439—6

© Издательское объединение
«Выща школа», 1989

ПРЕДИСЛОВИЕ

Электрификация, химизация, роботизация, компьютеризация, применение биотехнологий — эти пять направлений ускорения научно-технического прогресса определены Программой Коммунистической партии Советского Союза (новая редакция).

Наличие в производственной, проектной или научной системе ЭВМ само по себе не решает проблемы компьютеризации. Для функционирования автоматизированных систем управления технологическими процессами (АСУ ТП), систем автоматизированного проектирования (САПР) или автоматизированных систем научных исследований (АСНИ) необходимо их методическое и программное обеспечение, ориентированное на решение отраслевых инженерных задач строительного материаловедения и технологии. При этом определяющую роль в формировании надежных и полезных для научно-технической практики результатов играет этап математического моделирования объектов управления, проектирования, исследования и др. Эти математические модели могут быть разработаны только при непосредственном участии инженеров-технологов по производству строительных материалов и конструкций [8].

Новый для инженеров этого профиля курс «Численные методы решения строительно-технологических задач на ЭВМ», завершающий вузовскую математическую подготовку, ставит своей целью формирование у специалистов знаний, позволяющих сформулировать материаловедческие и технологические задачи в математических терминах, выбрать те или иные пути решения математической задачи и получить из результатов математического моделирования полезную инженерную информацию. В сочетании с лабораторным практикумом изложенный в учебнике материал должен способствовать формированию у инженера-технолога умения решать на ЭВМ типовые задачи анализа и оптимизации рецептурно-технологических ситуаций.

Учебник по данному курсу направлен, с одной стороны, на формирование достаточно конкретных инженерных знаний, а с другой — на повышение уровня общенаучной подготовки инженеров.

Методика изложения материала предусматривает прежде всего *самостоятельную работу* читателя с учебником. Сохраняя в качестве базовой обусловленную программой курса структуру, учебник в значительной степени отличается от традиционного построения

ния курсов по численным методам прежде всего тем, что он строго ориентирован на специалистов по *строительному материаловедению и технологии*, во-первых, по *уровню физико-математической подготовки* обучаемых, во-вторых по *кругу задач*, решаемых численными методами в отраслевой практике.

В связи с первым обстоятельством учебник включает анализируемый в гл. I ряд простейших понятий и методов численных решений (приближенные числа, приближенные решения, поиск корней уравнений, дифференцирование и интегрирование), которые могут быть знакомы читателям из курсов высшей математики (к сожалению, именно эти заключительные ее разделы, как правило, опускаются для студентов специальности «Производство строительных изделий и конструкций»), а также прил. I, в котором собраны базовые понятия линейной алгебры аналитической геометрии и теории вероятностей, необходимые для изучения, усвоения и практического применения содержащейся в данном учебнике информации. Читателю, недостаточно полно владеющему этим математическим аппаратом, рекомендуется изучить информацию, собранную в прил. I, сразу после ознакомления с предисловием и введением.

Второе обстоятельство потребовало, с одной стороны, отбора именно тех математических методов, которые доказали свою эффективность в отраслевых задачах (методы оптимизации, в том числе линейного программирования; методы математической теории эксперимента, в частности статистического анализа данных и принятия решений по экспериментально-статистическим моделям), а с другой — насыщения учебника примерами решения задач в области материаловедения и технологии. Следует отметить, что абсолютное большинство примеров в учебнике представляют собой фрагменты конкретных задач из практики авторов. Конечной формой представления в учебнике алгоритмов являются их блок-схемы, что делает рекомендации независимыми от языка программирования и класса ЭВМ.

Математические методы решения прикладных задач развиваются ускоренными темпами, поэтому авторы старались отметить те или иные проблемные вопросы как по ходу изложения основного материала (см., в частности, п. III.14, посвященный синтезу оптимальных планов эксперимента на ЭВМ), так и в специальной гл. IV, анализирующей основные принципы имитационного эксперимента и диалога ЭВМ — технолог (исследователь, проектировщик, оператор систем управления).

Авторы стремились к такому изложению курса, чтобы учебником можно было достаточно просто пользоваться как руководством при самостоятельной деятельности на всех этапах обучения, включая лабораторный и учебно-исследовательский практикум, курсовое проектирование по профилирующим дисциплинам, дипломное проектирование. При этом особую роль играют сквозные комплексные примеры, которые можно рассматривать как руководство по решению типовых материаловедческих и технологических задач.

Специфика научно-технической проблематики и ее новизна потребовали учесть при написании учебника необходимость переподготовки инженерных кадров, поскольку многие технологи слабо информированы о возможностях и методологии численных методов при решении строительно-технологических задач.

Предисловие, введение, гл. I (п. 1, 2), гл. III (п. 1 — 12 и 15 — 17), гл. IV (п. 1 — 4) написаны доктором технических наук, профессором В. А. Вознесенским; гл. II (п. 1 — п. 4), гл. III (п. 13, 14) и гл. IV (п. 5) — кандидатом технических наук, старшим научным сотрудником Т. В. Ляшенко; гл. I (п. 3 — 5) и гл. II (п. 6, 7) — кандидатом технических наук Огарковым Б. Л., а п. II.5 и приложения написаны В. А. Вознесенским и Б. Л. Огарковым совместно.

Авторы с благодарностью примут любые конструктивные замечания по данному учебнику. Их можно направлять в издательство по адресу: 252054, Киев, ул. Гоголевская, 7.

- A — коэффициент асимметрии
 A_{ij} — оценки коэффициентов приведенных полиномов
 \mathbf{A} — вектор
 $\|\mathbf{A}\|$ — норма вектора
 \mathbf{A} — матрица
 $A = \det \mathbf{A}$ — определитель матрицы
 \mathbf{A}^{-1} — обратная матрица
 \mathbf{A}^T — транспонированная матрица
 $\overset{\circ}{a}$ — точное число
 $a^{(0)}$ — первое приближение к $\overset{\circ}{a}$
 $\arg Y$ — аргумент функции Y
 b_i — оценка коэффициента ЭСМ при i -м нормализованном факторе
 $C = \text{const}$ — константа
 C — целевая функция в ЛП
 C_d — концентрация добавки
 C_n^m — число сочетаний из n по m
 c_{ii}, c_{ij} — элементы ковариационной матрицы \mathbf{D} , соответственно диагональной и внедиагональной
 $\text{cov}\{Y_1, Y_2\}$ — оценка ковариации двух случайных величин
 $D\{Y\}$ — дисперсия
 \mathbf{D} — ковариационная матрица
 d — мера точности предсказания модели
 $\det \mathbf{A} = |\mathbf{A}|$ — определитель матрицы \mathbf{A}
 E — модуль упругости материала
 $E(Y)$ — математическое ожидание
 E^* — оценка эксцесса
 F — критерий Фишера
 $F(Y)$ — закон распределения случайной величины Y
 f — число степеней свободы
 $f(X^*) = f^*$ — экстремум функции
 $f_N(Y)$ — плотность нормального распределения Y
 G — критерий Кокрена
 $g(x)$ — функция ограничения — неравенство
 H_0 — нуль-гипотеза
 $h\{b_i\}$ — гауссовская точность оценки коэффициента b_i
 $h(x)$ — функция ограничения — равенство
 I — инвариант
 k — число независимых переменных, число факторов в ЭС-модели
 $L\{b_i\}$ — число коэффициентов в ЭС-модели
 \mathbf{M} — матрица моментов плана
 \mathbf{M}_N — матрица плана информационной модели
 Me — медиана
 m_u — число параллельных измерений в u -ом опыте
 N — число опытов
 $P(A)$ — вероятность события A
 P_F — вероятность получения допустимого решения
 p — оценка частоты
 R — прочность материала
 R^k — евклидово k -мерное пространство
 $r\{Y_1, Y_2\}$ — оценка коэффициента корреляции между Y_1 и Y_2
 Q_α — квантиль, соответствующая риску α
 q — число линейно связанных факторов
 SS — сумма квадратов
 s — номер шага процесса (итерации)
 $s^2\{Y\}$ — оценка дисперсии Y
 $s\{Y\}$ — оценка среднеквадратического отклонения Y
 $s\{b_i\}$ — ошибка оценки коэффициента
 $s_{\text{ост}}^2$ — оценка остаточной дисперсии
 $s_{\text{на}}^2$ — оценка дисперсии неадекватности
 T — температура
 T_i — коэффициент при расчете b_i по симметричным планам
 t_α — критерий Стьюдента, соответствующий риску α
 $v\{Y\}$ — оценка коэффициента вариации Y
 v — доля компонента в смеси
 $W\{Y\}$ — размах в выборке Y
 W_i — выход квазиоднофакторной модели

X_i — натуральное значение i -го фактора (кг, м, с и т. п.)

$X_{0,i}$ — основной уровень фактора

x — нормализованное значение фактора ($|x| \leq 1$)

$x^{(0)}$ — корень

x^* — координата экстремума

\tilde{x} — главная ось при канонической форме поверхности

Y — выход, отклик системы

\hat{Y} — расчетное (прогнозируемое) значение отклика

y_{uv} — частное значение Y в v -ом измерении u -го опыта

Z — преобразованная переменная

α — риск ошибки первого ряда

β_i — истинный коэффициент полиномиальной ЭС-модели

$\gamma\{Y\}$ — истинный коэффициент вариации Y

γ — статистический критерий

ΔX_i — интервал варьирования i -го фактора

$\Delta\{X\}$ — абсолютная погрешность, ошибка

ΔY — прирост выхода за шаг процесса

$\delta(Y)$ — относительная погрешность, ошибка

E — план эксперимента

ϵ — малая величина (случайная или детерминированная)

ϵ_N — нормально распределенная случайная величина

$\eta(Y)$ — истинное среднее, математическое ожидание Y

η — вязкость технологической смеси

Θ — истинное значение параметра

Θ^* — оценка параметра

ν — истинная частота

ξ — случайное воздействие, случайная величина

$\rho\{Y_1, Y_2\}$ — истинный коэффициент корреляции между Y_1 и Y_2

ρ — плотность материала

σ^2 — дисперсия

σ — среднее квадратическое отклонение

τ — время

φ_N — плотность нормального распределения

Ω — множество, область

Ω_x — исследуемая область

$\Omega_{зпр}$ — запрещенная область

$\nabla f(x)$ — градиент функции

$\bar{1}, \bar{k}$ — отрезок натурального ряда

\otimes — тензорное произведение функционалов

\cup — объединение

\cap — пересечение

\vee — дизъюнкция («или ... или» ...)

\wedge — конъюнкция («и... и...»)

\subseteq — подмножество

\in — принадлежность

$[a, b]$ — отрезок

$\{a, b, \dots\}$ — множество

(00) — сумма $\sum_{u=1}^N x_{0u} = N$

(0i) — сумма $\sum_{u=1}^N x_{iu}$

(0Y) — сумма $\sum_{u=1}^N y_u$

(iY) — сумма $\sum_{u=1}^N x_{iu} y_u$

(ii) — сумма $\sum_{u=1}^N x_{iu}^2$

(ijY) — сумма $\sum_{u=1}^N x_{iu} x_{ju} y_u$

(iiY) — сумма $\sum_{u=1}^N x_{iu}^2 y_u$

$A_{\text{в}}$ — верхний предел A

$A_{\text{кр}}$ — критический уровень

$A_{\text{н}}$ — нижний предел

$A_{\text{норм}}$ — нормативный уровень

$A_{\text{общ}}$ — общее по массиву данных

$A_{\text{т}}$ — табличное значение

$A_{\text{тр}}$ — требуемое, заданное значение

$A_{\text{факт}}$ — фактическое значение

$A_{\text{э}}$ — экспериментальное значение

A_{ext} — экстремальное значение

A_{opt} — оптимальное значение

A — расчетное значение

$A_{0,05} = A_{05}$ — значение A с риском

ошибки $\alpha = 0,05$

ДФЭ — дробный факторный эксперимент

ДШМ — диссоциативно-шаговый метод

ЛП — линейное программирование

МНК — метод наименьших квадратов

МТЭ — математическая теория эксперимента

ОЗЛП — основная задача линейного программирования

ПРА — последовательный регрессионный анализ

ПФЭ — полный факторный эксперимент

ЭСМ — экспериментально-статистическая модель

В производстве строительных материалов и конструкций в связи с ускорением научно-технического прогресса как в народном хозяйстве в целом, так и непосредственно в этой отрасли происходит интенсивный процесс усложнения программных технических задач, технико-экономических условий их решений, научно-технических подходов к созданию и исследованию новых материалов и технологий, средств реализации оптимальных инженерных решений, обеспечивающих производительность труда, ресурсосбережение, гарантированное качество продукции и др.]

Усложнение основных объектов (материалов как конечной продукции, собственно технологических процессов и реализующих их аппаратов, технологических линий и комплексов, процессов создания новой техники и т. п.) приводит к прогрессирующему росту потерь от ошибочных или ненадежных решений по развитию и функционированию этих объектов (лат. *functio* — деятельность). Для уменьшения вероятности таких ошибок необходимо, с одной стороны, основывать решения на рекомендациях фундаментальных и прикладных наук при системном подходе к объектам, с другой — использовать возможности вычислительной техники для всестороннего анализа объекта и выбора путей оптимизации его структуры, свойств, поведения и пр. Диалектическая связь между этими сторонами процесса принятия инженерных решений обеспечивается математическими моделями объекта и программным обеспечением ЭВМ.

Обобщенная схема [52] *управляемой системы* показана на рис. В.1. Система как единое целое *выделена* из окружающей среды и в то же время *связана* с ней. Уровень целостности может быть самым разным: от макросистем (отрасль производства в целом, завод, технологическая линия и т. п.) до микросистем (структура композиционного материала, элементарный процесс переноса вещества или энергии и т. п.) с границами, определяемыми условиями конкретной задачи. Система с собственными параметрами $X_{s,i}$ (количество дефектов в структуре материала, размеры аппарата, температура кипения рабочего тела и т. п.) связана со средой через *входные факторы* (лат. *factor* — делающий) X_i , отражающие контролируемое (наблюдаемое) влияние среды на систему, через *возмущающие* (неконтролируемые, ненаблюдаемые, случайные) *воздействия* среды

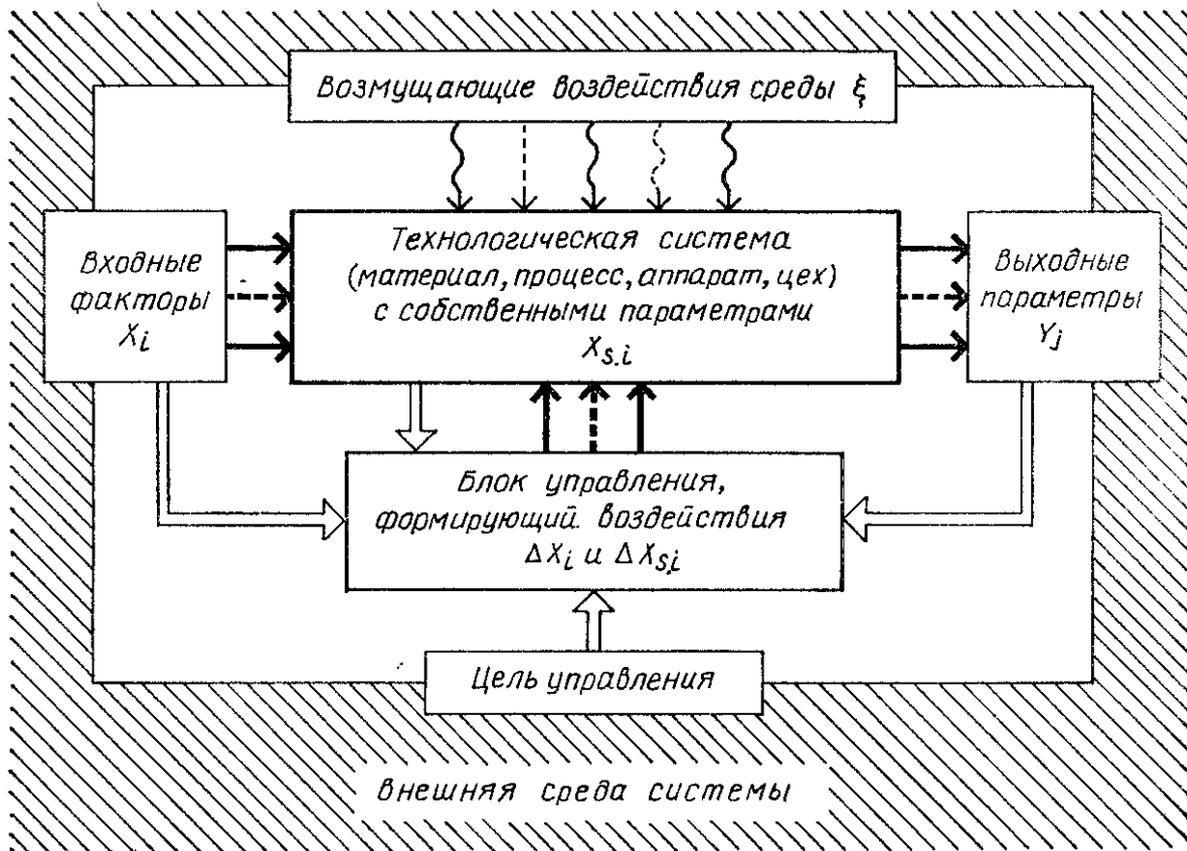


Рис. В.1. Принципиальная схема управляемой системы и ее связи с внешней средой

и через *выходные параметры* Y_j , отражающие *реакцию*, или *отклик*, объекта на изменения во внешней среде и характеризующие его влияние на эту среду.

В технологии строительных материалов целесообразно [6] выделить 9 взаимосвязанных *подсистем факторов* (в дальнейшем в это понятие включаются и параметры состояния системы $X_{s,i}$, если в их выделении нет особой необходимости).

Шесть подсистем характерны для заводской технологии:

X_c — параметры качества *сырьевых* компонентов;

X_p — *рецептура*, задаваемая, во-первых, как *абсолютная* в килограммах на единицу продукции или загрузку аппарата, во-вторых, как *относительная* в долях к общей массе или в долях к одному компоненту, принимаемому за единицу;

$X_{см}$ — параметры *режима смешивания* (скорости течения и деформации неньютоновских жидкостей, градиенты давлений и температур, продолжительность цикла и т. п.);

X_f — параметры *режима формования*, под которым понимаются не только агрегатирование частиц и формообразование материала в изделие, но и измельчение, если конечный продукт или полуфабрикат порошкообразный;

$X_{п.п}$ — параметры *режима переходных процессов*, приводящих к агрегатным или фазовым превращениям вещества (в частности, к гидратации, полимеризации, сушке, спеканию, сплавлению и т. п.); они определяют в отформованном изделии образование из подготовленной сырьевой смеси материала как такового;

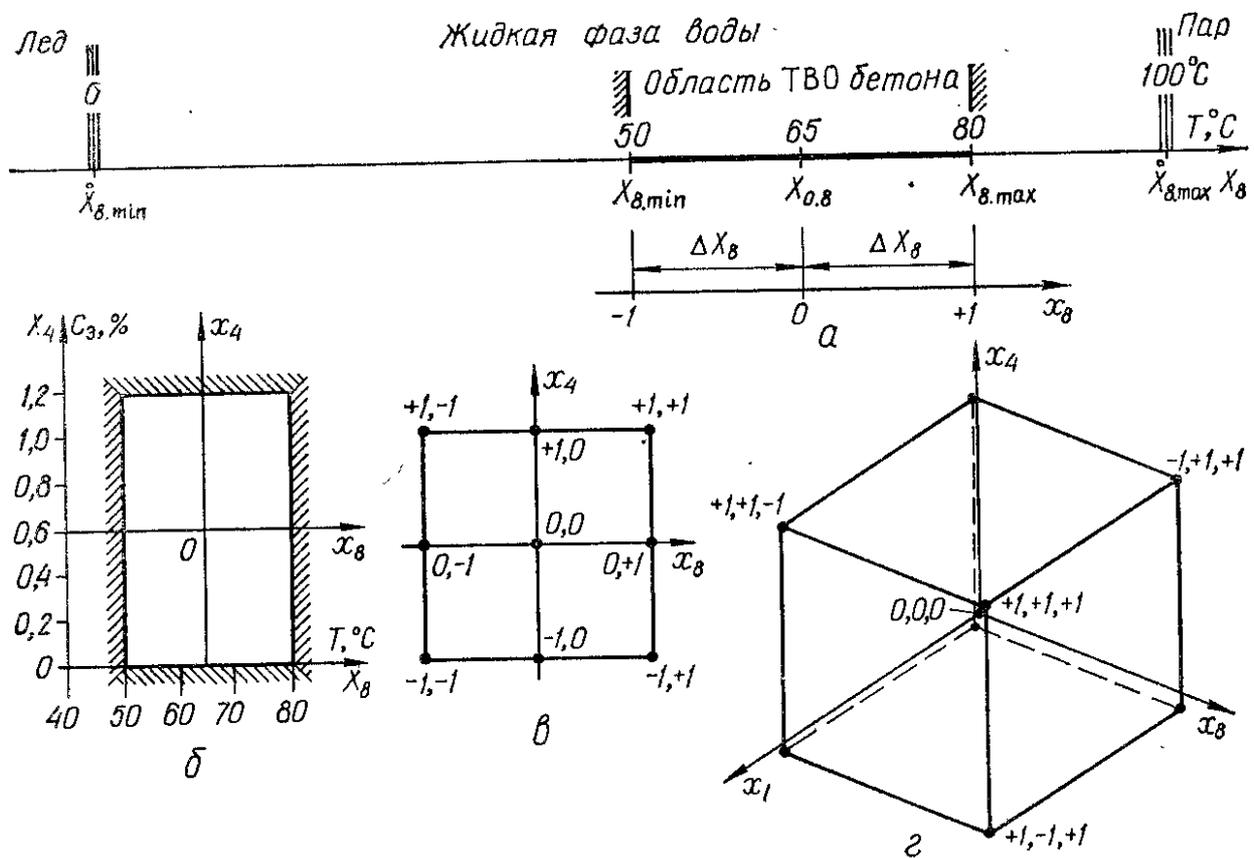


Рис. В.2. Факторное пространство:

а — один фактор (температура воды) натуральный X_8 и нормализованный $-1 \leq x_8 \leq +1$; б — двухфакторное пространство в натуральных переменных X_8 (температура ТВО) и X_4 (концентрация добавки электролита); в — то же, пространство в нормализованных переменных — квадрат; г — куб как трехфакторное пространство нормализованных переменных

X_m — параметры *режима модификации*, направленного на улучшение структуры материала без изменения формы изделия (например, пропитка полимеризующими составами).

Три группы факторов обособлены от заводского технолога, но должны быть им учтены для повышения народнохозяйственного эффекта от применения материалов и конструкций в сооружениях:

X_n — параметры *условий исследования* (контроля качества, испытаний и т. п.), обеспечивающих достоверную информацию о всей системе и отдельных ее элементах;

X_k — *конструктивные* параметры изделия, обеспечивающие рациональное использование свойств материала в макрообъемах;

X_s — параметры *условий эксплуатации*, определяющие применимость материала или изделия в конкретном сооружении.

Выходы системы образуют четыре связанные группы:

Y_t — *технологические показатели* свойств, необходимые для оценки воздействия факторов начального уровня на последующие технологические процессы (например, для оценки X_c и X_p , влияющих на X_{cm} , используется такой показатель Y_t , как удобоукладываемость бетонной смеси);

Y_s — *структурные показатели*, характеризующие внутреннее строение материала или изделия (например, пористость и т. п.);

Y_o — *показатели основного назначения* продукции, образующие его качество определенного сорта (класса и т. п.), *технико-экономические показатели* производства;

Y_α — вероятностные показатели качества и надежности (например, вероятность появления бракованных изделий в партии и т. п.).

Каждый фактор X_i как входная (переменная, управляемая) характеристика системы в силу объективных физических или нормативных экономических возможностей существует лишь в некотором диапазоне $\overset{\circ}{X}_{i.\min} \leq X_i \leq \overset{\circ}{X}_{i.\max}$. Так, концентрация любого компонента (в % к общей массе смеси), например содержание в щебне зерен со средним диаметром $d = 7,5$ мм, не может быть меньше 0 (т. е. отрицательной величиной) и больше 100 %; вода при атмосферном давлении находится в жидком состоянии в диапазоне от 0 до 100° С (рис. В.2, а) и т. п. Однако весь возможный диапазон существования фактора $\overset{\circ}{X}_{i.\min} \dots \overset{\circ}{X}_{i.\max}$ обычно не исследуется. Так как технолог стремится к конкретизации результатов, то ему достаточно проанализировать влияние фактора в более узком диапазоне, характерном для данной рецептурно-технологической ситуации:

$$\overset{\circ}{X}_{i.\min} \leq (X_{i.\min} \leq X_i \leq X_{i.\max}) \leq \overset{\circ}{X}_{i.\max}. \quad (\text{В.1})$$

Заклученный в скобках в (В.1) диапазон изменения X_i называется *размахом варьирования* фактора, половина размаха носит название *интервала варьирования* ΔX_i , а середина диапазона варьирования X_{0i} — *основного уровня* фактора (ГОСТ 24026—80):

$$\Delta X_i = 0,5 (X_{i.\max} - X_{i.\min}). \quad (\text{В.2})$$

$$X_{0i} = 0,5 (X_{i.\max} + X_{i.\min}). \quad (\text{В.3})$$

При экспериментально-статистическом моделировании целесообразно (как со статистической и вычислительной позиций, так и с точки зрения интерпретации, что будет показано далее) переходить от натуральных переменных (именованные величины с размерностью) к *безразмерным нормализованным переменным*

$$x_i = (X_i - X_{0i}) / \Delta X_i. \quad (\text{В.4})$$

Из (В.4) следуют три важных для практики моделирования соотношения:

$$\left. \begin{array}{l} \text{если } X_i = X_{i.\min}, \text{ то } x_i = -1, \\ \text{если } X_i = X_{0.i}, \text{ то } x_i = 0, \\ \text{если } X_i = X_{i.\max}, \text{ то } x_i = +1. \end{array} \right\} \quad (\text{В.5})$$

При нормализации факторов, во-первых, происходит центрирование переменных в результате переноса начала координат в точку $X_{0.i}$, во-вторых, единица масштаба новой шкалы равна интервалу варьирования фактора ΔX_i (рис. В.2, а). Например, при моделировании влияния тепловлажностной обработки (ТВО) бетона на его свойства [6] температура изотермической выдержки T меняется в диапазоне от $X_{8.\min} = 50^\circ\text{C}$ до $X_{8.\max} = 80^\circ\text{C}$ (детальная поста-

новка задачи — см. пр. * III.1), хотя диапазон существования фактора от $\overset{\circ}{X}_{8, \min} = 0^\circ\text{C}$ до $\overset{\circ}{X}_{8, \max} = 100^\circ\text{C}$. На основе (В.2)—(В.4) интервал варьирования ΔX_8 , основной уровень фактора $X_{0,8}$ и его нормализованная величина $|x_8| \leq 1$ определяются как

$$\left. \begin{aligned} \Delta X_8 &= 0,5(80 - 50) = 15^\circ\text{C}; & X_{0,8} &= 0,5(80 + 50) = 65^\circ\text{C}; \\ x_8 &= (X_8 - 65)/15. \end{aligned} \right\} \quad (\text{В.6})$$

При нормализации двух факторов ограничивающий факторное пространство прямоугольник трансформируется в квадрат, через центр которого проходят оси координат x_i . Например, если моделируется влияние ТВО на свойства бетона с химической добавкой-электролитом (см. пр. III.1), концентрация которой C_3 изменяется в диапазоне $0 \leq X_4 \leq 1,2\%$ (от массы цемента), то факторное пространство $\{T, C_3\}$ (рис. В. 2, б) после нормализации превращается в квадрат $\{x_8, x_4\}$ с координатами вершин $\{\pm 1, \pm 1\}$ (рис. I.2, в).

Трехфакторная прямоугольная призма после нормализации переменных трансформируется в куб с вершинами $\{\pm 1, \pm 1, \pm 1\}$ (рис. В. 2, г), а k -мерный прямоугольник — в гиперкуб (греч. «гипер» — сверх) с вершинами $\{\pm 1, \pm 1, \dots, \pm 1\}$ и осями координат, проходящими через его центр $\{0, 0, \dots, 0\}$.

Возврат от нормализованных безразмерных переменных x_i (В.5) к натуральным, необходимый при проведении экспериментальных работ и интерпретации моделей, происходит по вытекающему из (В.4) соотношению

$$X_i = x_i \Delta X + X_{0,i}. \quad (\text{В.7})$$

Например, точка $x_8 = -0,8$ (рис. В. 2, а) с учетом (В.6) будет соответствовать температуре $X_8 = -0,8 \cdot 15 + 65 = 53^\circ\text{C}$.

Число N_x факторов X , влияющих на результаты Y работы системы, весьма велико ($N_x \rightarrow \infty$). Управление системой ведется по ограниченному числу k тех контролируемых факторов X_i , которые, по мнению технолога, наиболее существенно влияют на результат Y . Остальные $N_x - k$ факторов входят в группу возмущающего воздействия среды ξ , уровень влияния которой случаен. Наличие в технологической системе случайных воздействий ξ обуславливает ее поведение как системы стохастического класса [4, 8], в которой случайность не только действует наряду с необходимостью, но и в значительной степени определяет причинно-следственные связи; это и приводит к тому, что выходы системы являются величинами не детерминированными, а случайными.

Поскольку система управляемая, то для целенаправленного изменения значений выходов Y_j и для компенсации возмущений ξ она подвергается управляющим воздействиям ΔX_i и $\Delta X_{s,i}$, формируемым на основе информации I об уровнях Y_j , X_i и $X_{s,i}$.

* Здесь и далее сокращение «пр» — пример.

Под информацией I понимаются фактические сведения о структуре системы и происходящих в ней явлениях, о состоянии и поведении и т. п. (как в виде количественных оценок, так и в виде качественных сообщений).

Управляющие воздействия ΔX_i и $\Delta X_{s,i}$ могут быть сформированы в блоке управления (человек, директорат, управляющая машина и т. п.) только после того, как поступившая в него информация $I(X_i, X_{s,i}, Y_j)$ сравнена с соответствующим набором требований к объекту ($X_{i,\min} \leq X_j \leq X_{i,\max}$, $X_{s,i,\min} \leq X_{s,i} \leq X_{s,i,\max}$, $Y_{j,\min} \leq Y_j \leq Y_{j,\max}$), который и образует цель управления объектом. Цель управления, так же как и блок управления, одновременно принадлежит и управляемой системе (поскольку без них она будет существовать как не управляемая) и внешней среде. Необходимые управляющие воздействия можно определить, если известна взаимосвязь между входами, собственными параметрами и выходами системы. Эта объективно существующая взаимосвязь

$$\psi(Y_j, \Theta_\psi, X_\psi, \tau_\psi, \xi_\psi) = 0 \quad (\text{В.8})$$

по аналогии с термодинамикой может быть названа уравнением состояния системы. Однако в технологических, технико-экономических, экологических и других реальных сложных системах неизвестны ни вид функций ψ , ни полный набор факторов X_ψ , ни числовые значения параметров (коэффициентов) Θ_ψ , ни законы распределения случайных величин ξ_ψ . Отсутствие знаний об истинном уравнении (В.8) не позволяет решать инженерные задачи, поэтому неопределенность на практике устраняется за счет замены анализа выражения (В.8) на анализ некоторого приближенного с известной точностью приближения) описания взаимосвязей в системе. Формальное абстрактно-знаковое описание системы (в виде набора чисел, графиков, уравнений, неравенств, логических выражений, алгоритмов и т. п.), позволяющее выносить суждение о некоторых чертах поведения этой системы с помощью формальных процедур над этим описанием, можно назвать математической моделью M [4, 6].

Как и модель любого класса, математическая модель M должна соответствовать объекту-прототипу O (в частности, системе на рис. В.1) лишь ограниченно в рамках, обусловленных целями анализа или управления поведением O . Критерием истинности модели служит инженерная или технико-экономическая полезность новой информации I_M , полученной по M при последующей проверке ее в системе O [6]. Чтобы модель M оказалась полезной, она должна обладать двумя диалектически связанными свойствами.

Во-первых, модель должна быть экономичной, т. е. позволять достичь поставленной цели при меньшем расходе ресурсов (дешевле, быстрее, безопаснее и т. п.), чем при использовании прототипа O .

Во-вторых, модель должна обеспечивать возможность распространения без существенного искажения (с известным риском ошибки [6]) информации I_M на систему O , т. е. обладать свойством транзитивности (лат. *transitus* — переход).

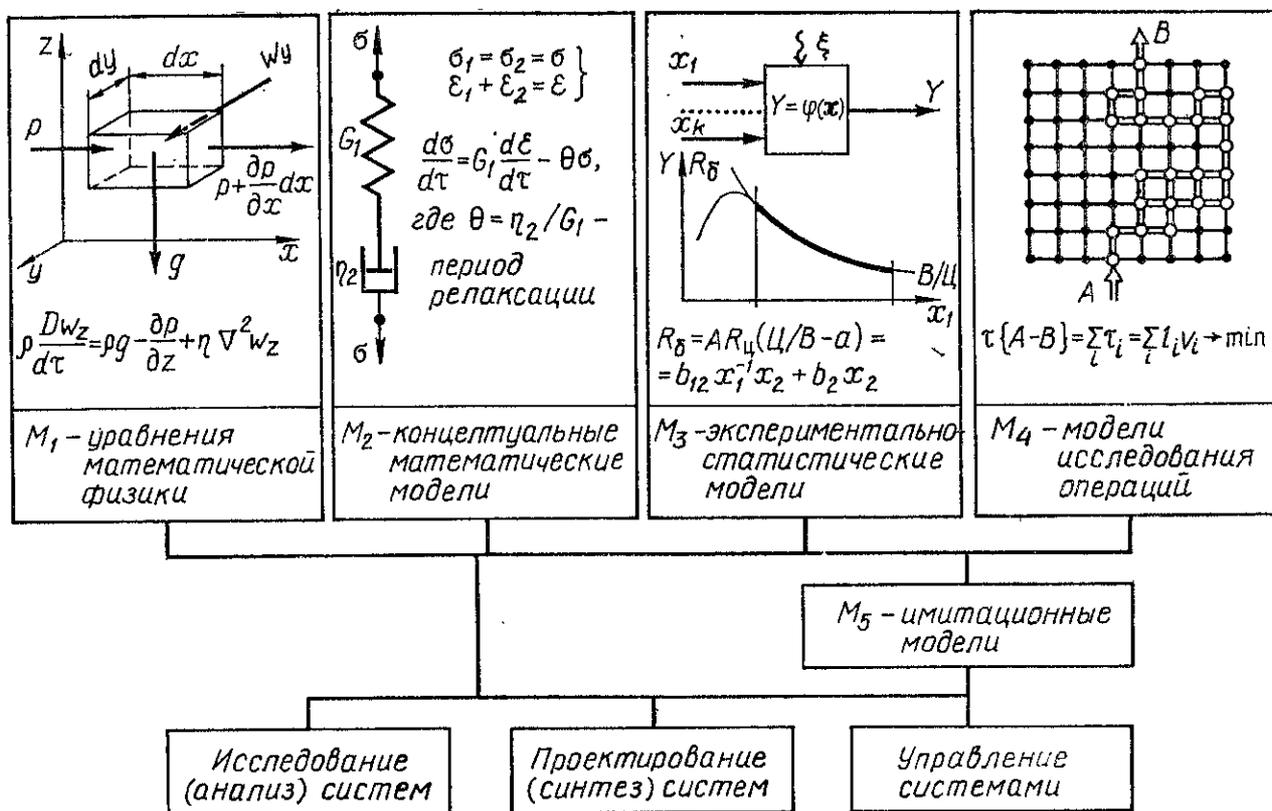


Рис. В.3. Пять типов различных по своей природе моделей и их связь с целями моделирования

При математическом моделировании используются практически все разделы математики. В основе выбора того или иного математического метода должны лежать целевая установка моделирования, теоретические представления об объекте, накопленная информация о нем, ограничения на материальные и временные ресурсы и т. п. с учетом сложившейся практики решения инженерных задач в области строительного материаловедения и технологии целесообразно выделить [8] пять основных типов математических моделей (рис. В.3), различных по своей природе.

Модели M_1 (например, уравнение движения вязкой жидкости Навье — Стокса [29, 52]) основываются на анализе бесконечно малых изменений в физической системе бесконечно малого объема. При этом число допущений мало (например, «поток неразрывен» и вязкость жидкости зависит только от температуры, т. е. жидкость — «ньютоновская» [52]), а при составлении таких моделей — уравнений математической физики — используются фундаментальные законы природы (в частности, знак равенства в уравнении Навье — Стокса отражает закон сохранения энергии). Для инженерных расчетов требуется переход от бесконечно малых к конечным (измеримым) изменениям в реальной системе, однако именно этот переход, как правило, аналитически столь сложен (или вообще невозможен в общем виде), что интегрирование заменяется теми или иными приближенными методами. В качестве инженерного средства преодоления математико-аналитических «барьеров» в конце XIX — начале XX вв. сформировалось специальное научное направление — теория подобия, соединившая дифференциальные урав-

нения с экспериментальной информацией о системе. В связи с развитием ЭВМ первостепенную роль приобрели численные методы решения задач математической физики.

Модели M_2 имеют в своем основании прежде всего некоторую инженерную *концептуальную* (лат. *conceptus* — представление, мысль) модель, выраженную первоначально словами в терминах данной науки (материаловедение, технология). Эта концептуальная модель описывается далее абстрактно-знаково с помощью дифференциальных или алгебраических уравнений, геометрических соотношений, логических операций и т. п. На рис. В. 3 рассмотрено реологическое тело Максвелла [52] с периодом релаксации, равным отношению между ньютоновской вязкостью η_2 и модулем упругости Гука G_1 .

В основе математической модели лежит представление о том, что поведение реального материала под нагрузкой аналогично поведению последовательно соединенной пружины и поршня, движущегося в вязкой жидкости. Преимущества моделей M_2 ясны — они легко поддаются «физической» интерпретации технологом, ибо эта интерпретация уже заложена в концептуальном основании модели. В то же время именно в этом же основании заложен и коренной недостаток моделей M_2 , поскольку можно корректно описать математическими символами концепцию, весьма далекую от реальности. Следует подчеркнуть, что для использования моделей M_2 необходимо опытное определение различных параметров объекта (в частности, вязкости η_2 и упругости G_1 конкретной жидкости и т. п.).

Модели M_3 (рис. В. 3 отражает известную в строительном материаловедении зависимость прочности бетона от водосодержания, обнаруженную в 1885 г. проф. И. Г. Малюгой и описанную формулой Болломея — Скрамтаева в 1922 — 1936 гг.) описывают с известной точностью связь между входами X и выходом Y системы на основе *экспериментальных данных без анализа внутренней структуры* системы, которую можно рассматривать как «черный ящик» [6, 52]. Модели M_3 отличаются крайне *простыми* исходными *предпосылками* («связь может существовать и ее можно описать формулой или алгоритмом»), что является источником и достоинств и недостатков. Одно из основных достоинств — *универсальность* методологии оптимального сбора экспериментальной информации и построения моделей. К числу недостатков следует отнести сложность всесторонней интерпретации параметров моделей M_3 и ограниченность области их действия с заданной точностью рамками тех конкретных объектов, для которых они построены. Модели M_3 всегда занимали, и по-видимому, будут занимать основное место в инженерной практике, причем их позиции значительно укрепляются в связи с расширением выпуска микроЭВМ и персональных компьютеров.

Модели M_4 *исследования операций* как совокупности действий, направленных на достижение некоторой цели [8, 59], характерны главным образом для технико-экономических задач (управление запасами, распределение ресурсов между производствами, задачи замены оборудования и контроля качества и т. п.). Они пока еще

только начинают входить в практику решения задач строительного материаловедения и технологии. На рис. В. 3 показан один из наиболее интересных случаев применения моделей M_4 , когда фильтрация жидкости через бетон рассматривается как движение через сеть случайной конфигурации с запертыми узлами [50].

В реальных задачах, как правило, нет четких границ между M_i , причем наиболее часто объединяются модели $M_1 + M_2$ и $M_2 + M_3$. Однако действительное объединение моделей всех типов для всестороннего описания, анализа и оптимизации поведения объекта стало возможным только благодаря применению вычислительной техники.

Имитационная модель M_5 включает в единую программу для ЭВМ модели любых типов, объединенные правилами взаимоперехода, а также *генератор случайных чисел*, моделирующий возмущающие воздействия среды ξ на систему (см. рис. В. 1). Можно создать имитационные модели, в состав которых вошли бы программные блоки, обеспечивающие активный диалог человека и ЭВМ [44 и др].

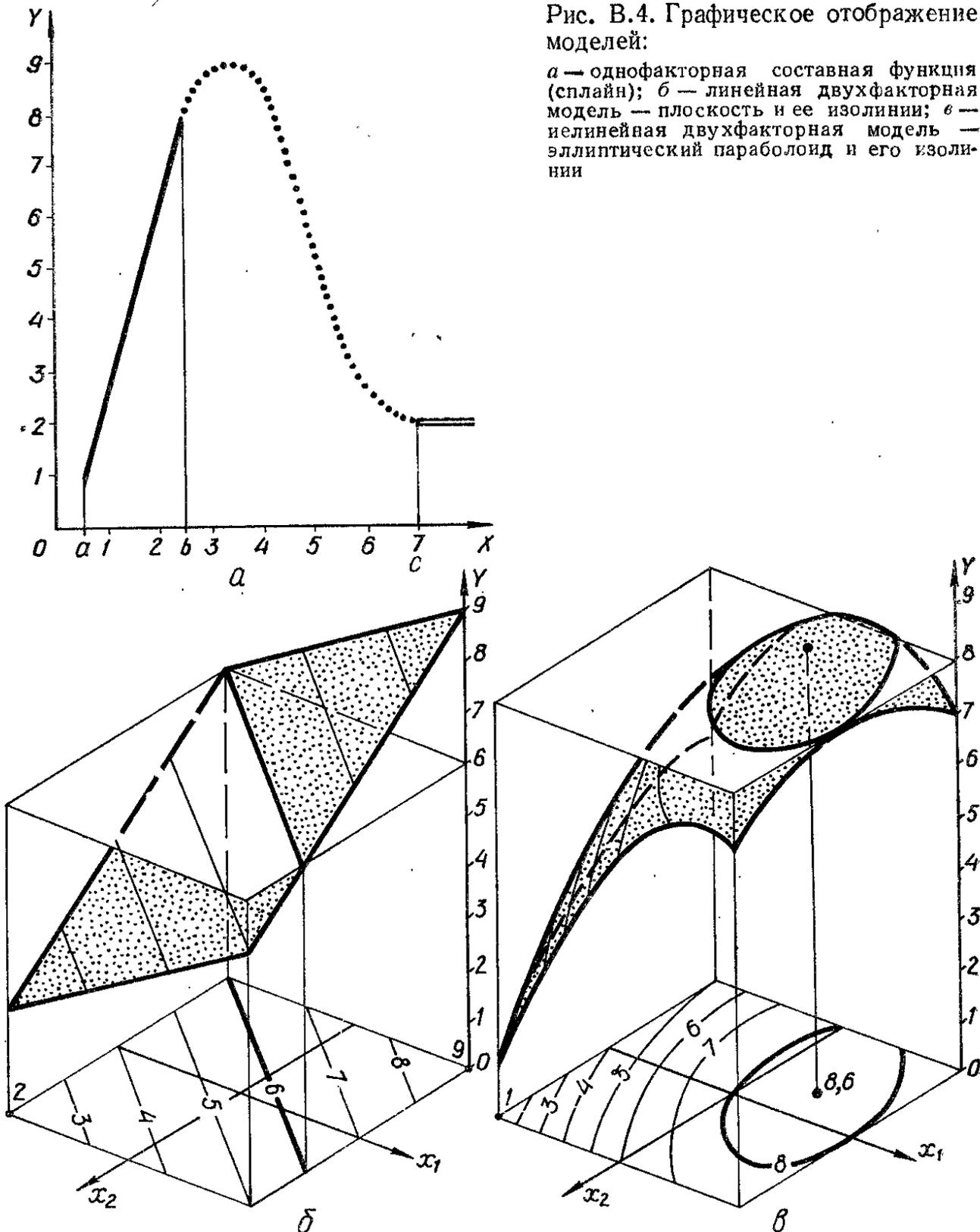
Математические модели являются эффективным средством решения задач во всех трех основных областях деятельности инженеров-материаловедов и технологов — при исследовании и проектировании систем, при управлении их функционированием (рис. В. 3). Однако роль каждой из моделей в разных областях деятельности не равноценна. В задачах анализа системы, особенно направленного на познание внутреннего механизма происходящих в ней явлений, особое значение имеют модели M_1 и M_2 , природа которых обеспечивает углубленную интерпретацию результатов моделирования. В то же время в задачах управления системой, когда главной задачей является достижение конкретной цели функционирования конкретной системой в конкретных условиях, приоритет переходит к моделям $M_3—M_5$, в которых именно эту конкретность удается, как правило, отразить с требуемой инженеру полнотой и точностью. При проектировании системы на этапе поискового синтеза предпочтительнее модели M_1 и M_2 , отражающие более общую инженерную информацию о системах данного класса, а на стадии подготовки синтезированной системы к реализации повышается значение моделей M_3 , M_4 и особенно M_5 .

Работа с моделями всех классов существенно упрощается, если можно использовать их *графические отображения*. Так, однофакторная модель отображается единственной линией (рис. В. 4, а) — графиком функции $Y = f(X)$. При этом она может быть кусочной: в разных диапазонах изменения фактора X функции, составляющие модель, — различны ($a \leq X \leq b$ — линейное увеличение; $b \leq X \leq c$ — экстремальная кривая; $c \leq X$ — константа; при граничных уровнях X значения выхода по обеим «стыкующимся» функциям равны). Такая составная функция получила название сплайн [21].

Двухфакторная модель может быть отображена семейством однофакторных функций $Y = f_1(X_1)$ при нескольких фиксированных значениях другой переменной $X_2 = C_x = \text{const}$ (и наоборот). Однако наиболее удобное ее отображение — плоскостная диаграм-

Рис. В.4. Графическое отображение моделей:

а — однофакторная составная функция (сплайн); б — линейная двухфакторная модель — плоскость и ее изолинии; в — нелинейная двухфакторная модель — эллиптический параболоид и его изолинии



ма, составленная из *линий равного уровня отклика системы*, так называемых *изолиний*. Диаграммы из изолиний $Y = f(X_1, X_2) = C_Y = \text{const}$ широко распространены в науке и технике: в географии на карту наносятся *изогипсы* — линии равных высот поверхности земли над уровнем моря (другой термин — *горизонталы*), в метеорологии — *изобары*, т. е. линии равного атмосферного давления на карте; в химии — на трехкомпонентной диаграмме «состав — свойство» линии равной температуры плавления веществ — *изотермы* и т. п. Изолинии образуются (рис. В. 4, б и в) как линии *пересечения* поверхности отклика Y с горизонтальными плоскостями,

проведенными через определенные отметки вертикальной шкалы Y ; при этом, как правило, шаг ΔY (или расстояние между секущими плоскостями) должен быть постоянным. Изолинии с поверхности отклика проектируются на плоскость факторного пространства и образуют диаграмму. Если поверхность отклика-плоскость (рис. В. 4, б), то при условии $\Delta Y = \text{const}$ диаграмма будет составлена из параллельных равноотстоящих прямых; во всех остальных случаях диаграммы составлены или из кривых (рис. В. 4, в), или из неравноотстоящих прямых. Следует отметить, что если заданному набору координат $X_1 = C_1$ и $X_2 = C_2$ соответствует одно значение отклика Y , то одному заданному уровню $Y_{\text{тр}} = C_Y$ соответствует множество $\{X_1, X_2\}$ точек, лежащих на этой изолинии. Для того чтобы при изменении одного из факторов (в частности, X_1) уровень $Y_{\text{тр}}$ сохранялся, необходимо регулировать уровень другого фактора X_2 , поскольку они жестко взаимосвязаны функцией $Y_{\text{тр}} = f(X_1, X_2)$, т. е. $X_2 = \varphi(X_1, Y_{\text{тр}})$.

Несмотря на многообразие объектов и целей, рассматриваемых в реальных задачах строительного материаловедения и технологии, все задачи, по существу, содержат требование о *доказательном* выборе в *ограниченных* пределах наиболее *целесообразного варианта* инженерного решения и, таким образом, должны рассматриваться как задачи *оптимизации*.

Цели оптимизации в общем могут быть двух альтернативных (лат. *alternare* — чередовать) типов. *Первый* — достигнуть *максимума* одного из выходов системы $Y_j \rightarrow \text{max}$, выбранного в качестве *критерия оптимизации* при оценке степени приближения к цели, израсходовав при этом *любую часть* доступного (нормативного) ресурса $X \leq X_{\text{норм}}$. В повседневной инженерной практике значительно чаще встречаются задачи *второго* типа — достигнуть *требуемого уровня* одного из показателей качества $Y_j \geq Y_{\text{тр}} = C_Y = \text{const}$ функционирования системы при *минимальном расходе ресурсов*. При этом требуемый уровень $Y_{\text{тр}} = \text{const}$ — это лишь *нижнее предельное* состояние системы; естественно, что во многих задачах допускается превышение такого состояния.

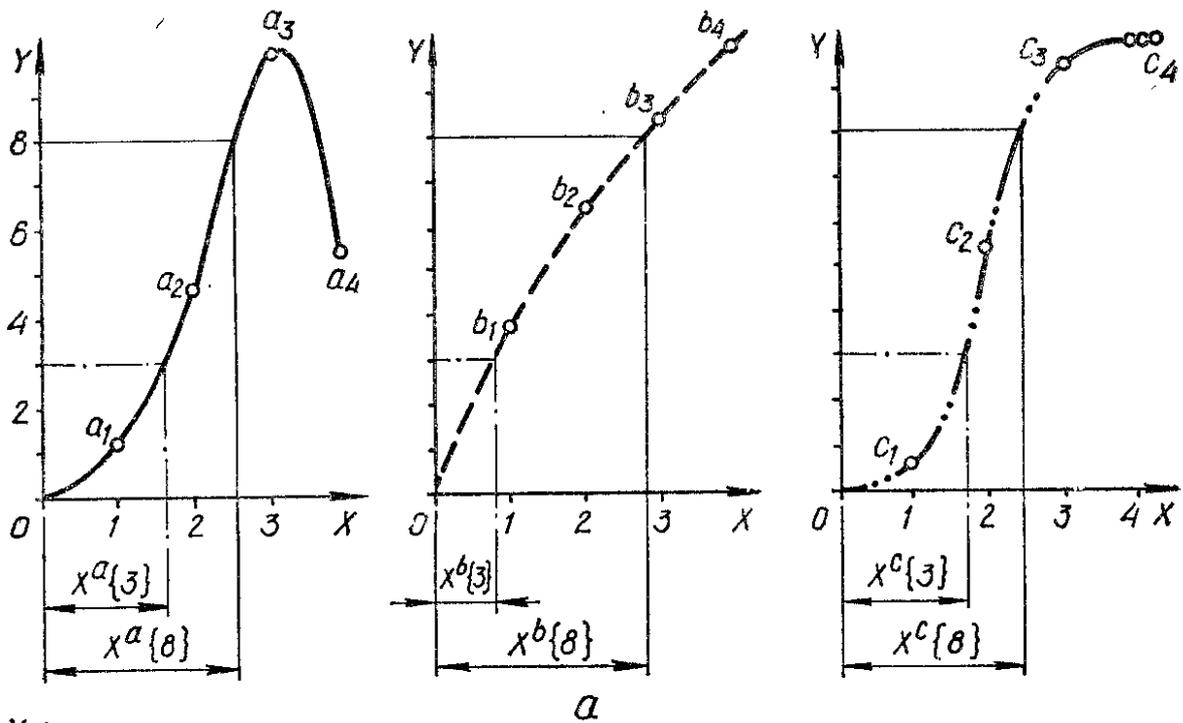
Возникающие при решении задач оптимизации простейшие типовые ситуации иллюстрируются примерами на рис. В. 5, а, где показаны три характерные для строительно-технологических задач функции (траектории достижения цели).

Задача первого типа: если возможен расход ресурсов $X_{\text{норм}} = 4$, то три результата ранжируются (фр. *ranger* — ставить по росту) как

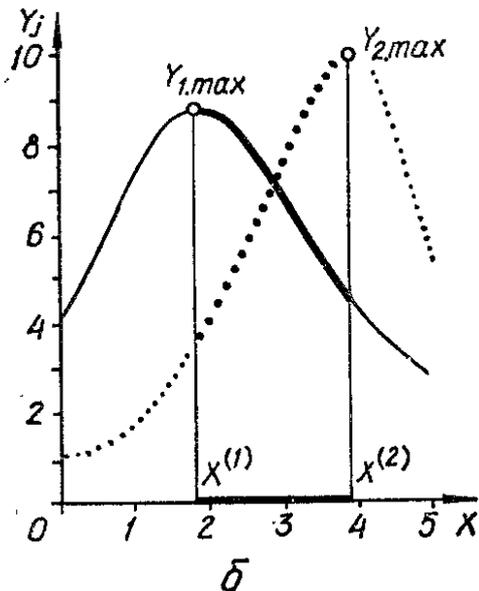
$$a_{3(4)} = b_4 = c_4, \quad (\text{В.9})$$

следовательно, все траектории приводят к равному результату. Однако при изменении нормы ресурса приоритеты меняются:

$$\left. \begin{array}{l} \text{если } X_{\text{норм}} = 3, \text{ то } a_3 > c_3 > b_3, \\ \text{если } X_{\text{норм}} = 2, \text{ то } b_2 > c_2 > a_2, \\ \text{если } X_{\text{норм}} = 1, \text{ то } b_1 > a_1 > c_1. \end{array} \right\} \quad (\text{В.10})$$



a



б

Рис. В.5. Простейшие типовые задачи оптимизации при трех траекториях достижения цели (а) и в ситуации (б), требующей компромиссного решения

Задача второго типа: при заданном уровне качества $Y_{тр} = 8$ по минимальному расходу ресурса три технологии ранжируются как

$$X^c \{8\} < X^a \{8\} < X^b \{8\}, \quad (B.11)$$

но при уменьшении требований до $Y_{тр} = 3$ приоритеты трансформируются:

$$X^b \{3\} < X^a \{3\} < X^c \{3\}. \quad (B.12)$$

Результаты анализа (B.9) — (B.12) отражают диалектическую природу управления технологической системой: *наилучшее* решение (траекторию управления) *выбрать нельзя* до тех пор, пока не определены цели и границы области допустимых действий как по X , так и по Y .

Если в задаче обусловлены два и более критерия оптимизации Y_j , то цель становится неопределенной [43]. В общем случае два критерия Y_1 и Y_2 (рис. В. 5, б) имеют оптимумы при разных расходах ресурса $X^{(1)}$ и $X^{(2)}$, поэтому можно определить лишь зону *компро-*

мисса (лат. *compromissum* — соглашение) между ними. Выбор решения в зоне $X^{(1)} \leq X \leq X^{(2)}$ не может быть формализован без введения дополнительных правил, например о приоритете критерия Y_1 над Y_2 или наоборот, об оценке потерь при выборе неоптимального расхода ресурса $X^{(i)}$ с помощью штрафов, о свертке критериев Y_1 и Y_2 в один с помощью новой функции $V = f(Y_j)$ и т. п. [9, 23, 43, 48].

Полезность применения математических моделей во всех областях человеческой деятельности резко возросла после расширения производства эффективных вычислительных машин, ставших сегодня необходимым инструментом решения задач науки и практики. Эти машины могут быть построены по двум разным принципам — по континуальному (лат. *continuum* — сплошное, непрерывное) и по дискретному (лат. *discretus* — разрыв).

Первый принцип реализуется в *аналоговой* (лат. *analogikos* — сходный, соответственный) вычислительной технике (АВТ). При ее создании и использовании *непрерывным* тепловым, механическим, гидродинамическим, электрическим и другим *процессам* [33, 52], проходящим на объекте O , ставятся в соответствие непрерывные процессы, проходящие, как правило, в электронной модели M . Это соответствие может быть основано [52] на простой аналогии, когда процессы описываются дифференциальными уравнениями одного вида, отличающимися для O и M лишь физическим смыслом постоянных коэффициентов. Аналогия может быть и сложной, когда процессы в АВТ непрерывно воспроизводят те математические операции (суммирование, интегрирование и т. п.), которые соответствуют математической модели процесса на объекте O , записанной в форме уравнений математической физики. АВТ имеет много достоинств при решении задач анализа и оптимизации динамических систем, в том числе в реальном масштабе времени. Однако для них характерны специфичность целевого назначения, ограниченность классов моделируемых систем, сравнительно невысокая точность результатов.

Универсальны цифровые вычислительные машины (ЦВМ), реализующие дискретный принцип представления и переработки информации в виде *конечных множеств цифр*. Если необходимо отразить непрерывный процесс, проходящий на объекте O , то он специально дискретизируется и представляется в виде последовательности цифр. Все задачи в ЦВМ решаются в результате проведения простейших арифметических и логических *операций* с цифрами. Быстродействие персональных компьютеров (микропроцессорных ЭВМ) достигает 1 млн. элементарных операций в секунду, что соответствует выполнению около 100 тыс. полноценных операций с числами с плавающей запятой, а для суперкомпьютеров уже получено быстродействие в 100 раз большее. Последовательность операций с цифрами управляется специальной *программой* — *инструкцией*, записанной на языке вычислительной машины, т. е. в конечном счете в цифровых кодах. Цифровые коды формируются транслятором, который переводит в них программу, составленную на одном из

языков программирования (ФОРТРАН, ПАСКАЛЬ, БЕЙСИК и др). Программа описывает алгоритм, т. е. точно определенное правило действий, для которого задано указание, как и в какой последовательности это правило необходимо применять к исходным данным задачи, чтобы ее решить.

Получение на ЭВМ с помощью конечного количества арифметических и логических действий результата в виде числа обуславливает применение к решению инженерных задач (а значит, и задач прикладной математики) специальных методов, отличных от классического анализа, оперирующего функциями непрерывных переменных. Эти методы создаются и изучаются быстро развивающимся разделом математики — *вычислительной математикой*, включающей вычисление значений функций, вычислительные методы линейной алгебры, численное решение уравнений, численное дифференцирование и интегрирование, численные методы оптимизации и др.

Общим для всех численных методов является *дискретизация исходной задачи*, которая решается за конечное число действий. Полученное решение дискретной задачи принимается за приближенное решение исходной задачи.

На рис. В. 6 показан пример применения простейших элементов численных методов к операциям над континуальной моделью в виде трансцендентной функции

$$Y = 11e^{-x} - 9 \ln X, \quad (\text{В.13})$$

для которой необходимо рассчитать значения функции с шагом $\Delta X = 0,2$ и определить с точностью 0,01 координату $X^{(0)}$ пересечения с осью абсцисс, т. е. корень уравнения $Y = 0$.

Функция (В.13) представляет собой алгебраическую сумму двух трансцендентных функций. Для каждой из них произведена дискретизация с заданным шагом, в результате чего непрерывная кривая $Y_i = f_i(X)$ превращена в набор отрезков Y_i в шести точках X ; числовое значение Y_i известно (взято, без потери общности, из таблиц функций). Значения функции (В.13) определены в заданных точках X как разность двух чисел $Y_1 - Y_2$. Ясно, что значение $Y = 0$ находится внутри того шага-отрезка ΔX , на концах которого Y имеет численные значения разных знаков — это отрезок $1,2 \leq X \leq 1,4$. Первое приближение к значению корня уравнения $Y = 0$ можно найти, используя известное правило пропорциональных частей (в физике «правило рычага»): если обозначить через $\Delta\{\Delta X\}$ расстояние от начала отрезка $X = 1,2$ до точки пересечения $Y = 0$, то

$$\left. \begin{aligned} 1,67/(1,67 + 0,30) &= \Delta\{\Delta X\}/0,2, \\ \Delta\{\Delta X\} &= 0,17; X\{Y = 0\} = 1,2 + 0,17 = 1,37. \end{aligned} \right\} \quad (\text{В.14})$$

Повторив операции определения Y по (В.14), можно найти, что в точке $X = 1,37$ значение функции равно $Y = -0,04$ вместо $Y = 0$. Если такая точность решения задачи удовлетворяет иссле-

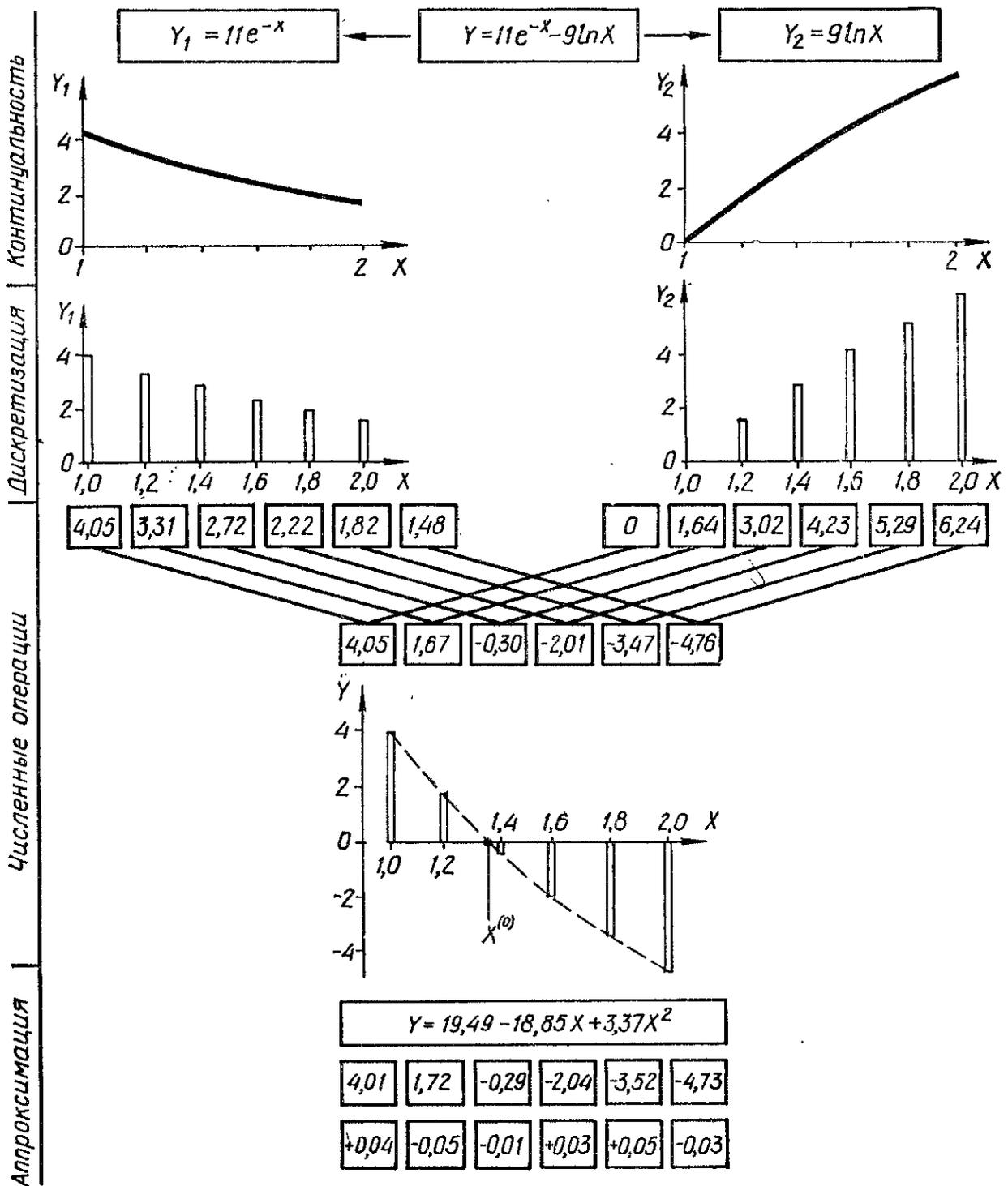


Рис. В.6. Принципиальная схема применения простейших элементов численных методов в операциях над континуальной моделью

дователя, то процедура прекращается, а если нет, то повторяется на новом отрезке $\Delta\{\Delta X\}$, т. е. при $1,2 \leq X \leq 1,37$. Такой процесс *последовательного, пошагового приближения* к решению с *заданной точностью*, основанный на результатах предшествующих шагов применения процедуры, носит название *итерационного* (лат. *iteratio* — повторение).

На рис. В. 6 приведена еще одна важная для инженеров материаловедов и технологов задача, решаемая на основе численных методов. При экспериментальном исследовании систем, как правило, исходная континуальная функция, в частности (В.13), *неизвестна*, а существует лишь информация о ее численных значениях

на отрезке (на рис. В.6 при $1 \leq X \leq 2$) — именно в таком виде представляются результаты испытаний образцов, измерения температур в точках аппарата и т. п. Эту информацию необходимо сжать в компактную форму, т. е. вернуться к континуальной модели. Но это будет не истинная функция (В.13), а ее функциональная модель типа M_3 в заданном диапазоне изменения X . Модель M_3 в виде квадратичного полинома на рис. В. 6 *аппроксимирует* (лат. *approximare* — приближаться) истинную функцию. Расхождение между истинной трансцендентной функцией и полиномом в шести точках не превышает $\pm 0,05$; если такая точность не удовлетворяет исследователя, то процесс аппроксимирующей модели будет продолжен (например, в направлении повышения степени полинома и т. п.).

I.1. Приближенные числа

В результате реализации численных методов на ЭВМ всегда будет получено приближенное по отношению к исходной задаче решение. Это приближение обусловлено погрешностями, воздействием комплекса объективных причин [4, 37 и др.].

Математические модели $M_1 - M_5$ (см. рис. В. 2) описывают реальный объект приближенно, поскольку истинное уравнение состояния системы, как правило, неизвестно; при этом параметры модели оцениваются по экспериментальным данным, определяемым приближенно с точностью, которая зависит от качества измерительной аппаратуры, организации исследований [4, 8]. Эти погрешности можно и нужно уменьшать до начала вычислительной работы с моделями, поскольку по отношению к численным методам решения задач эти погрешности являются неустранимыми. Их оценка необходима для ответа на вопрос о требуемой точности вычислительного процесса, так как назначать его погрешность намного меньше неустранимой погрешности входных данных нецелесообразно.

При замене исходной задачи дискретной, бесконечных математических процессов конечными (например, при разложении рядов) и т. п. возникает *погрешность численного метода*.

Конечная разрядность чисел, представленных в ЭВМ, и выполнение операций с такими приближенными числами приводят к *вычислительным погрешностям*, называемым также *погрешностями округления*.

Абсолютной погрешностью (или *ошибкой*) $\Delta\{a\}$ приближенного числа a называется разность между ним и точным значением числа \hat{a} , а *относительной погрешностью* $\delta\{a\}$ — отношение абсолютной погрешности к модулю точного числа:

$$\Delta\{a\} = |a - \hat{a}|, \quad (I.1)$$

$$\delta\{a\} = \Delta\{a\} / |\hat{a}|. \quad (I.2)$$

В практических задачах точное число \hat{a} , как правило, неизвестно, поэтому, во-первых, $\Delta\{a\}$ оценивают как верхний предел погрешности приближенного числа, базируясь на теоретических рассуждениях или опыте работы с аналогичными числами, во-вторых, заменяют в знаменателе (I.2) точное число на приближенное. Поэтому под $\Delta\{a\}$ и $\delta\{a\}$ понимают *предельные погрешности*, заведомо

превышающие истинные (в крайнем случае равные им). Информация о приближенном числе a записывается с одинаковым числом знаков после запятой в виде

$$\dot{a} = a \pm \Delta\{a\} = a(1 \pm \delta\{a\}). \quad (I.3)$$

Пример I.1. Плотность цементного клинкера $\rho = 3,1$ г/см³; предельная погрешность этого числа (по литературным данным) $\Delta\{\rho\} = 0,1$ г/см³, следовательно, точное значение плотности находится в интервале $\overset{\circ}{\rho} = 3,1 \pm 0,1$ г/см³ или $3 \leq \overset{\circ}{\rho} \leq 3,2$ г/см³. Относительная предельная погрешность $\delta\{\rho\} = 0,1/3,1 = 0,032$, или 3,2 %.

Значащими цифрами числа a называются все цифры в его записи начиная с первой ненулевой слева (ее десятичный разряд обозначается m). Значащие цифры могут быть верными и сомнительными. Понятие о верной значащей цифре связано со способом выполнения расчетов. При *ручных* вычислениях пользуются верными значащими цифрами в *узком смысле* — приближенное число имеет n верных значащих цифр, если погрешность $\Delta\{a\}$ не превышает половины соответствующего данной цифре $(m + 1 - n)$ -го десятичного разряда:

$$\Delta\{a\} \leq 0,5 \cdot 10^{m+1-n}. \quad (I.4)$$

При округлении числа a до n значащих цифр в узком смысле отбрасываются сомнительные цифры, стоящие справа от n -й значащей цифры. При этом последняя значащая цифра увеличивается на единицу, если первая из отбрасываемых цифр а) больше 5, б) равна 5, но среди следующих за ней цифр есть отличные от нуля, в) равна 5 и за ней следуют нули, а значащая цифра нечетная (правило образования четной цифры).

При *машинных* вычислениях отбрасываются при округлении a все сомнительные цифры, стоящие за n -й значащей цифрой без ее изменения. Поэтому она называется значащей в *широком смысле* и для нее выполняется требование

$$\Delta\{a\} \leq 1 \cdot 10^{m+1-n}. \quad (I.5)$$

Относительная погрешность $\delta\{a\}$ связана с числом n верхних значащих цифр в широком смысле и старшей значащей цифрой α_m соотношением

$$\delta\{a\} = \frac{\Delta\{a\}}{\alpha_m \cdot 10^m} = \frac{1 \cdot 10^{m+1-n}}{\alpha_m \cdot 10^m} = \frac{1}{\alpha_m \cdot 10^{n-1}}. \quad (I.6)$$

Пример I.2. Массовая доля связующего в полимерном композите $p = 0,375$ определена с $\Delta\{p\} = 0,008 = 0,8 \cdot 10^{-2}$ (количество значащих цифр — 3; первая ненулевая цифра имеет разряд «-1», поскольку $0,3 = 3 \cdot 10^{-1}$). Из соотношения

$$0,8 \cdot 10^{-2} > 0,5 \cdot 10^{-1+1-2} = 0,5 \cdot 10^{-2}$$

следует, что цифра 7 сомнительна в узком смысле (значит, сомнительна и цифра 5 в меньшем разряде). В то же время ясно, что цифра 7 верна в широком смысле, хотя цифра 5, как видно из соотношения

$$0,8 \cdot 10^{-2} > 1 \cdot 10^{-1+1-3} = 0,1 \cdot 10^{-2},$$

сомнительна и в широком смысле. Округление при ручном счете дает приближенное число $p = 0,38$, а при машинном счете — $p \approx 0,37$. Относительная ошибка $\delta \{p\}$ (I.6) в первом варианте $1 / (8 \cdot 10^2 - 1) = 0,0125$, во втором — $1 / (7 \cdot 10^2 - 1) = 0,0143$.

При вычислении на ЭВМ в режиме с плавающей запятой округленные числа имеют определенное количество значащих цифр, следовательно, должна контролироваться относительная погрешность округления. В режиме с фиксированной запятой округленные числа имеют определенное количество разрядов после запятой, т. е. контролируется абсолютная погрешность округления.

При выполнении математических операций с приближенными числами в результате получается приближенное число, которое можно рассматривать как функцию $y = f(X_1, \dots, X_i, \dots, X_k)$, определенную с погрешностью $\Delta \{y\}$ по k известным с погрешностью $\Delta \{X_i\}$ аргументам X_i^* . Истинная погрешность функции как разность функций от приближенных X_i и истинных аргументов $\overset{\circ}{X}_i$

$$\Delta \{y\} = f(\overset{\circ}{X}_1, \dots, \overset{\circ}{X}_i, \dots, \overset{\circ}{X}_k) - f(X_1, \dots, X_i, \dots, X_k) \quad (I.7)$$

с учетом $\overset{\circ}{X}_i = X_i + \Delta \{X_i\}$ имеет вид

$$\Delta \{y\} = f(X_1 + \Delta \{X_1\}, \dots, X_i + \Delta \{X_i\}, \dots, X_k + \Delta \{X_k\}) - f(X_1, \dots, X_i, \dots, X_k). \quad (I.8)$$

Поскольку погрешности $\Delta \{X_i\}$ малы по сравнению со значениями аргументов X_i , то к первому слагаемому в (I.8) можно применить разложение в ряд Тейлора (см. прил. I.3) в окрестностях точки $X_{i,0}$, ограничиваясь только членами первого порядка и заменяя дифференциалы dX_i на приращения ΔX_i :

$$\begin{aligned} \Delta \{y\} &= f(X_1, \dots, X_i, \dots, X_k) + \left(\frac{\partial y}{\partial X_1}\right)_0 \Delta X_1 + \dots + \\ &+ \left(\frac{\partial y}{\partial X_i}\right)_0 \Delta X_i + \dots + \left(\frac{\partial y}{\partial X_k}\right)_0 \Delta X_k - f(X_1, \dots, X_i, \dots, X_k) = \\ &= \left(\frac{\partial y}{\partial X_1}\right)_0 \Delta X_1 + \dots + \left(\frac{\partial y}{\partial X_i}\right)_0 \Delta X_i + \dots + \left(\frac{\partial y}{\partial X_k}\right)_0 \Delta X_k = \\ &= \sum_{i=1}^k \left(\frac{\partial y}{\partial X_i}\right)_0 \Delta X_i. \end{aligned} \quad (I.9)$$

Относительная погрешность функции с учетом $d(\ln y) = (dy)/y$ на основании (I.9) определяется как

$$\delta \{y\} = \frac{\Delta \{y\}}{y} = \sum_{i=1}^k \frac{1}{y} \left(\frac{\partial y}{\partial X_i}\right)_0 \Delta X_i = \sum_{i=1}^k \left(\frac{\partial (\ln y)}{\partial X_i}\right)_0 \Delta X_i. \quad (I.10)$$

*) Здесь и далее в п. I.1 подразумевается, что аргумент X_i может быть как детерминированной величиной, так и случайной (см. прил. I.4); для случайных аргументов под погрешностью $\Delta \{X_i\}$ понимаются их дисперсии $\sigma^2 \{X_i\}$, характеризующие рассеяние случайной величины по отношению к среднему значению (прил. I. 4 и п. III. 3).

На основе общих формул погрешностей функции легко получить выражение для оценки погрешностей любых математических операций. Так, для алгебраической суммы по (I.9) будет получена абсолютная погрешность функции как сумма абсолютных погрешностей аргументов

$$\Delta \{y\} = \Delta \{X_1\} + \dots + \Delta \{X_i\} + \dots + \Delta \{X_k\} = \sum_{i=1}^k \Delta X_i, \quad (\text{I.12})$$

что дает относительную погрешность

$$\delta \{y\} = \Delta \{y\} / |y| = \sum_{i=1}^k \Delta X_i / \left| \sum_{i=1}^k X_i \right|, \quad (\text{I.13})$$

а для произведения

$$y = (X_1 \dots X_i) (X_{i+1} \dots X_k)^{-1} \quad (\text{I.14})$$

по (I.10) будет получена относительная погрешность функции как сумма относительных погрешностей аргументов

$$\begin{aligned} \delta \{y\} &= \left(\frac{1}{X_1} \right) \Delta X_1 + \dots + \left(\frac{1}{X_i} \right) \Delta X_i + \left(\frac{1}{X_{i+1}} \right) \Delta X_{i+1} + \\ &+ \dots + \left(\frac{1}{X_k} \right) \Delta X_k = \delta \{X_1\} + \dots + \delta \{X_i\} + \delta \{X_{i+1}\} + \\ &+ \dots + \delta \{X_k\} = \sum_{i=1}^k \delta \{X_i\}, \end{aligned} \quad (\text{I.15})$$

что дает абсолютную погрешность

$$\Delta \{y\} = \delta \{y\} |y| = \left(\sum_{i=1}^k \delta \{X_i\} \right) (X_1 \dots X_i) / (X_{i+1} \dots X_k). \quad (\text{I.16})$$

Для некоторых однофакторных функций погрешности приведены в табл. I.1 [2, 4]. При оценке погрешностей функций, состоящих из ряда математических операций, к ним последовательно применяются соотношения, полученные для погрешностей простейших действий над приближенными числами. Из вышеизложенного ясно, что любые операции с приближенными числами (как с детерминированными, так и со случайными) ведут к накоплению погрешностей.

Пример I.3. А. Предел прочности бетона на сжатие после длительного пребывания образцов в агрессивной среде составил $R^a = 25,2$ МПа при $\Delta R^a = 1,3$ МПа, а для бетона того же состава, находившегося в нормальных условиях, $R^H = 31,2$ МПа при $\Delta R^H = 1,2$ МПа. Коэффициент стойкости рассчитывается как отношение $K_{ст} = R^a / R^H = 25,2 / 31,2 = 0,808$. Погрешность числового значения этого коэффициента, согласно (I.15) и (I.16),

$$\begin{aligned} \delta \{K_{ст}\} &= \delta \{R^a\} + \delta \{R^H\} = 1,3/25,2 + 1,2/31,2 = 0,05 + 0,04 = 0,09; \\ \Delta \{K_{ст}\} &= 0,09 \cdot 0,808 = 0,072. \end{aligned}$$

Верной значащей цифрой в широком смысле является $8 \cdot 10^{-1}$, значит, результат следует отразить как $K_{ст} = 0,8 \pm 0,07$.

Таблица 1.1. Погрешности однофакторных функций ($c = \text{const}$)

Функция $y = f(X)$	Абсолютная ошибка $\Delta\{y\}$		Относительная ошибка $\delta\{y\}$	
	от $\Delta\{X\}$	от $\delta\{X\}$	от $\Delta\{X\}$	от $\delta\{X\}$
X^c	$cX^{(c-1)} \Delta\{X\}$	$cX^c \delta\{X\}$	$c \frac{1}{X} \Delta\{X\}$	$c \delta\{X\}$
$\frac{1}{X^c}$	$\frac{c}{X^{c+1}} \Delta\{X\}$	$\frac{c}{X^c} \delta\{X\}$	$c \frac{1}{X} \Delta\{X\}$	$c \delta\{X\}$
$\frac{1}{X}$	$\frac{1}{X^2} \Delta\{X\}$	$\frac{1}{X} \delta\{X\}$	$\frac{1}{X} \Delta\{X\}$	$\delta\{X\}$
$\sqrt[c]{X}$	$\frac{c\sqrt[c]{X}}{cX} \Delta\{X\}$	$\frac{c\sqrt[c]{X}}{c} \delta\{X\}$	$\frac{1}{cX} \Delta\{X\}$	$\frac{1}{c} \delta\{X\}$
c^X	$c^X \ln c \Delta\{X\}$	$Xc^X \ln c \delta\{X\}$	$\ln c \Delta\{X\}$	$X \ln c \delta\{X\}$
e^X	$e^X \Delta\{X\}$	$Xe^X \delta\{X\}$	$\Delta\{X\}$	$X \delta\{X\}$
$\log_c X$	$\frac{1}{X \ln c} \Delta\{X\}$	$\frac{1}{\ln c} \delta\{X\}$	$\frac{\Delta\{X\}}{X \ln c \log_e X}$	$\frac{\delta\{X\}}{\ln c \log_e X}$
$\lg X$	$0,434 \frac{\Delta\{X\}}{X}$	$0,434 \delta\{X\}$	$\frac{0,434 \Delta\{X\}}{X \lg X}$	$\frac{0,434 \delta\{X\}}{\lg X}$
$\ln X$	$\frac{1}{X} \Delta\{X\}$	$\delta\{X\}$	$\frac{1}{X \ln X} \Delta\{X\}$	$\frac{1}{\ln X} \delta\{X\}$

Б. При расчете состава бетона по методу абсолютных объемов расход песка определяется по соотношению, учитывающему плотности компонентов бетона:

$$\Pi = \left[1000 - \left(\frac{\text{Ц}}{\rho_{\text{ц}}} + \text{В} + \frac{\text{Щ}}{\rho_{\text{щ}}} \right) \right] \rho_{\text{п}}. \quad (1.17)$$

Если значения ρ_i , кг/дм³, оценены с предельной погрешностью как $\rho_{\text{ц}} = 3,1 \pm 0,01$, $\rho_{\text{щ}} = 2,6 \pm 0,01$ и $\rho_{\text{п}} = 2,63 \pm 0,01$, а массовые расходы, кг/м³, компонентов как $\text{Ц} = 240 (1 \pm 0,01)$, $\text{В} = 130 (1 \pm 0,01)$ и $\text{Щ} = 1420 (1 \pm 0,02)$, то предельная погрешность числового значения Π ,

$$\Pi = \left[1000 - \left(\frac{240}{3,1} + 130 + \frac{1420}{2,6} \right) \right] 2,63 = 246,43 \cdot 2,63 = 648 \text{ кг/м}^3,$$

определяется так:

$$\Delta \left\{ \frac{\text{Ц}}{\rho_{\text{ц}}} \right\} = \left(\frac{\text{Ц}}{\rho_{\text{ц}}} \right) \left(\delta\{\text{Ц}\} + \frac{\Delta\rho_{\text{ц}}}{\rho_{\text{ц}}} \right) = \left(\frac{240}{3,1} \right) \left(0,01 + \frac{0,01}{3,1} \right) = 1,02,$$

$$\Delta \left\{ \frac{\text{Щ}}{\rho_{\text{щ}}} \right\} = \left(\frac{\text{Щ}}{\rho_{\text{щ}}} \right) \left(\delta\{\text{Щ}\} + \frac{\Delta\rho_{\text{щ}}}{\rho_{\text{щ}}} \right) = \left(\frac{1420}{2,6} \right) \left(0,02 + \frac{0,01}{2,6} \right) = 13,02,$$

$$\Delta \left\{ \left[1000 - \left(\frac{\Pi}{\rho_{\text{ц}}} + B + \frac{\text{Ш}}{\rho_{\text{ш}}} \right) \right] \right\} = 0 + 1,02 + 130 \cdot 0,01 + 13,02 = 15,34,$$

$$\Delta \Pi = 648 \left(\frac{15,34}{246,43} + \frac{0,01}{2,63} \right) = 42,8,$$

следовательно, $\Pi = 650 \pm 43$ кг/м³; относительная погрешность $\delta\{\Pi\} = 42,8/648 = 0,066$ достаточно высока, поскольку значение Π определяется как разность, меньшая по величине, чем вычитаемое.

В. Коэффициент теплоотдачи α газа к поверхности зернового материала определяется через критерий Нуссельта $\text{Nu} = \alpha d_3 / \lambda_r$ (где d_3 — диаметр зерен слоя, λ_r — коэффициент теплопроводности газа) с помощью [52] критерияльного уравнения

$$\text{Nu} = 0,286 \text{Re}^{0,8}, \quad (\text{I.18})$$

содержащего критерий Рейнольдса $\text{Re} = \omega d_3 / \nu_r$ (где ω — скорость газа, отнесенная к полному сечению аппарата, ν_r — кинематическая вязкость газа) и две эмпирические константы $A = 0,286 \pm 0,001$ и $B = 0,8 \pm 0,02$.

Если исходная информация определена с предельной погрешностью как

$$d_3 = 0,5 (1 \pm 0,05) \text{ мм}, \quad \lambda_r = 0,0395 (1 \pm 0,001) \text{ Вт/(м} \cdot \text{К)},$$

$$\omega = 0,5 (1 \pm 0,05) \text{ м/с}, \quad \nu_r = 0,0349 \cdot 10^{-3} (1 \pm 0,02) \text{ м}^2/\text{с},$$

то погрешности составляют:

при расчете критерия Рейнольдса

$$\text{Re} = 0,5 \cdot 0,5 \cdot 10^{-3} / (0,0349 \cdot 10^{-3}) = 7,16,$$

$$\delta\{\text{Re}\} = 0,05 + 0,05 + 0,02 = 0,12;$$

при расчете критерия Nu с учетом (I.10) и табл. I.1

$$\text{Nu} = 0,286 \cdot 7,16^{0,8} = 1,38,$$

$$\delta\{\text{Nu}\} = \delta\{A\} + B\delta\{\text{Re}\} + \ln \text{Re} \cdot \Delta\{B\} =$$

$$= 0,001 / 0,286 + 0,8 \times 0,12 + \ln 7,16 \cdot 0,02 = 0,139;$$

при расчете коэффициента теплоотдачи

$$\alpha = \text{Nu} \lambda_r / d_3 = 1,38 \cdot 0,0395 / (0,5 \cdot 10^{-3}) = 109,$$

$$\delta\{\alpha\} = 0,139 + 0,001 + 0,05 = 0,19,$$

$$\Delta\{\alpha\} = 109 \cdot 0,148 = 20,7.$$

Расчетное значение $\alpha = 110 \pm 21$ Вт/(м² · К).

Зависимости (I.9) — (I.10) должны использоваться не только для прямого расчета абсолютной и относительной погрешностей функции по известным значениям погрешностей аргументов $\Delta\{X_i\}$, но и для решения обратной задачи — для определения допустимых погрешностей аргументов по заданной величине $\Delta\{y\}$ или $\delta\{y\}$ (см. подробно в [2]). Это позволяет инженеру выбрать необходимую измерительную аппаратуру, схему организации эксперимента и вычислений и т. п.

1.2. Приближенные решения

В инженерной практике одним из главных элементов содержания многих задач является отыскание на основе экспериментально-теоретической информации тех или иных соотношений между факторами X и выходными параметрами системы Y_j с целью оценки по этим соотношениям тех сочетаний X с Y_j , для которых отсутствие

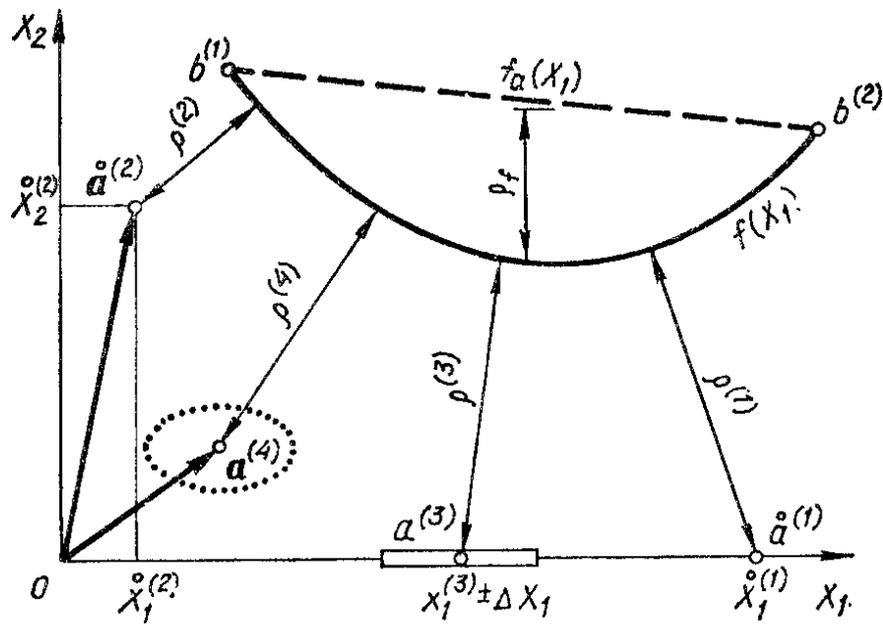


Рис. I.1. Графические образы точного $\overset{\circ}{a}^{(1)}$ и приближенного $\overset{\circ}{a}^{(3)}$ чисел, точного $\overset{\circ}{a}^{(2)}$ и приближенного $\overset{\circ}{a}^{(4)}$ векторов, расстояний ρ_i от точек a_i до кривой $b^{(1)}$, $b^{(2)}$ и нормы ρ_f , аппроксимирующей эту кривую прямой

вала исходная информация. Поскольку экспериментально-теоретическая информация, как правило, задается и (или) обрабатывается на ЭВМ в виде множества приближенных чисел, то, естественно, что и конечные решения будут приближенными причем погрешность этого приближения должна быть оценена и, по возможности, минимизирована.

Точная скалярная величина $\overset{\circ}{a}^{(1)}$ отражается точкой на числовой оси X_1 (рис. I.1), а точная векторная величина отражается точкой $\overset{\circ}{a}^{(2)}$ с точными координатами $X_1^{(2)}$ и $X_2^{(2)}$. Вектор можно охарактеризовать обобщающей числовой величиной — его длиной от начала координат $\|X\| > 0$ ($\|X\| = 0$ только при $X = 0$), называемой нормой (прил. I.1). При этом норма вектора в k -мерном пространстве может быть построена по разным принципам. При суммировании абсолютных значений координат получается *октаэдральная* норма

$$\|X\|_1 = \sum_{i=1}^k |X_i|, \quad (I.19)$$

которая для ситуации на рис. I.1 приобретает вид

$$\|X\|_1 = |X_1^{(2)}| + |X_2^{(2)}|. \quad (I.20)$$

При суммировании квадратов координат, что исключает необходимость перехода к абсолютным числам, получается *сферическая* (квадратичная, эвклидова) норма

$$\|X\|_2 = \left(\sum_{i=1}^k X_i^2 \right)^{1/2}, \quad (I.21)$$

$$\|X\|_2 = \sqrt{(X_1^{(2)})^2 + (X_2^{(2)})^2}. \quad (I.22)$$

При выборе в качестве меры длины наибольшей абсолютной координаты, так называемой *мажоранты*, будет получена *равномерная* (кубическая) норма

$$\|X\|_{\infty} = \max_{1 \leq i \leq k} |X_i|, \quad (I.23)$$

которая применительно к рис. I.1 имеет вид

$$\|X\|_{\infty} = |X_1^{(2)}|. \quad (I.24)$$

Выбор той или иной формы нормы обуславливается особенностями формулировки и решения математической задачи.

Приближенное число $a^{(3)}$ на числовой оси отражается (рис. I.1) уже не точкой, а отрезком с серединой $X_1^{(3)}$ и границами: нижней $X_1^{(3)} - \Delta\{X_1\}$ и верхней $X_1^{(3)} + \Delta\{X_1\}$. Конец приближенного вектора отражается координатами $X_1^{(4)} \pm \Delta X_1$ и $X_2^{(4)} \pm \Delta X_2$, следовательно, он может попасть не в точку $a^{(4)}$, а в окрестность точки, ограниченную (без потери общности) эллипсом с главными радиусами $\Delta\{X_1\}$ и $\Delta\{X_2\}$. Естественно, что приближенно будет определена и норма такого вектора.

Расстояние между элементами такого множества, как концы векторов $a^{(1)}, a^{(2)}, a^{(3)}, \dots$ в k -мерном пространстве, называется *метрикой*. Она определяется как норма разности

$$\rho(\overset{\circ}{a}^{(i)}, \overset{\circ}{a}^{(j)}) = \|X^{(i)} - X^{(j)}\|. \quad (I.25)$$

В частном случае для точек $\overset{\circ}{a}^{(1)}$ и $\overset{\circ}{a}^{(2)}$ на рис. I.1 квадратичная метрика, нередко обозначаемая L_2 , определяется так:

$$L_2(\overset{\circ}{a}^{(1)}, \overset{\circ}{a}^{(2)}) = + \sqrt{(X_1^{(2)} - X_1^{(1)})^2 + (X_2^{(2)})^2}. \quad (I.26)$$

Метрика используется как характеристика степени близости не только между дискретными точками $a^{(i)}$ или точкой и кривой $X_2 = f(X_1)$ (см., например, $\rho^{(2)}$ на рис. I.1), но и между двумя функциями $Y = f(X_1)$ и $Y_a = f_a(X_1)$. В этом случае каждая из функций рассматривается как точка функционального пространства, в котором, например, равномерная метрика определится по наибольшему расстоянию между функциями на заданном отрезке:

$$\|f(X_1) - f_a(X_1)\| = \max |f(X_1) - f_a(X_1)|. \quad (I.27)$$

Так, приложение (I.27) к ситуации на рис. I.1, когда функция $f(X) = X_2 = f(X_1)$ на отрезке $[b^{(1)}, b^{(2)}]$ заменяется прямой $f_a(X) = X_2 = c_0 + c_1 X_1$, дает в качестве метрики расстояние

$$\rho_f = \|f(X_1) - f_a(X_1)\| = \max |f(X_1) - f_a(X_1)|. \quad (I.28)$$

Функция $f_a(X_1)$ *аппроксимирует* на отрезке $[b^{(1)}, b^{(2)}]$ функцию $f(X_1)$. При аппроксимации роль *абсолютной погрешности* играет норма разности $\|f(X) - f_a(X)\|$, обозначаемая $\|\Delta(X)\|$, а величина $\|\Delta\{X\}\| / \|f\{X\}\|$ — роль *относительной погрешности*.

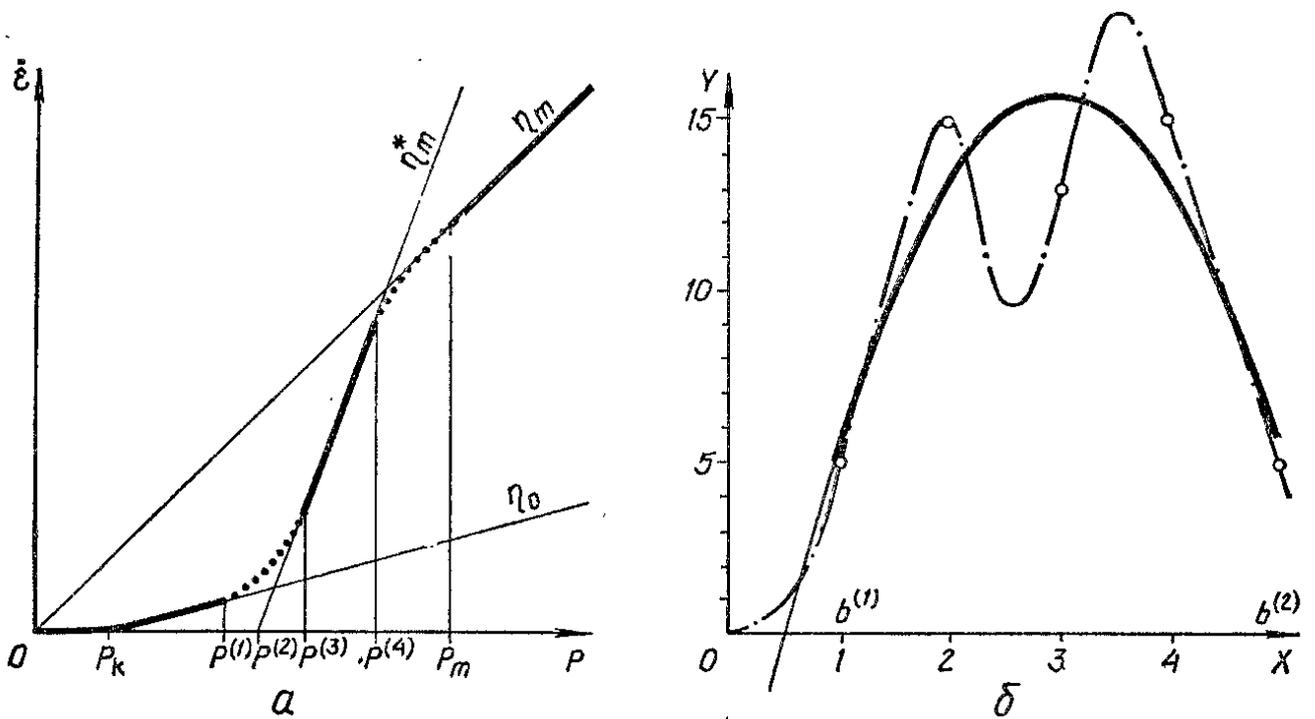


Рис. 1.2. Аппроксимация реологической кривой прямыми (а) и экспериментальных точек (б) — параболой

Замена точной функции на приближенную необходима для упрощения как вычислительных операций, так и принятия инженерных решений (включая интерпретацию результатов моделирования) в ограниченной области действий.

Пример 1.4. Полная реологическая кривая на рис. 1.2, а — функция «напряжение P , Па, — скорость деформации тела $\dot{\epsilon}$, c^{-1} », отражающая весьма сложное поведение (течение) под нагрузкой структурированного незатвердевшего материала, аппроксимируется набором прямых

$$P = C_0 + C_1 \dot{\epsilon}.$$

В этом семействе аппроксимирующих линейных функций коэффициент пропорциональности C_1 , Па · с, интерпретируется как динамическая вязкость неньютоновской жидкости [6, 13, 52] с разной степенью деструкции. При малых нагрузках до предельного напряжения P_k скорость деформаций равна нулю (т. е. как в упругом теле деформации ϵ пропорциональны P , а вязкость «бесконечна»). На участке от P_k до $P^{(1)}$ течение происходит при высокой постоянной начальной вязкости η_0 и описывается уравнением

$$P = P_k + \eta_0 \dot{\epsilon}.$$

Следует отметить, что так называемая «эффективная» вязкость $\eta = P/\dot{\epsilon}$ будет при этом непрерывно уменьшаться. При $P > P^{(3)}$ структура тела начинает интенсивно разрушаться, эффективная вязкость продолжает падать, а закономерность течения описывается уравнением

$$P = P^{(2)} + \eta_m^* \dot{\epsilon},$$

где $\eta_m^* < \eta_0$ — «пластическая» вязкость. При высоких напряжениях структура полностью разрушена, и жидкость начинает вести себя как ньютоновская при $\eta = \eta_m$, что описывается так:

$$P = \eta_m \dot{\epsilon}.$$

Если использовать линейную аппроксимацию на рис. 1.2, *a* в пределах, расширенных до точек пересечения прямых, то будет получена четырехзвенная ломаная, или *сплайн-функция*.

Среди семейства аппроксимирующих кривых выбирается та наилучшая функция $f_a(X)$, которая *минимизирует* в заданной области меру $\|\Delta\{X\}\|$. Если величина $\|\Delta\{X\}\|$ окажется меньше заданного малого числа ε , которое является некоторым *нормативом точности аппроксимации*, назначенным из инженерных соображений, то функция $f_a(X)$ принимается для дальнейшего использования в задачах материаловедения и технологии; если же $\min \|\Delta\{X\}\| > \varepsilon$, то необходимо расширить семейство функций, используемых для аппроксимации. Естественно, что выход за заданную область не гарантирует, что выбранная функция сохранит свои лучшие свойства. Так, достаточно точные аппроксимирующие функции на рис. 1.2, *a* при выходе за границы $P^{(i)}$ могут давать физически бессмысленные результаты вплоть до отрицательной скорости деформаций $\dot{\varepsilon} < 0$.

Особый интерес для инженеров-материаловедов и технологов представляет аппроксимация *непрерывной* функцией $f_a(X)$ *дискретного* множества точек (рис. 1.2, *б*). Это конечное множество обычно отражает в K -мерном пространстве результаты N экспериментов с k -факторной ($k = K - 1$) системой $y_u = f(X_{1u}, X_{2u}, \dots, X_{ku})$. Необходимо в условиях неполной информации об истинной функции $f(X)$ найти такую аппроксимирующую функцию $f_a(X)$, чтобы норма $\|f(X) - f_a(X)\|$ минимизировалась. Наиболее часто для этой цели используется квадратичная норма $\|\Delta\{X\}\|_2 \rightarrow \min$, поэтому такой способ аппроксимации получил название *метода наименьших квадратов* — МНК (см. п. III.7). Задачу замены дискретного множества на аналитическую функцию можно считать обратной к вычислению значений функции в N заданных точках, причем численные методы служат главным образом для оценки числовых значений параметров выбранной для восстановления функции.

Общий вид аппроксимирующей функции — модели — выбирается на основе инженерных предпосылок по неформальным процедурам. Так, для аппроксимации пяти точек на рис. 1.2, *б* выбрана простейшая форма — парабола $Y = b_0 + b_1 X_1 + b_{11} X_1^2$, для которой по МНК на отрезке $1 \leq X \leq 5$ определены оценки параметров (МНК-оценки): $b_0 = -7,4$; $b_1 = 15,4$, $b_{11} = -2,6$. Следует отметить, что «хорошее» приближение аппроксимирующей функции $f_a(X)$ к N точкам, вообще говоря, еще не гарантирует того, что «хорошим» в смысле $\|\Delta\{X\}\| < \varepsilon$ будет приближение к неизвестной истинной функции $f(X)$, если она примет значения X , отличные от координат аппроксимируемых точек. На рис. 1.2, *б* такая ситуация возникает, когда оказывается истинной функция, нанесенная штрихпунктиром. Однако абсолютное большинство функций, связывающих входные факторы X и выходные параметры Y_j в областях, характерных для реальных материаловедческих и технологических систем (см. «Введение» рис. В.1 и далее), не имеет таких неприятных особенностей.

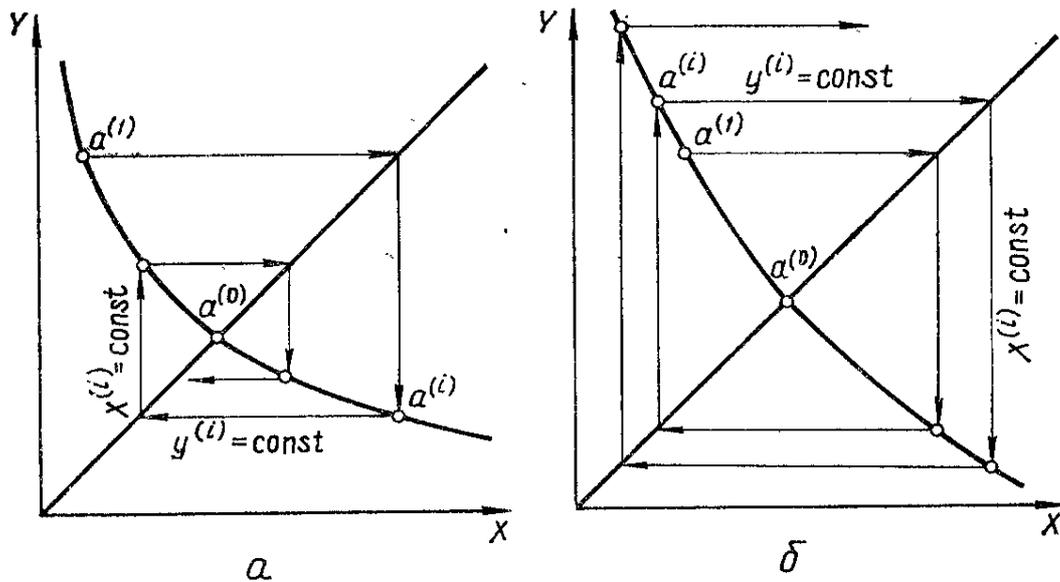


Рис. 1.3. Итерационные процессы:
 а — сходящийся; б — расходящийся

Функциональные зависимости (в абсолютном большинстве материаловедческих и технологических задач аппроксимирующие, а не точные) используются для принятия инженерных решений. Так, одна из типовых инженерных задач (см. п. III.15) — определение значений аргумента X , при котором функция $f(X)$ достигает заданного значения C , что равносильно определению корней уравнения $F(X) = f(X) - C = 0$.

Многие численные методы основаны на *последовательном приближении* к точному решению $X^{(0)}$, причем каждый последующий s -й шаг основан на *результате* $X^{(s-1)}$ *предыдущего*. Этот итерационный процесс полезен в том случае, если после каждого шага расстояние $\rho(X^{(s)}, X^{(0)})$ уменьшается, т. е. пределом $X^{(s)}$ является $X^{(0)}$. Такой итерационный процесс называется *сходящимся* и его можно прервать, когда изменение расстояний $[\rho(X^{(s-1)}, X^{(s)}) - \rho(X^{(s)}, X^{(s+1)})]$ станет меньше заданной величины ε .

На рис. 1.3, а показан *сходящийся процесс* приближения от произвольной точки $a^{(1)}$ к точке $a^{(0)}$ пересечения прямой $Y = X$ и кривой $Y = f(X)$, заключающийся в поэтапном движении от точки на кривой при $y^{(i)} = \text{const}$ до прямой и далее от прямой к кривой при $X^{(i)} = \text{const}$ и т. д. На рис. 1.3, б процесс, проводимый по аналогичному алгоритму сходимостью не обладает. Этот алгоритм используется, в частности, при определении корней уравнения (см. п. I.3), поэтому в последнем случае решение получено не будет.

На основе вышеизложенного следует заключить, что *инженерное решение* \hat{X} , *базирующееся* на эмпирических данных, будет получено приближенно с *погрешностью*, состоящей из $\Delta_{\text{в.т.}}\{X\}$ — погрешности, связанной с накоплением ошибки округления при использовании вычислительной техники; $\Delta_{\text{в.а}}\{X\}$ — погрешности сходящегося вычислительного алгоритма численного метода решения задачи по математической модели; $\Delta_{\text{а}}\{X\}$ — погрешности построения аппроксимирующей модели [6]; $\Delta_{\text{э}}\{X\}$ — погрешности исходной

экспериментальной информации, на основе которой построена модель [2, 4, 6].

Общим требованием является получение приближенного решения с допустимой, по условиям инженерной задачи, погрешностью $\Delta_{\text{доп}}\{X\}$, т. е. обеспечение

$$\Delta\{\hat{X}\} = \|\Delta_{\text{в.т}}\{X\}\| + \|\Delta_{\text{в.а}}\{X\}\| + \|\Delta_{\text{а}}\{X\}\| + \|\Delta_{\text{э}}\{X\}\| \leq \Delta_{\text{доп}}\{X\}. \quad (1.29)$$

Прямая оценка полной погрешности по (1.29) принципиальной трудности не представляет, поскольку допустимо назначать предельные (мажорантные — см. (1.23) — (1.24)) погрешности слагаемых, хотя в этом случае значение $\Delta\{X\}$ может оказаться сильно завышенным (лучше оперировать статистическими оценками слагаемых — см. п. III.3 и далее). Значительно сложнее обратная задача о назначении оптимального соотношения между слагаемыми и минимизации каждого из них. Она рассматривается в тех разделах научно-технического знания (вычислительная математика, математическая теория эксперимента, приборостроение и т. п.), где формируются средства решения определенного этапа инженерной задачи.

1.3. Методы решения уравнений

Во многих приложениях возникает задача отыскания *корней алгебраических и трансцендентных уравнений*. Если известна математическая модель поведения объекта (см. «Введение»)

$$Y - \psi(X) = f(X) = 0 \quad (1.30)$$

как функция одной переменной, то зачастую требуется знать значение $X^{(0)}$, при котором $\psi(X)$ достигает заданного уровня b , т. е. нужно найти X из уравнения $b - \psi(X) = 0$ или $\psi(X) = b$. Подстановка значения корня уравнения $X = X^{(0)}$ превращает уравнение модели объекта (1.30) в тождество.

Отыскание корней уравнений, в которых функция f имеет вид полинома не выше четвертой степени, не представляет затруднений. Корни полиномов до четвертой степени включительно можно выразить через конечное число арифметических операций, включая извлечение корня [21]. Однако в большинстве практических ситуаций функция f не сводится к указанному случаю. Определить корни уравнения (1.30) точно удастся лишь в некоторых частных случаях, а если учесть, что коэффициенты в модели являются приближенными числами, полученными при решении практических задач, то постановка задач отыскания точных значений $X^{(0)}$ лишена смысла.

Вычислительные алгоритмы приближенного определения корней таких уравнений основаны на том, что известное заранее некоторое значение X^0 (начальное приближение) одного из корней с помощью итерационной процедуры последовательно улучшается до

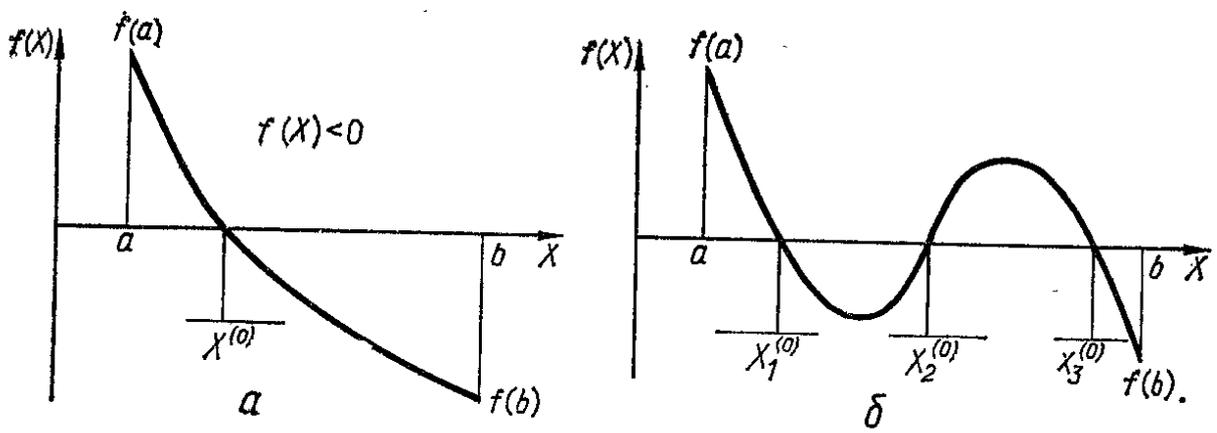


Рис. 1.4. Графики функций с единственным (а) и несколькими (б) корнями

заданной точности нахождения корня. Как правило, алгоритмы решения нелинейных уравнений разбиваются на два этапа: 1) отделение корней уравнения — отыскание достаточно малого отрезка $[a; b]$, принадлежащего области определения X , в которой заключен один из корней; 2) этап уточнения приближенного значения корня.

Первый этап в общем случае плохо формализуем, поэтому для отделения корней часто пользуются качественными методами исследования уравнений.

Пусть требуется найти приближенное значение действительного корня уравнения (1.30), где функция $f(X)$ определена и непрерывна на числовом множестве $\{X\}$. Говорят, что корень $X^{(0)}$ уравнения $f(X) = 0$ отделен на отрезке $[a; b]$, если $X^{(0)} \in [a; b]$, и на этом отрезке нет других корней.

Поиск отрезка $[a; b]$ основан на том, что если непрерывная функция $f(X)$ на концах отрезка $[a; b] \in \{X\}$ принимает значения разных знаков, то внутри этого отрезка содержится, по крайней мере, один корень уравнения $f(X) = 0$. Данное правило не дает ответа на вопрос о количестве корней на отрезке $[a; b]$. Однако, если положить, что всюду в $[a; b]$ первая производная не равна нулю ($f'(X) \neq 0$), то число корней на отрезке определяется следующим условием: если $f'(X)$ на отрезке $[a; b]$ не изменяет знака и при этом выполняется условие $f(a)f(b) < 0$, то $f(X)$ имеет на $[a; b]$ единственный корень $X^{(0)}$ (рис. 1.4, а). На рис. 1.4, б показана функция с несколькими корнями на $[a; b]$; в таких случаях целесообразно разбить $[a; b]$ на частные интервалы соответственно интервалам монотонности функции $f(X)$ и провести аналогичный анализ.

Пример 1.5. Отделить корни уравнения

$$f(X) = X^3 - 2X. \quad (1.31)$$

Областью определения X является $\{X\} =]-\infty; \infty[$. Из равенства производной нулю

$$f'(X) = 3X^2 - 2 = 0$$

следуют критические точки $X_1^{кр} = -0,82$ и $X_2^{кр} = 0,82$. Интервалы монотонности: $]-\infty; -0,82]$; $[-0,82; 0,82]$; $[0,82; +\infty[$.

Исследуются знаки функции на концах интервалов монотонности:

$$\lim_{X \rightarrow -\infty} f(X) = -\infty; \quad f(-0,82) = +1,1; \quad f(+0,82) = -1,1; \quad \lim_{X \rightarrow +\infty} f(X) = +\infty.$$

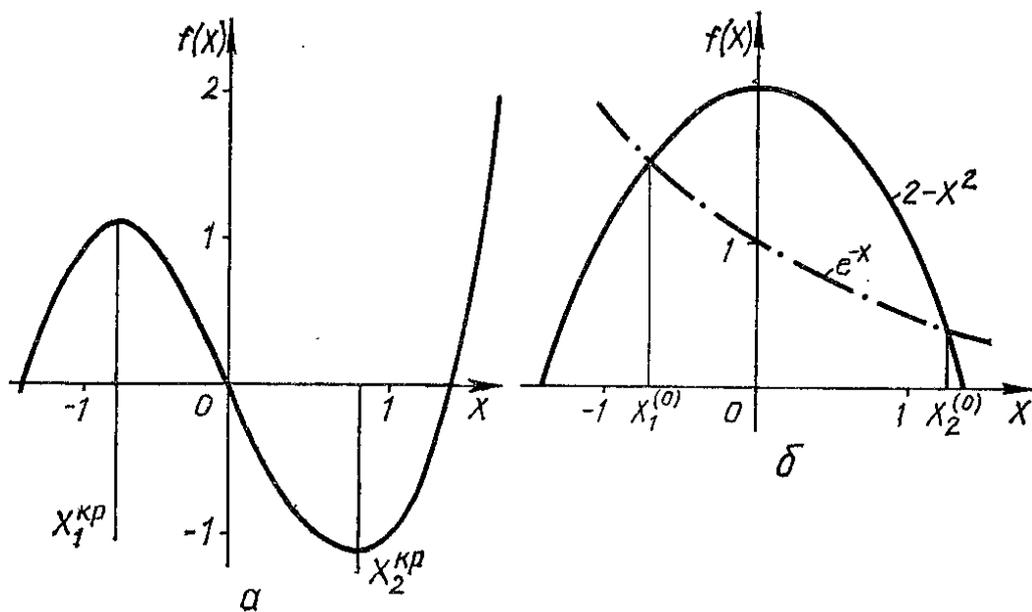


Рис. 1.5. Определение графическим методом корней функции:
 а — по графику $f(X)$; б — по параметрическим функциям $w(X)$ и $u(X)$

Следовательно, уравнение (1.31) имеет три корня. Отрезки изоляции их соответственно: $]-\infty; -0,82]$; $[-0,82; 0,82]$; $[0,82; +\infty[$.

Так как первый и второй интервалы весьма велики, то логично их сузить, воспользовавшись, например, подстановкой различных значений X и вычисляя значения функций:

$$\begin{aligned} f(-5) &= -115; & f(2) &= +4; \\ f(-2) &= -4; & f(5) &= +115. \end{aligned}$$

Искомые отрезки изоляции: $X_1^{(0)} \in [-2; -0,82]$; $X_2^{(0)} \in [-0,82; +0,82]$; $X_3^{(0)} \in [+0,82; 2]$.

Корни можно отделить также *графическим методом*, используя график функции $f(X)$. Так, на рис. 1.5, а показана функция (1.31), значения отрезков изоляции корней которой совпадают с ранее найденными аналитическим методом в пр. 1.5. Иногда удобно представить функцию $f(X)$ в виде $w(X) = u(X)$ и построить графики этих функций. Очевидно, что проекция точки пересечения графиков этих функций на ось абсцисс даст значение корня (корней) уравнения.

Пример 1.6. Отделить корни уравнения

$$f(X) = e^{-X} + X^2 - 2 = 0.$$

Строятся графики функций $w(X) = e^{-X}$ и $u(X) = 2 - X^2$ (рис. 1.5, б). В результате построения выделяются соответственно двум корням два отрезка изоляции $[-1; 0]$ и $[1; 2]$.

Использование современных быстродействующих ЭВМ позволяет отделить действительные корни уравнения методом последовательного перебора, подробный алгоритм которого изложен в [35].

Из алгоритмов уточнения корня наиболее надежным является *итерационный метод бисекции*, или *половинного деления*. Алгоритм представлен блок-схемой на рис. 1.6. Исходное уравнение

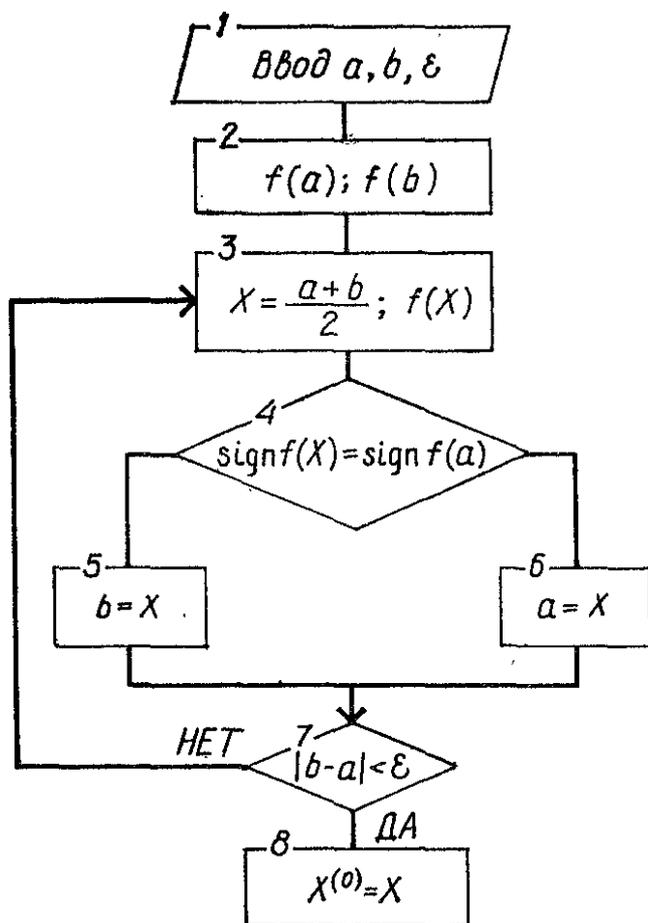


Рис. 1.6. Блок-схема алгоритма поиска корней методом половинного деления

Условием применимости метода итераций к численному решению уравнений является вопрос о сходимости последовательности $\{X_s\}$. Последовательность сходится (см. рис. 1.3, а), если существует предел

$$X^{(0)} = \lim_{s \rightarrow \infty} X_s. \quad (I.34)$$

Сходимость итерационного процесса обеспечивается неравенством

$$|\varphi'(X)| \leq q < 1, \quad (I.35)$$

выполняемом при всех $X \in [a; b]$ ($q = \max_{[a; b]} |\varphi'(X)|$).

Условием окончания процедуры последовательных приближений является неравенство

$$|X_s - X_{s-1}| \leq \varepsilon(1 - q)/q, \quad (I.36)$$

где ε — наперед заданное достаточно малое положительное число.

Пример 1.7. Методом итераций с точностью 10^{-3} вычислить на отрезке $[1; 2]$ корень уравнения

$$11e^{-X} - 9\ln X = 0,$$

тождественного приведенному на рис. В. 6. Преобразованием к виду $X = \varphi(X)$ исходное уравнение сводится к

$$X = \exp [11 \exp (-X)/9].$$

(I.30) предварительно приводится к виду

$$X = \varphi(X). \quad (I.32)$$

При известном отрезке изоляции корня $[a; b]$ начальное приближение, равное

$$X_0 = (a + b)/2,$$

подставляется в правую часть уравнения (I.32). Получается первое приближение

$$X_1 = \varphi(X_0).$$

В качестве второго приближения берется

$$X_2 = \varphi(X_1).$$

Если продолжить этот процесс далее, то получится числовая последовательность $\{X_s\}$, задаваемая рекуррентной формулой

$$X_{s+1} = \varphi(X_s), \quad s = 1, 2, \dots \quad (I.33)$$

Значение модуля производной правой части

$$|\Phi'(X)| = \left| -\exp \left[\frac{11}{9} \exp(-X) \right] \frac{11}{9} \exp(-X) \right|$$

на отрезке $[1; 2]$ всегда меньше единицы, следовательно, итерационная последовательность будет сходящейся. В качестве нулевого приближения берется значение $X_0 = 1$. Последовательные приближения X_s ($s = 1, 2, \dots$) вычисляются по формуле

$$X_{s+1} = \exp \left[\frac{11}{9} \exp(-X_s) \right]$$

до тех пор, пока два последовательных значения не совпадут друг с другом в пределах заданной точности. Численный анализ дает следующую последовательность:

$$X_1 = 1,5677; X_2 = 1,2903; X_3 = 1,3998; X_4 = 1,3518; X_5 = 1,3720; \\ X_6 = 1,2608; X_7 = 1,3670; X_8 = 1,3654; X_9 = 1,3661; X_{10} = 1,3658.$$

Значения X_9 и X_{10} отличаются на 0,0002, что меньше заданной точности 0,001. Приближенное значение корня $X^{(0)} = (X_9 + X_{10})/2 = 1,366$.

Исходное уравнение можно также преобразовать к виду $X = \Phi(X)$:

$$X = \ln \left(\frac{11}{9 \ln X} \right),$$

что соответствует итерационному процессу

$$X_{s+1} = \ln \frac{11}{9} - \ln \ln X_s.$$

Абсолютное значение производной

$$|\Phi'(X)| = \left| -\frac{1}{X \ln X} \right|$$

не при всех значениях X , принадлежащих $[1; 2]$, меньше единицы (в чем можно убедиться, используя, например, метод проб). Поэтому такое представление функции дает расходящийся итерационный процесс.

Если функция f достаточно гладкая (имеет одну или две непрерывные производные), то можно применить другие методы поиска корня, например метод хорд или метод Ньютона (касательных) [3, 7, 35, 60].

Для сравнения методов следует рассмотреть ошибку e_s приближения значения X_s к корню $X^{(0)}$ ($e_s = (X_s - X^{(0)})$). Итерационный процесс имеет сходимость порядка p , если ошибка e_s удовлетворяет соотношению

$$\lim_{s \rightarrow \infty} |e_{s+1}/e_s^p| = c \neq 0. \quad (I.37)$$

Определив аналитически порядок p , можно оценить скорость достижения решения для разных методов. Оказывается, что наиболее медленный алгоритм — метод бисекции ($p = 1$), а наиболее быстрый — метод Ньютона ($p = 2$). Однако первый метод гарантированно дает решение, тогда как метод касательных при неправильном выборе начального приближения может оказаться расходящимся.

1.4. Численные методы решения задач линейной алгебры

Одной из фундаментальных вычислительных задач линейной алгебры является решение системы n линейных уравнений вида

$$\left. \begin{aligned} a_{11}X_1 + a_{12}X_2 + \dots + a_{1n}X_n &= b_1; \\ a_{21}X_1 + a_{22}X_2 + \dots + a_{2n}X_n &= b_2; \\ \dots & \dots \dots \dots \dots \\ a_{n1}X_1 + a_{n2}X_2 + \dots + a_{nn}X_n &= b_n. \end{aligned} \right\} \quad (1.38)$$

Система (1.38) в матричном виде

$$\mathbf{aX} = \mathbf{b}, \quad (1.39)$$

где $\mathbf{a} = [a_{ij}]_1^n$ — заданная квадратная матрица порядка n ; $\mathbf{b} = [b_i]_1^n$ — заданный вектор-столбец с n компонентами и $\mathbf{X} = [X_j]_1^n$ — неизвестный вектор-столбец с n компонентами.

К решению систем линейных уравнений сводятся задачи аппроксимации дифференциальных и интегральных уравнений конечными, линеаризации нелинейных уравнений, приближенных функций, подбора полиномов или других кривых специального вида по заданной информации (задачи решения систем нормальных уравнений, полученных по методу наименьших квадратов) и пр. [4, 37 и др.].

Система уравнений (1.39) имеет единственное решение тогда и только тогда, когда матрица \mathbf{a} является обратимой (другие формулировки: $\det \mathbf{a} \neq 0$, столбцы \mathbf{a} суть независимые векторы). В таком случае система (1.39) называется *совместной*, и, согласно правилу Крамера, единственное решение ее задается формулами

$$X_1 = \frac{D_1}{D}; X_2 = \frac{D_2}{D}; \dots; X_n = \frac{D_n}{D}, \quad (1.40)$$

где $D = \det \mathbf{a}$; $D_j = \det \mathbf{a}_j$ — дополнительные определители, получаемые из матрицы \mathbf{a} [21]. Однако метод решения линейных систем с помощью определителей оказывается весьма громоздким при $n > 3$ (требует $n!n$ арифметических операций) и зачастую ведет к чрезмерным ошибкам округления даже при небольших n , что искажает результат вычислений.

Методы решения систем линейных уравнений типа (1.38) делятся на точные и приближенные. Из *точных методов* наиболее употребима схема Гаусса с ее различными модификациями. Эти схемы позволяют получить точное решение (набор $X_1^{(0)}, X_2^{(0)}, \dots, X_n^{(0)}$) после конечного числа математических операций. *Приближенные* или *итерационные* методы основаны на построении итерационной последовательности вычислений, сходящейся к искомому решению. В этом случае после определенного числа итераций можно получить приближенное решение с заданной точностью ϵ .

Выбор метода решения зависит от вида матрицы \mathbf{a} , порядка системы n , характеристик применяемой ЭВМ и программного обеспечения.

Метод приведения строк, или метод последовательного исключения неизвестных Гаусса, является одним из наиболее универсальных и эффективных. Его реализация по схеме единственного деления требует выполнения $n(4n^2 - 3n - 4)/6$ арифметических операций.

Этот метод состоит в преобразовании матрицы a к верхней треугольной матрице

$$v = \begin{pmatrix} v_{11} & v_{12} & \dots & v_{1n} \\ 0 & v_{22} & \dots & v_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & v_{nn} \end{pmatrix}. \quad (I.41)$$

Достигается это при помощи цепочки невырожденных элементарных преобразований, при которых из каждой строки вычитаются некоторые кратные величины выше расположенных строк. В результате обращаются в нуль все элементы, расположенные ниже диагонали. Поскольку произведение невырожденных преобразований также невырождено, по существу находится такая невырожденная матрица K размера $n \times n$ (которая оказывается нижней треугольной матрицей), что

$$Ka = v, \quad (I.42)$$

а система (I.39) теперь имеет вид

$$vX = KaX = Kb = d \quad (I.43)$$

и элементы вектора X легко получаются обратной подстановкой.

При решении по методу Гаусса на первом этапе (прямой ход) система (I.38) приводится к треугольному виду. На втором этапе (обратном ходе) последовательно определяются неизвестные из этой треугольной системы.

Считается, что коэффициент a_{11} отличен от нуля (если $a_{11} = 0$, то меняются местами строки 1 и i , где элемент $a_{i1} \neq 0$). Этот элемент называется *ведущим элементом первого шага*.

На первом шаге с помощью первого уравнения X_1 исключается из других уравнений. Это достигается прибавлением первого уравнения, умноженного на $M_{21} = -a_{21}/a_{11}$, ко второму уравнению, а первого уравнения, умноженного на $M_{31} = -a_{31}/a_{11}$, к третьему и т. д. Величины M_{ij} называются *множителями*. Преобразованная система имеет вид

$$\left. \begin{aligned} a_{11}X_1 + a_{12}X_2 + \dots + a_{1n}X_n &= b_1; \\ a'_{22}X_2 + \dots + a'_{2n}X_n &= b'_2; \\ \dots & \dots \\ a'_{n2}X_2 + \dots + a'_{nn}X_n &= b'_n. \end{aligned} \right\} \quad (I.44)$$

Здесь $a'_{ij}, b'_j (i, j = \overline{2, n})$ — измененные в результате первого шага коэффициенты и правые части, которые получаются после выполнения первого шага (') прямого хода метода Гаусса. Основным

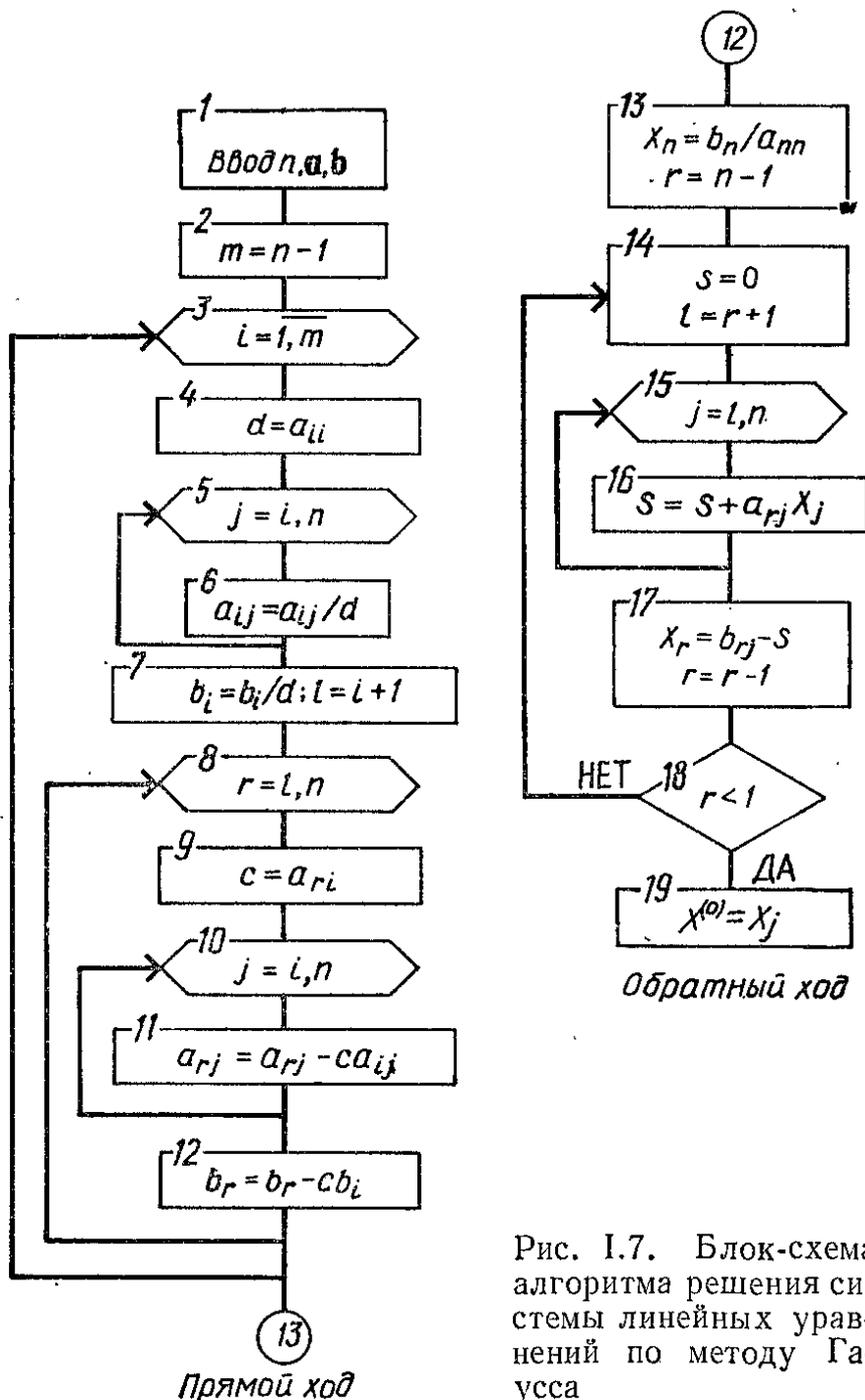


Рис. 1.7. Блок-схема алгоритма решения системы линейных уравнений по методу Гаусса

результатом первого шага является система $(n - 1)$ -го порядка с неизвестными X_2, X_3, \dots, X_n , образованная последними $n - 1$ уравнениями системы (I.44).

Аналогичные преобразования проводятся с остальными уравнениями системы (см. пр. 1.7).

После $n - 1$ шагов получается треугольная система:

$$\left. \begin{aligned} a_{11}X_1 + a_{12}X_2 + \dots + a_{1n}X_n &= b_1; \\ a'_{22}X_2 + \dots + a'_{2n}X_n &= b'_2; \\ \dots &\dots \\ a^{(n-1)}_{nn}X_n &= b^{(n-1)}_n. \end{aligned} \right\} \quad (I.45)$$

На этом завершается первый этап решения системы — приведение к верхней диагональной форме. Теперь матрица коэффициентов левой части системы (I.45) по виду совпадает с выражением для v .

Второй этап — обратный ход — заключается в решении треугольной системы (1.45), из последнего уравнения которой прямо следует значение $X_n^{(0)}$. Значение $X_{n-1}^{(0)}$ получается при подстановке найденного $X_n^{(0)}$ в $(n-1)$ -е уравнение и так далее. Блок-схема изложенного алгоритма приведена на рис. 1.7.

Пример 1.8. Решить систему линейных уравнений методом Гаусса:

$$\begin{aligned} 2X_1 + 8X_2 + 6X_3 &= 20; \\ 4X_1 + 2X_2 - 2X_3 &= -2; \\ 4X_1 - X_2 + X_3 &= 11. \end{aligned}$$

После деления 1-й строки на 2-ю получается

$$\left. \begin{aligned} X_1 + 4X_2 + 3X_3 &= 10; \\ 4X_1 + 2X_2 - 2X_3 &= -2; \\ 3X_1 - X_2 + X_3 &= 11. \end{aligned} \right\}$$

Удалить члены с X_1 из 2-го и 3-го уравнений можно, вычитая из 2-й строки 1-ю, умноженную на 4, а из 3-й — первую, умноженную на 3 ($M_{21} = -4/1$; $M_{31} = -3/1$):

$$\left. \begin{aligned} X_1 + 4X_2 + 3X_3 &= 10; \\ -14X_2 - 14X_3 &= -42; \\ -13X_2 - 8X_3 &= -19; \end{aligned} \right\}$$

2-я и 3-я строки делятся соответственно на -14 и -1 :

$$\left. \begin{aligned} X_1 + 4X_2 + 3X_3 &= 10; \\ X_2 + X_3 &= 3; \\ 13X_2 + 8X_3 &= 19. \end{aligned} \right\}$$

Вторая строка умножается на 13 ($M_{32} = -13/14$) и вычитается из 3-й:

$$\left. \begin{aligned} X_1 + 4X_2 + 3X_3 &= 10; \\ X_2 + X_3 &= 3; \\ -5X_3 &= -20. \end{aligned} \right\}$$

Наконец, 3-я строка делится на -5 :

$$\left. \begin{aligned} X_1 + 4X_2 + 3X_3 &= 10; \\ X_2 + X_3 &= 3; \\ X_3 &= 4. \end{aligned} \right\}$$

Процедура обратного хода дает искомое решение: $X_3^{(0)} = 4$; $X_2^{(0)} = -1$; $X_1^{(0)} = 10 - 4X_2^{(0)} - 3X_3^{(0)} = 2$.

При выборе ведущих элементов предполагалось отличие их от нуля, что дало теоретически правильное решение по методу Гаусса. Однако для получения надежного результата недостаточно избегать только нулевых ведущих элементов, необходимо также избегать выбора относительно малых ведущих элементов [7, 28, 62], операции с которыми приводят к увеличению ошибки округления при расчетах на ЭВМ.

Чтобы избежать нежелательного роста значений элементов матрицы во время прямого хода и тем самым предотвратить большую потерю точности, обычно метод Гаусса применяется в сочетании с какой-нибудь схемой выбора ведущего элемента на каждом шаге.

Существуют две стандартные стратегии выбора. Первая стратегия состоит в частичном выборе ведущего элемента: на k -м шаге прямого хода в качестве ведущего берется наибольший (по абсолютному значению) элемент в неприведенной части k -го столбца, т. е. $|a_{kk}| = \max |a_{ik}|$, ($i = k, k+1, \dots, n$). Вторая стратегия состоит в полном выборе ведущего элемента: на k -м шаге прямого хода в качестве ведущего берется наибольший по абсолютному значению элемент в неприведенной части матрицы $|a_{kk}| = \max |a_{ij}|$ ($i = \overline{k, n}; j = \overline{k, n}$).

Ошибки округления вызваны, с одной стороны, чувствительностью к ним самой задачи, с другой — алгоритмом решения. Первая трудность является практически непреодолимой, можно лишь попытаться идентифицировать такого рода задачи. Выбор ведущего элемента позволяет устранить вторую трудность, связанную с эффектом увеличения ошибки округления.

Существуют более сложные и более надежные схемы выбора ведущих элементов. Любая специальная схема выбора усложняет программу и требует дополнительных вычислительных ресурсов. Однако в конечном счете это оправдывается благодаря повышению точности решения задачи.

Если в методе исключения Гаусса продолжить элементарные преобразования строк после приведения матрицы \mathbf{a} к треугольной форме \mathbf{v} , то последняя приводится к диагональной форме (*модификация Жордана*). Например, последняя строка системы (1.45), умноженная на соответствующие множители, может вычитаться из предыдущих строк, чтобы обратить в нуль элементы последнего столбца матрицы \mathbf{v} . Затем предпоследняя строка матрицы \mathbf{v} может быть использована, чтобы обратить в нуль элементы предпоследнего столбца и т. д. Такие преобразования называют жордановыми исключениями. По методу Гаусса — Жордана расчет производится аналогично способу Гаусса, но решение системы (ввиду того, что в левой части единичная матрица) получается сразу, без обратного хода.

Пример 1.9. Решить систему уравнений пр. 1.8 методом Гаусса — Жордана.

Первый шаг исключения проводится аналогично тому, как это выполнено в пр. 1.8:

$$\left. \begin{aligned} 1X_1 + 4X_2 + 3X_3 &= 10; \\ 0X_1 + 14X_2 - 14X_3 &= -42; \\ 0X_1 - 13X_2 - 8X_3 &= -19. \end{aligned} \right\}$$

После деления 2-й строки на -14 получается

$$X_2 + X_3 = 3.$$

Последнее уравнение умножается на 4 и на 13, вычитается из первого уравнения и прибавляется к третьему соответственно:

$$\left. \begin{aligned} 1X_1 + 0X_2 - 1X_3 &= -2; \\ 0X_1 + 1X_2 + 7X_3 &= -3; \\ 0X_1 + 0X_2 + 5X_3 &= 20. \end{aligned} \right\}$$

Из третьей строки непосредственно следует $X_3^{(0)} = 4$. Теперь это уравнение, умноженное на $+1$ и на -1 , складывается соответственно с первым и вторым уравнениями:

$$\left. \begin{aligned} 1X_1 + 0X_2 + 0X_3 &= 2; \\ 0X_1 + 1X_2 + 0X_3 &= -1; \\ 0X_1 + 0X_2 + 1X_3 &= 4. \end{aligned} \right\}$$

Окончательное решение ввиду диагональности последней системы $X_1^{(0)} = 2$; $X_2^{(0)} = -1$; $X_3^{(0)} = 4$ совпадает с решением пр. 1.8.

Нередко в процессе решения систем уравнений $\mathbf{aX} = \mathbf{b}$ помимо ошибок округления, связанных с особенностями ЭВМ, вносятся погрешности в определении коэффициентов и правых частей, зачастую определяемых экспериментально. Кроме того, в задачах обработки данных пассивного эксперимента расположение точек в факторном пространстве может вызывать сильную корреляцию между значениями разных факторов. Это приводит к тому, что матрица близка к вырожденной и найденное решение $X^{(0)}$ может оказаться неверным. Близость матрицы к вырожденной оценивается с помощью так называемой *меры обусловленности матрицы* $\mu(\mathbf{a})$ [7, 88, 62].

Следует отметить, что программы, реализующие ту или иную модификацию метода Гаусса, не дают возможности избежать больших ошибок в определении решения системы $\mathbf{aX} = \mathbf{b}$ при больших $\mu(\mathbf{a})$. В этих случаях рекомендуется применение методов, основанных на матричной факторизации [7, 62].

Во многих линейных системах с большим числом уравнений и неизвестных большая часть матрицы \mathbf{a} оказывается заполненной нулевыми элементами. Сюда относятся трехдиагональные матрицы, ленточные, матрицы, элементы которых задаются простой формулой и могут вычисляться в нужный момент. Естественно желание при работе с такими матрицами сэкономить память ЭВМ; в этой ситуации линейные системы решаются специальными методами: итерационным методом, методом прогонки и др. [37, 67].

Суть вычислений *итерационными методами* состоит в следующем. Расчет начинается с некоторого заранее выбранного приближения X^0 (начального приближения). Вычислительный процесс, использующий матрицу \mathbf{a} , векторы \mathbf{b} и X^0 , приводит к новому вектору X^1 . Затем процесс повторяется, только вместо X^0 используется новое значение X^1 . На s -м шаге итерационного процесса по \mathbf{a} , \mathbf{b} , X^{s-1} получают X^s . При выполнении некоторых заранее оговоренных условий процесс сходится к решению $X^{(0)}$ при $s \rightarrow \infty$.

Специфичность каждой конкретной задачи сказывается на выборе того или иного итерационного метода, поэтому в математических библиотеках ЭВМ программы, реализующие методы итерации, практически отсутствуют. Вместе с тем все итерационные методы в основном являются модификациями двух основных: метода

простой итерации и метода Зейделя. При этом систему линейных уравнений (I.39) приводят к виду

$$X = QX + q, \quad (I.46)$$

где Q — заданная квадратная матрица n -го порядка; q — свободный член, входящий в линейное пространство R^n (см. прил. I.1).

Метод простых итераций (или метод Якоби) заключается в построении итерационной последовательности векторов, начиная с некоторого начального $X^0 \in R^n$:

$$X^s = QX^{s-1} + q \quad (s = 1, 2, \dots). \quad (I.47)$$

Этот метод зависит от правильной нумерации уравнений и неизвестных. Можно показать, что если норма матрицы $\|Q\| < 1$, то система уравнений (I.46) имеет единственное решение и сходится к нему со скоростью геометрической прогрессии.

Пример I.10. Представить систему линейных уравнений

$$\left. \begin{aligned} 3X_1 + 4X_2 - X_3 &= 7; \\ 2X_1 + 6X_2 - 3X_3 &= -2; \\ -X_1 + X_2 + 4X_3 &= 4 \end{aligned} \right\}$$

к виду, пригодному для построения итерационного процесса методом Якоби.

Все члены, кроме X_i , переносятся в правую часть:

$$\left. \begin{aligned} X_1 &= \frac{1}{3} (7 - 4X_2 + X_3); \\ X_2 &= \frac{1}{6} (-2 - 2X_1 - 3X_3); \\ X_3 &= \frac{1}{4} (4 + X_1 - X_2) \end{aligned} \right\}$$

или в матричных обозначениях (I.46)

$$Q = \begin{pmatrix} 0 & -\frac{4}{3} & \frac{1}{3} \\ -\frac{2}{6} & 0 & -\frac{3}{6} \\ \frac{1}{4} & -\frac{1}{4} & 0 \end{pmatrix}$$

и $q = (7/3, -2/6, 4/4)^T$.

Задается начальное приближение вектора $X = (X_1^0, X_2^0, X_3^0)$, которое подставляется в правую часть.

Погрешность s -го приближения можно оценить через норму разности двух последовательных приближений и норму матрицы

$$\|X^{(0)} - X^s\| \leq \frac{\|Q\|}{1 - \|Q\|} \|X^s - X^{s-1}\|. \quad (I.48)$$

В итерационном процессе метода Гаусса — Зейделя от системы $aX = b$ переходят к эквивалентной системе вида $X = cX + d$ в предположении, что у матрицы a все диагональные элементы отличны от нуля. Для этого i -е уравнение системы $aX = b$ делится на a_{ii} , а затем все неизвестные, кроме X_i , переносятся вправо:

$$\mathbf{d} = (d_1, d_2, \dots, d_n), \quad d_i = \frac{b_i}{a_{ii}}, \quad \mathbf{c} = [c_{ij}]_{i,j=1}^n,$$

$$c_{ij} = \begin{cases} \frac{a_{ij}}{a_{ii}}, & j \neq i; \\ 0, & j = i. \end{cases} \quad (1.49)$$

Последовательность итераций задается формулой

$$X_i^s = \sum_{j=1}^{i-1} c_{ij} X_j^s + \sum_{j=i+1}^n c_{ij} X_j^{s-1} + d_i, \quad (1.50)$$

где $i = \overline{1, n}$; $j = \overline{1, n}$. При этом необходимо также задать вектор начального приближения \mathbf{X}^0 .

Вычисления в этом методе почти такие же, как и в методе Якоби, с той лишь разницей, что в последнем новые значения \mathbf{X}^s не используются до следующей итерации. В методе Гаусса — Зейделя при нахождении i -й компоненты используются уже найденные компоненты этой же итерации с меньшими номерами.

Сходимость метода Гаусса — Зейделя обеспечивается при выполнении условия

$$\sum_{j \neq i} |a_{ij}| < |a_{ii}| \quad (i = \overline{1, n}), \quad (1.51)$$

либо если матрица \mathbf{a} симметрична и положительно определена (см. прил. I.1).

Пример I.11. Задать итерационный процесс по методу Гаусса — Зейделя для нахождения решений линейной системы уравнений из пр. I.10.

Вычисляются элементы матрицы \mathbf{c} и вектора \mathbf{d} :

$$\mathbf{c} = \begin{pmatrix} 0 & -\frac{4}{7} & \frac{1}{7} \\ -\frac{1}{3} & 0 & \frac{1}{2} \\ \frac{1}{4} & -\frac{1}{4} & 0 \end{pmatrix}; \quad \mathbf{d} = \begin{pmatrix} \frac{7}{3} \\ -\frac{1}{3} \\ 1 \end{pmatrix}.$$

Достаточное условие сходимости выполняется:

$$\begin{aligned} |a_{12}| + |a_{13}| &= 4 + 1 < |a_{11}| = 7; \\ |a_{21}| + |a_{23}| &= 2 + 3 < |a_{22}| = 6; \\ |a_{31}| + |a_{32}| &= 1 + 1 < |a_{33}| = 4. \end{aligned}$$

После задания начального приближения $\mathbf{X}^0 = \mathbf{d} = \left(\frac{7}{3}; -\frac{1}{3}; 1\right)$ выражение для первого приближения имеет вид

$$X_1^1 = -\frac{4}{7} X_2^0 + \frac{1}{7} X_3^0 + \frac{7}{3};$$

$$X_2^1 = -\frac{1}{3} X_1^1 - \frac{1}{2} X_3^0 - \frac{1}{3};$$

$$X_3^1 = \frac{1}{4} X_1^1 - \frac{1}{4} X_2^1 + 1.$$

Аналогично выполняются и последующие итерации.

1.5. Интерполяция, численное интегрирование и дифференцирование

Полное описание функциональной зависимости предполагает задание формулы или бесконечного набора чисел. Однако при решении задач на ЭВМ операции производятся только с конечными совокупностями чисел. Возникает необходимость приближенно охарактеризовать функцию *конечным набором чисел*.

Согласно *методу сеток*, функции описываются их значениями в конечном числе s -х точек. Это означает, что в соответствие функции $f(X)$, определенной на $[a; b]$, ставится совокупность ее значений в узлах сетки X_1, X_2, \dots, X_n :

$$f_s = f(X_s) \quad (s = \overline{1, n}). \quad (1.52)$$

Можно показать, что любая функция непрерывного аргумента имеет сеточное представление.

В практических приложениях наиболее часто встречается *обратная задача*: восстановить функцию по ее сеточному представлению. Ее можно решить при помощи интерполяции.

Задача одномерной интерполяции состоит в построении некоторой такой функции f (вид ее заранее задается), что $f(X_s) = y_s$ для всех s , и при этом $f(X)$ должна принимать разумные значения для X , лежащих между заданными точками. Критерий разумности для каждой задачи свой, и ему, возможно, никогда не будет дано точного определения.

Практически наиболее важным классом интерполирующих функций является множество алгебраических полиномов. Задача полиномиальной интерполяции: построить полином $P_{n-1}(X; f) = P_{n-1}(X; X_1, \dots, X_n; f_1, \dots, f_n)$ степени $n - 1$, принимающий в узлах X_1, X_2, \dots, X_n значения f_1, f_2, \dots, f_n . Из курса математического анализа известно, что любая непрерывная функция может быть представлена на конечном интервале $[a; b]$ полиномом P_n степени n , а через n точек можно провести единственный полином степени $n - 1$.

Восстановление коэффициентов полинома $P_{n-1}(X; f)$ можно провести, решая систему n уравнений относительно неизвестных линейных коэффициентов $a_s (s = \overline{0, n-1})$. Однако во многих случаях уравнения чрезвычайно плохо обусловлены. Гораздо более удовлетворительный способ вычисления полинома, интерполирующего точки (X_s, f_s) , состоит в использовании базиса лагранжевых полиномов на множестве $\{X_s\}$. Это такие полиномы $l_j(X) (j = \overline{0, n-1})$ степени $n - 1$, что

$$l_j(x) = \begin{cases} 1, & \text{если } j = s; \\ 0, & \text{если } j \neq s. \end{cases}$$

Этим условиям удовлетворяет полином

$$l_j(X) = \frac{(X - X_1) \dots (X - X_{j-1})(X - X_{j+1}) \dots (X - X_n)}{(X_j - X_1) \dots (X_j - X_{j-1})(X_j - X_{j+1}) \dots (X_j - X_n)}. \quad (1.53)$$

Полином (I.53) определяется единственным образом. Каждый множитель числителя обращает $l_j(X_s)$ в нуль при некотором $s \neq j$. Соответствующие множители знаменателя нормируют полином так, что $l_j(X_j) = 1$. Полином $l_j(X) f_j$ принимает значение f_j в точке X_j и равен нулю во всех точках $X_s (s \neq j)$. Таким образом, интерполяционный полином степени $n - 1$, который проходит через n точек (X_s, f_s) , выражается формулой

$$P_{n-1}(X, f) = \sum_{j=1}^n l_j(X) f(X_s) \quad (I.54)$$

и называется *интерполяционным полиномом Лагранжа*.

Погрешность полиномиальной интерполяции в интервале $[a; b]$ оценивается по формуле

$$f(X) - P_{n-1}(X, f) = \frac{M_n}{n!} |(X - X_1) \dots (X - X_n)|, \quad (I.55)$$

где $M_n = \max_{[a; b]} |f^{(n)}(X)|$, $f^{(n)}$ — производная порядка n .

Целесообразный выбор числа узлов n , а следовательно, и степени интерполирующего полинома основывается на имеющихся сведениях о существовании и оценках производных интерполируемых функций. Если производные высоких порядков велики по абсолютному значению, то точность аппроксимации, согласно (I.55), будет мала, и не имеет смысла выбирать большое число узлов. Кроме того, при повышении степени полинома (I.54) результат из-за плохой обусловленности становится весьма чувствительным к ошибкам округления. При недостаточной гладкости функции $f(X)$ отрезок $[a; b]$ разбивают на частичные интервалы и на каждом из них применяют интерполяцию невысокого порядка. Например, если известно, что $f(X)$ имеет ограниченной только вторую производную $|f''(x)| \leq M_2$, целесообразно ограничиваться линейной интерполяцией. Если провести кусочно-линейную интерполяцию на равномерной сетке с шагом h , то ее погрешность не превышает $M_2 h^2 / 8$. Таким образом, процесс кусочно-линейной интерполяции сходится при $h \rightarrow 0$ со скоростью h^2 , тогда как процесс интерполяции в целом может расходиться.

Для приближенного восстановления функции по ее значениям на сетке используются кроме полиномов также другие функции. Например, для представления периодических функций целесообразно применять тригонометрическую интерполяцию.

Пример 1.12. По опытным данным о зависимости водопоглощения ω от температуры обжига T :

$T, ^\circ\text{C}$	1000	1200	1300
$\omega, \%$	15	7	2

определить водопоглощение для промежуточного значения $T = 1250 ^\circ\text{C}$.

Интерполяционный полином Лагранжа (I.54) для трехточечной интерполяции

$$P_2 = \frac{(X - X_2)(X - X_3)}{(X_1 - X_2)(X_1 - X_3)} f_1 + \frac{(X_1 - X_1)(X - X_3)}{(X_2 - X_1)(X_2 - X_3)} f_2 + \frac{(X - X_1)(X - X_2)}{(X_3 - X_1)(X_3 - X_2)} f_3. \quad (I.56)$$

При подстановке в (I.56) значений $X = T = 1250$, $X_1 = 1000$, $X_2 = 1200$, $X_3 = 1300$, $f_1 = \omega_1 = 15$, $f_2 = 7$ и $f_3 = 2$ получено

$$P_2(1250) = \frac{(1250 - 1200)(1250 - 1300)}{(1000 - 1200)(1000 - 1300)} \cdot 15 + \frac{(1250 - 1000)(1250 - 1300)}{(1200 - 1000)(1200 - 1300)} \cdot 7 + \frac{(1250 - 1000)(1250 - 1200)}{(1300 - 1000)(1300 - 1200)} \cdot 2 = 4,6.$$

Таким образом, при $T = 1250$ °С водопоглощение $\omega = 4,6$ %.

К числу других важнейших применений сеточного представления функций относятся численное интегрирование и дифференцирование.

Основная идея *численного интегрирования* заключается в следующем. При вычислении определенного интеграла

$$I = \int_a^b f(X) dX \quad (I.57)$$

на отрезке $[a; b]$ пользуются, как правило, *кусочно-полиномиальной интерполяцией*. Отрезок $[a; b]$ разбивают на несколько частичных интервалов и, заменяя f приближенно интерполяционным полиномом, вычисляют интеграл по каждому частичному интервалу.

Если (I.57) — интеграл от непрерывной на $[a; b]$ функции, то приближенное равенство

$$\int_a^b f(X) dX \approx \sum_{j=1}^n q_j f(X_j) \quad (I.58)$$

(где q_j — некоторые числа, а X_j — некоторые точки отрезка $[a; b]$) называется *квадратурной формулой*, определяемой весами q_j и узлами X_j . Квадратурная формула точна для многочленов степени n , если при замене f на произвольный алгебраический многочлен степени m приближенное равенство (I.58) становится точным. Наиболее употребительными в численном интегрировании являются квадратурные формулы прямоугольников, трапеций и Симпсона, которые можно получить, заменяя площадь под кривой BFC (рис. I. 8, а) площадью прямоугольников ($AB \cdot AD$), ($CD \cdot AD$), ($FE \times \times AD$), трапеции — $AD(AB + CD)/2$ или другой геометрической фигуры.

Квадратурная формула Симпсона является следствием интерполяции $f(X)$ на интервале (X_k, X_{k+2}) . Пусть частичная область состоит из двух интервалов длины h : $X_{k+1} = X_k + h$, $X_{k+2} = X_k + 2h$. Тогда, заменяя истинную кривую BFG интерполяционной параболой,

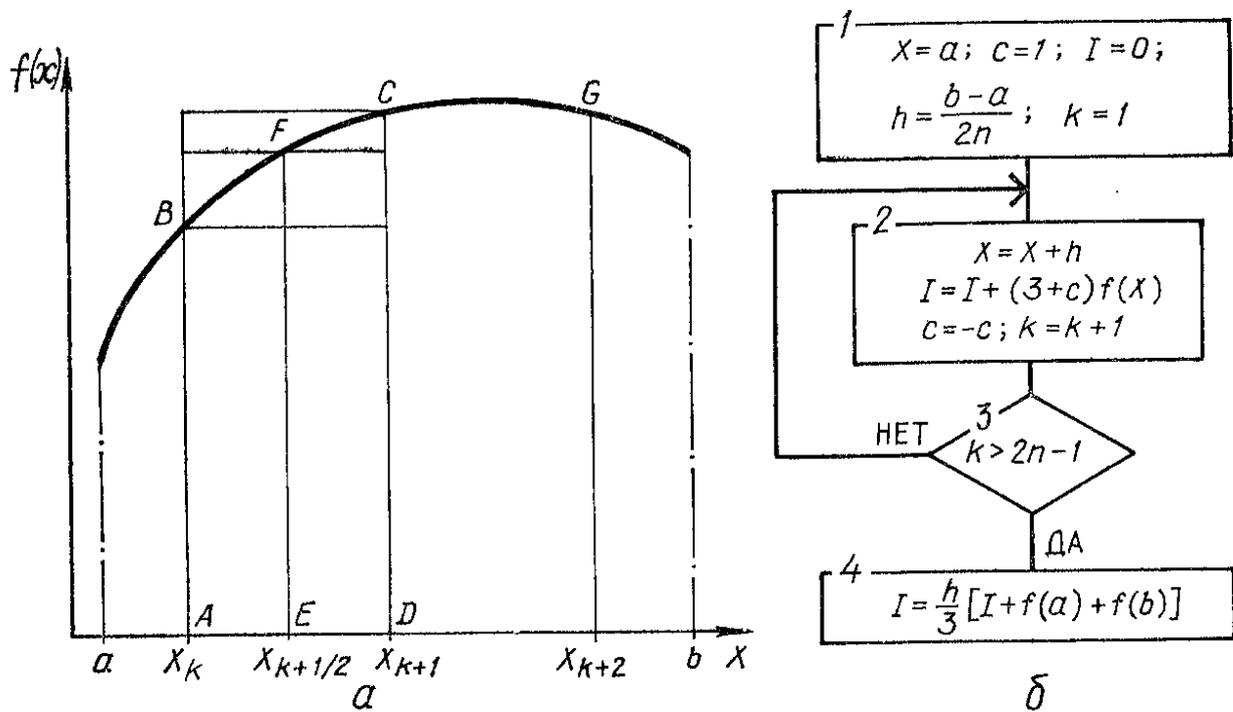


Рис. 1.8. Графическое отображение (а) численного интегрирования по квадратурной формуле и блок-схема алгоритма вычислений (б)

проведенной через точки f_k, f_{k+1}, f_{k+2} , получают формулу Симпсона

$$I \approx \frac{h}{3} [f(X_k) + 4f(X_{k+1}) + f(X_{k+2})]. \quad (I.59)$$

На отрезке $[a; b]$ при

$$h = \frac{b-a}{2n}; X_k = a + kh; y_k = f(X_k) \quad (k = \overline{0, 2n})$$

формулу Симпсона можно представить в таком виде:

$$I \approx \frac{h}{3} \left[y_0 + y_{2n} + \sum_{k=1}^{2n-1} (3 + c_k) y_k \right], \quad (I.60)$$

где

$$c_k = \begin{cases} 1 & \text{при нечетном } k; \\ -1 & \text{при четном } k. \end{cases}$$

При грубой оценке интеграла от остаточного члена квадратичной интерполяции погрешность формулы составляет $\epsilon(h^4)$. Блок-схема алгоритма вычисления представлена на рис. 1.8, б.

Пример 1.13. По опытным данным о зависимости энтропии s от температуры процесса T :

$T(s), \text{ K}$	278	303	323	340	350	354	347	330	325
$s, \text{ кДж}/(\text{кг} \cdot \text{K})$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9

определить среднеинтегральную температуру газа $T_{с.и.}$, используя для этого термодинамическое соотношение [29, 52 и др.]

$$T_{с.и.} = \left(\int_{t_0}^{s_{2n}} T(s) ds \right) (s_{2n} - s_0). \quad (I.61)$$

Интеграл в (I.61) вычисляется по методу Симпсона:

$$I \approx \frac{h}{3} [f(s_0) + f(s_8) + 2f(s_2) + 4f(s_3) + 2f(s_4) + 4f(s_5) + 2f(s_6) + 4f(s_7)] = \\ = \frac{0,1}{3} [278 + 325 + 4 \cdot 303 + 2 \cdot 323 + 4 \cdot 340 + 2 \cdot 350 + 4 \cdot 354 + 2 \cdot 347 + \\ + 4 \cdot 330] = 265.$$

Тогда

$$T_{с.и} = I/(s_{2n} - s_0) = \frac{265}{0,8} = 331 K.$$

Другой, не менее важной в прикладном анализе задач, базирующейся на сеточном представлении функции, является *численное дифференцирование*, или *вычисление производных*.

Пусть известны значения функции $f(X)$ в точках X_1, X_2, \dots, X_n , близких к точке X (точка X может совпадать с одной из точек X_1, X_2, \dots, X_n). Требуется найти приближенное значение производной $f'(X)$.

Один из способов решения этой задачи основан на замене функции $f(X)$ ее интерполяционным полиномом

$$\frac{d}{dX} f \approx \frac{d}{dX} P_{n-1}(X; f). \quad (I.62)$$

В другом способе — *неопределенных коэффициентов* — используется формула разложения в ряд Тейлора функции $f(X)$, непрерывной в окрестности X_k (см. прил. I.3):

$$f_k = f(X_k) = f(X) + (X_k - X)f'(X) + \frac{1}{2!}(X_k - X)^2 f''(X) + \dots \quad (I.63)$$

Составляется линейная комбинация

$$\bar{f}' = \sum_{k=1}^n a_k f_k.$$

Коэффициенты a_k подбираются так, чтобы разность $(\bar{f}' - f')$ стремилась к нулю при стягивании к X узлов X_1, X_2, \dots, X_n .

Используя алгебраический метод неопределенных коэффициентов, можно получить трехточечную одностороннюю аппроксимацию первой производной, имеющую второй порядок точности. На основании формулы Тейлора

$$f(X+h) = f(X) + hf'(X) + \frac{h^2}{2} f''(X) + \varepsilon(h^3); \\ f(X+2h) = f(X) + 2hf'(X) + 2h^2 f''(X) + \varepsilon(h^3).$$

Вычитанием учетверенной первой строки из второй находится

$$f'(X) = \frac{-f(X+2h) + 4f(X+h) - 3f(X)}{2h} + \varepsilon(h^2), \quad (I.64)$$

где $\varepsilon(h^2)$ — погрешность вычисления производной.

Пример 1.14. Найти симметричную аппроксимацию второй производной функции $f(X)$.

Пусть $X_1 = X - h$; $X_2 = X$; $X_3 = X + h$. На основании формулы Тейлора (I.63)

$$f(X + h) = f(X) + hf'(X) + \frac{h^2}{2} f''(X) + \frac{h^3}{6} f'''(X) + \varepsilon(h^4);$$

$$f(X - h) = f(X) - hf'(X) + \frac{h^2}{2} f''(X) - \frac{h^3}{6} f'''(X) + \varepsilon(h^4).$$

Сложив эти равенства и перенеся $f''(X)$ в левую часть, получим

$$f''(X) = \frac{f(X + h) - 2f(X) + f(X - h)}{h^2} + \varepsilon(h^2). \quad (\text{I.65})$$

Однако при всей простоте и очевидности формул численного дифференцирования функций, заданных сеточным представлением, существует опасность внесения больших погрешностей в аппроксимацию производных. Это вызвано причинами так называемой «численной неустойчивости». Оказывается, что формулы (I.64), (I.65) имеют способность умножать эффекты помех как при чрезмерно малом шаге сетки h , так и при больших его значениях. Очевидно, в каждом случае необходим подбор оптимального шага для максимально возможного исключения помех.

Для уменьшения действия возмущений при численном дифференцировании пользуются различными приемами регуляризации. В качестве простейшего оператора регуляризации функции можно использовать сглаживание с параметром ζ , подбираемым в предварительных расчетах:

$$f'_k = (1 - 2\zeta) f_k + \zeta f_{k-1} + \zeta f_{k+1}. \quad (\text{I.66})$$

Пример 1.15. По данным о зависимости предела прочности бетона R от водоцементного отношения В/Ц:

В/Ц	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75
R , МПа	32	36	41	41	38	36	34	32	30	28

оценить по значению $f'(B/C)$ чувствительность R к изменению В/Ц при значениях последнего: 0,3; 0,45 и 0,6.

Подстановкой соответствующих значений в формулу (I.64) получено:

$$f'(0,3) \approx [-f(0,3 + 2 \cdot 0,05) + 4 f(0,3 + 0,05) - 3f(0,3)] / (2 \times$$

$$\times 0,05) = (-41 + 4 \cdot 36 - 3 \cdot 32) / 0,1 = 70;$$

$$f'(0,45) \approx (-36 + 4 \cdot 38 - 3 \cdot 41) / 0,1 = -70;$$

$$f'(0,6) \approx (-30 + 4 \cdot 32 - 3 \cdot 34) / 0,1 = -40.$$

Следовательно, в большей части рассмотренного диапазона В/Ц по мере его увеличения прочность бетона снижается, причем при больших В/Ц темп такого снижения меньше.

Необходимо отметить, что ряд численных методов может быть реализован на программируемых микрокалькуляторах [16, 49 и др.] за вполне приемлемое для технологов время и с заданной, исходя из инженерной потребности, точностью.

Глава II. МЕТОДЫ ПОИСКА ОПТИМУМА В СТРОИТЕЛЬНО-ТЕХНОЛОГИЧЕСКИХ ЗАДАЧАХ

II.1. Общая постановка и классификация задач оптимизации

Большинство задач в строительном материаловедении и технологии, как и в любой практической деятельности, предполагает выбор решения из множества возможных вариантов. *Оптимизация* — определение наиболее целесообразного варианта решения, т. е. лучшего с точки зрения намеченной цели. Сравнение вариантов по отношению к цели осуществляется на основе критерия предпочтения — *критерия оптимальности* (эффективности, качества — см. «Введение»).

В строительной-технологической практике содержание задач оптимизации может быть самым разным: организовывать поставку сырьевых материалов заводу железобетонных конструкций и вывоз готовых изделий таким образом, чтобы транспортные расходы были минимальны; обеспечить как можно более низкий процент брака при назначении параметров технологического процесса; снизить расход портландцемента при производстве изделий требуемого качества за счет использования добавок и наполнителей; определить соотношение между компонентами полимерной композиции, обеспечивающее максимальную прочность материала при изгибе, и т. д.

Решить содержательно разные задачи оптимизации позволяют общие математические методы. Для этого необходимо конкретную задачу сформулировать как математическую, т. е. описать в математических терминах оба обязательных элемента оптимизации — множество вариантов решений и критерий оптимальности, придавая, таким образом, количественные оценки возможным вариантам и количественную меру их «близости» к цели.

Большинство разных по содержанию задач приводится к следующей общей математической задаче оптимизации: найти среди элементов заданного множества Ω тот элемент x^* , для которого заданная функция $f(x)$ принимает наименьшее или наибольшее значение $f(x^*)$. Множество Ω называется *множеством допустимых решений*, его элемент x^* — *оптимальным решением*; $f(x)$ — *целевая функция*, или *критерий оптимальности*; $f(x^*) = f^*$ — *минимум* или *максимум*, а в общем случае — *оптимум (экстремум)*.

Обычно при постановке задачи оптимизации используют запись

$$f(x) \rightarrow \min, \quad x \in \Omega, \quad (\text{II.1})$$

а оптимальное решение записывается как

$$x^* = \min_{x \in \Omega} f(x) \quad \text{или} \quad x^* = \operatorname{argmin}_{x \in \Omega} f(x). \quad (\text{II.2})$$

Очевидно, при определении максимума задача (II.1) примет вид

$$-f(x) \rightarrow \min, \quad x \in \Omega, \quad (\text{II.3})$$

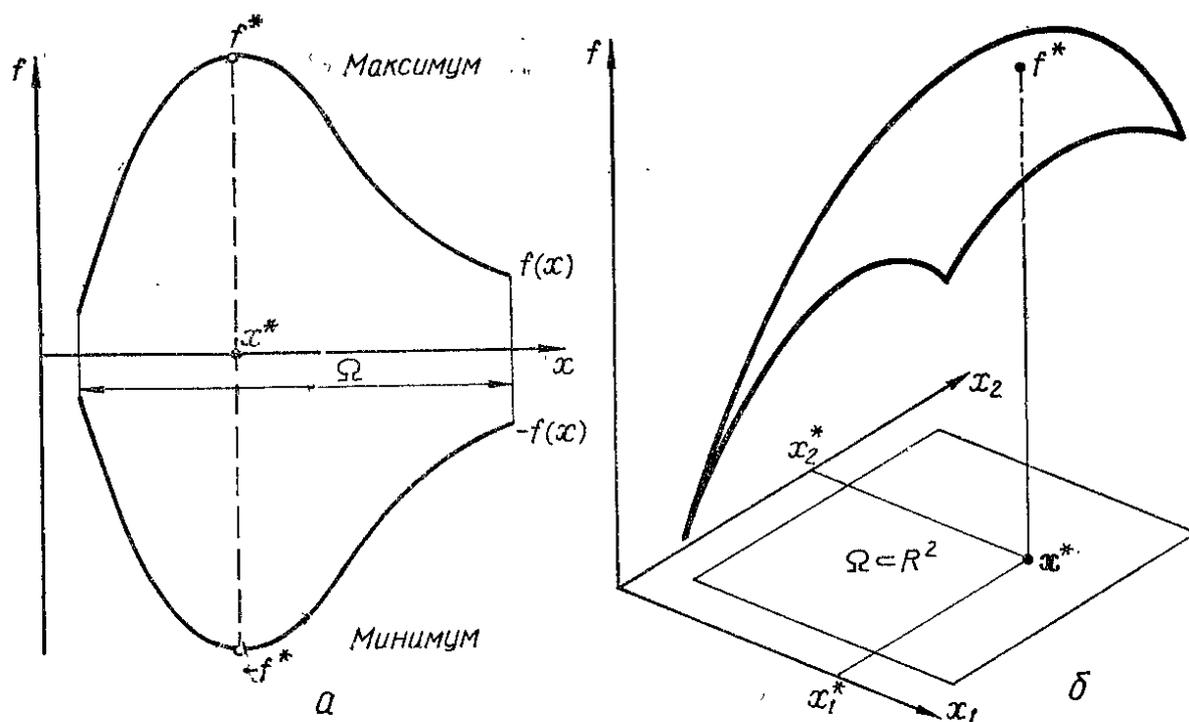


Рис. 11.1. Максимум и минимум однофакторной функции (а) и максимум двухфакторной функции (б) в области допустимых решений Ω

поскольку $\max f(x) = \min(-f(x))$, как показано на рис. 11.1, а для функции числа x .

Для того чтобы практическую задачу оптимизации привести к задаче (11.1), необходимо определить конкретное множество Ω допустимых решений x и указать конкретную целевую функцию $f(x)$. Особенности множеств Ω и виды целевых функций в разных задачах оптимизации лежат в основе их классификации и многообразия методов решения.

Если задача сформулирована так, что допустимые и оптимальные решения представляют собой число или группу k чисел x_i , то говорят о *параметрической оптимизации*. В этом случае множество решений Ω определяется в k -мерном евклидовом пространстве R^k , а его элемент — точка x — представляет k -мерный вектор оптимизируемых параметров $x = (x_1, x_2, \dots, x_k)$; $f(x)$ — принимающая значение в R^1 числовая функция числовых аргументов. В частности, для $k = 2$ эти представления иллюстрирует рис. 11.1, б на эллиптическом параболоиде, аналогичном рис. В. 4, в.

Такие задачи (в исследованиях, при проектировании, на производстве) формулируются во всех тех случаях, когда структура оптимизируемых систем задана: известен набор элементов конструктивной схемы или список компонентов смеси, зафиксирована последовательность технологических операций; выбран вид математической модели объекта или вид технологического оборудования и т.п. Необходимо определить оптимальные значения управляемых параметров заданных элементов систем: такие геометрические размеры конструктивных элементов изделий, при которых минимальна его масса; такие значения коэффициентов при элементах математической модели, чтобы рассчитанные по ней значения характери-

стик объекта как можно меньше отличались от истинных значений (см. п. III.7); содержания компонентов смеси, продолжительности технологических операций по ее переработке и режимы работы оборудования, обеспечивающие максимальную прочность готового материала, и т. п.

Пример II.1. Расход жидкости Q , м³/с, при установившемся истечении через кольцевой зазор между двумя концентрическими трубками с радиусами $R > r$ определяется формулой (ламинарное течение несжимаемой жидкости [52])

$$Q = K \{ R^4 - r^4 - (R^2 - r^2) / [\ln (R/r)] \},$$

где $K = \pi \Delta p / (8\mu)$; Δp — падение давления на единице длины; μ — вязкость жидкости. Требуется найти такие значения R и r в интервале $R_{\text{н}} \ll R \ll R_{\text{в}}$ и $r_{\text{н}} \ll r \ll r_{\text{в}}$, при которых расход жидкости через поперечное сечение заданной площади S максимален.

В этой конкретной задаче в роли компонентов двумерного вектора x выступают числовые параметры R и r ; $Q(R, r)$ — целевая функция. Множество допустимых значений параметров Ω в двумерном пространстве вещественных чисел ограничено границами интервалов для R и r , а также взаимосвязью этих факторов в элементарном геометрическом соотношении:

$$\pi(R^2 - r^2) = S.$$

Пример II.2. Изменение прочности на растяжение R , МПа, композиции для герметизации стыков конструкций в зависимости от содержания минерального $H_{\text{м}}$ и органического $H_{\text{о}}$ наполнителей, % к массе полимера, описывается моделью (см. рис. В.4, в)

$$R(x_1, x_2) = 8 + 2x_1 - 2x_1^2 + x_1x_2 - x_2 - x_2^2. \quad (\text{II.4})$$

Безразмерные нормализованные переменные $|x_i| \leq 1$ задают изменение содержания наполнителя в интервалах $0 \leq H_{\text{м}} = X_1 \leq 50$ и $0 \leq H_{\text{о}} = X_2 \leq 30$; натуральные X_i и нормализованные x_i факторы связаны соотношениями (В.2) — (В.4). Могут быть сформулированы оптимизационные задачи двух типов (см. «Введение»).

А. Требуется определить значения x_1 и x_2 (количество наполнителей X_1 и X_2 в композиции), при которых R максимальна, т. е. найти

$$x^* = (x_1^*, x_2^*) = \underset{x \in \Omega}{\operatorname{argmax}} R(x_1, x_2) = \underset{x \in \Omega}{\operatorname{argmin}} [-R(x_1, x_2)],$$

если область Ω определена записанными выше неравенствами — ограничениями на содержания компонентов x_i и X_i .

Б. Требуется определить x_1 и x_2 , которые обеспечивают минимальное содержание минерального наполнителя при условии, что прочность не ниже нормативного уровня $R_{\text{тр}} = 7$ МПа.

В этой постановке критерием оптимальности является одна из независимых управляемых переменных x_1 , а функция (II.4) задает дополнительное ограничение на область Ω допустимых значений x :

$$R(x_1, x_2) \geq 7 \text{ или } 7 - R(x_1, x_2) \leq 0.$$

Если задача оптимизации формулируется таким образом, что элементы множества Ω в (II.1), т. е. допустимые решения x , представляют собой функции (времени, координат в пространстве и пр.), а критерий оптимальности $f(x)$ является функционалом (числовой функцией на множестве функций), то это задача *вариационного исчисления* или *оптимального управления*. Так, в частности,

может быть сформулирована задача: определить оптимальную кривую скорости сушки [52] структурированного материала, которая обеспечивает его максимальную влагоотдачу без разрушения структуры. Подобный подход к оптимизации естественен при определении: оптимальной формы поверхностей; оптимальной по времени достижения цели или по энергетическим затратам траектории движения; оптимального, с точки зрения выхода конечного продукта, распределения температуры по длине реактора и т. д.

Задачи такого рода рассматриваются в специальных курсах [5, 19, 65]. В то же время они часто сводятся к *параметрической оптимизации*, которая осуществляется более простыми методами и рассматривается в данной главе. Термин «параметрическая» в дальнейшем опускается, поскольку речь идет только об оптимизации в пространстве R^k вещественных чисел — значений k управляемых параметров. Следует остановиться на основных особенностях перехода от конкретной содержательной задачи к ее математической формулировке, т. е. к записи целевой функции и области ее определения Ω .

Часто по сути задачи необходимо или желательно улучшить систему по нескольким зависящим от \mathbf{x} показателям качества $Y_j(\mathbf{x})$ — по *вектору критериев* $Y(\mathbf{x}) = (Y_1(\mathbf{x}), \dots, Y_j(\mathbf{x}), \dots)$. Так, в задаче пр. II.2 наряду с увеличением прочности целесообразно снизить модуль упругости композиций. Как уже отмечено во введении и отражено в общей постановке задачи оптимизации, подобные пожелания должны быть свернуты к одному *скалярному критерию оптимальности* $f(\mathbf{x})$, поскольку только в этом случае возможно однозначное сравнение решений.

В таких «многокритериальных» ситуациях в качестве целевой функции чаще всего назначают один из показателей качества (*частный критерий*), а требования по остальным показателям учитывают в виде функциональных ограничений для множества допустимых значений управляемых параметров, таких как нормативное ограничение по прочности в пр. II.2, Б. В частности, дополнительное требование к модулю упругости композиций может быть выражено ограничением вида $E(\mathbf{x}) \leq E_{\text{тр}}$.

В некоторых задачах формируется *комплексный критерий оптимальности* [26, 39, 43] — новая функция от \mathbf{x} , включающая все или часть частных критериев $Y_j(\mathbf{x})$. Так, желание увеличить R и одновременно снизить E может быть отражено в записи взвешенной аддитивной целевой функции $f(\mathbf{x}) = a_E E(\mathbf{x}) - a_R R(\mathbf{x})$, где a_E и a_R — положительные весовые коэффициенты, или функцией $f(\mathbf{x}) = E(\mathbf{x})/R(\mathbf{x})$, характеризующей жесткость материала. Однако в такой постановке задачи недостаточно полно учитываются требования, предъявляемые к отдельным показателям качества.

Построение подобных *интегральных критериев* и назначение в них коэффициентов таким образом, чтобы нарушение требований по одному показателю не нивелировалось за счет улучшения другого, представляют собой отдельную сложную проблему, которая рассматривается специальным научным направлением — *квалимет-*

рией. В инженерной практике более целесообразно, особенно при работе в диалоге с ЭВМ (см. п. IV.5), назначать частные критерии и ограничения, естественно следующие из конкретных требований к объекту оптимизации (см. далее пр. II.5).

Выбранные в соответствии с содержанием разных задач критерии оптимальности могут характеризоваться *разной степенью определенности* зависимости $f(\mathbf{x})$. Целевая функция может быть однозначно определена на множестве Ω точной формулой, как в пр. II.1. Если связь между показателями качества объекта Y_j и управляемыми параметрами \mathbf{x} статистическая, то значения критерия оптимальности и показателей, по которым область решений ограничивается, определяются как математические ожидания $E(Y_j/\mathbf{x})$ (см. прил. I.4); зависимость $f(\mathbf{x})$ и функциональные ограничения в таких случаях могут быть заданы *экспериментально-статистическими моделями* (которые подробно рассматриваются в гл. III). Так, в частности, в пр. II.2 задана зависимость прочности композиций от концентрации наполнителей.

Задача оптимизации формулируется и в том случае, если вид функции $f(\mathbf{x})$ *неизвестен*, но ее значение может быть определено для любого аргумента $\mathbf{x} \in \Omega$: либо в эксперименте на реальном объекте (пр. II.3), либо в численном эксперименте (в результате численного решения некоторой математической задачи, при реализации ситуации на имитационной модели — см. гл. IV).

Пример II.3. Время τ набора заданной начальной прочности бетонных изделий, а следовательно, и производительность линии по их изготовлению определяются температурой T тепловлажностной обработки изделий. Вид функции $\tau(T)$ зависит от используемого сырья (от активности цемента, качества минеральных добавок и т. д.). Требуется, получив очередную партию сырья, так назначить T из интервала $T_n \leq T \leq T_v$, чтобы продолжительность операции была как можно меньше при выполнении требований к качеству изделий, т. е. необходимо определить

$$T^* = \operatorname{argmin}_{T \in \Omega_T} \tau(T).$$

Поскольку формула для расчета τ неизвестна, определение τ для ряда значений T с целью сравнения и выбора лучшего варианта проводится экспериментально на опытных образцах.

Другой обязательный элемент постановки задачи оптимизации — множество допустимых решений Ω — определяется разными *ограничениями* на значения, которые могут принимать управляемые параметры \mathbf{x} . Эти ограничения, как правило, выражают условия, связанные с физической сущностью конкретной задачи. Выделяют три группы ограничений, которые могут иметь место по отдельности или в сочетании друг с другом: прямые ограничения, функциональные типа равенств и типа неравенств.

Прямые, или факторные, ограничения имеют вид

$$x_{н. i} \leq x_i \leq x_{в. i} \quad (i = \overline{1, k}), \quad (\text{II.5})$$

где $x_{н. i}$ и $x_{в. i}$ — наименьшее (нижнее) и наибольшее (верхнее) возможные значения i -й переменной. Так, в частности, для многих

объектов оптимизации значения управляемых параметров не могут быть отрицательными, т. е. $X_{n,i} > 0$ (геометрические размеры, доли компонентов смеси, стоимости и т. д. — см. «Введение»).

Функциональные ограничения-равенства имеют вид

$$h(x_1, x_2, \dots, x_i, \dots, x_k) = 0 \quad (\text{II.6})$$

и, как правило, отражают объективно существующие взаимосвязи между параметрами (см. пр. II.1). Если соотношение типа (II.6) удастся разрешить относительно одной из переменных x_i , то ее можно исключить из задачи, уменьшив размерность пространства параметров с k до $k - 1$.

Пример II.4. При изготовлении цилиндрической емкости (резервуара без крышки) заданного объема V необходимо так выбирать высоту H и радиус r цилиндра, чтобы площадь поверхности емкости была минимальна (для уменьшения теплоотдачи), т. е. требуется решить задачу

$$\begin{aligned} S(r, H) &= \pi r^2 + 2\pi rH \rightarrow \min; \\ 0 < r < \infty, 0 < H < \infty, &\text{ или просто } r > 0, H > 0; \\ \pi r^2 H - V &= 0. \end{aligned}$$

Если выразить высоту через радиус $H = V / (\pi r^2)$ и подставить полученное выражение в формулу для S , то можно от исходной двухпараметрической задачи оптимизации с одним ограничением-равенством перейти к более простой одномерной задаче без функциональных ограничений:

$$S(r) = \pi r^2 + 2V / r \rightarrow \min; \quad r > 0.$$

Функциональные ограничения-неравенства записываются в виде

$$g(x_1, x_2, \dots, x_i, \dots, x_k) \leq 0 \quad (\text{II.7})$$

и выражают, как правило, требования к функционированию оптимизируемой системы, не отраженные в целевой функции. Так, в пр. II. 2, Б ограничение-неравенство выражает условие, при котором прочность композиций не ниже нормативного уровня. Очевидно, что неравенство типа $E(\mathbf{x}) \leq E_{\text{тр}}$ также приводится к общему виду (II.7): $E(\mathbf{x}) - E_{\text{тр}} \leq 0$.

Иногда задача оптимизации формулируется таким образом, что элементы вектора \mathbf{x} принимают дискретные (в частности, целочисленные) значения, и область определения целевой функции Ω является дискретным множеством точек. Однако в большинстве строительно-технологических задач рассматриваются непрерывные параметры и Ω — непрерывная область в k -мерном пространстве вещественных чисел R^k .

В каждой точке пространства R^k целевая функция $f(\mathbf{x})$ и ограничивающие функции $h(\mathbf{x})$ и $g(\mathbf{x})$ принимают определенные значения. Следовательно, каждая из таких функций представляет в этом пространстве скалярное поле (критерия оптимальности или ограничивающих показателей — см. прил. I.1), которое может быть отражено семейством изоповерхностей (см. «Введение»). В пространстве двух переменных R^2 , которое отображается плоскостью прямоугольных координат (x_1, x_2) , функции f , h и g удобно представить совокупностью изолиний (см. «Введение»). Такие геометрические

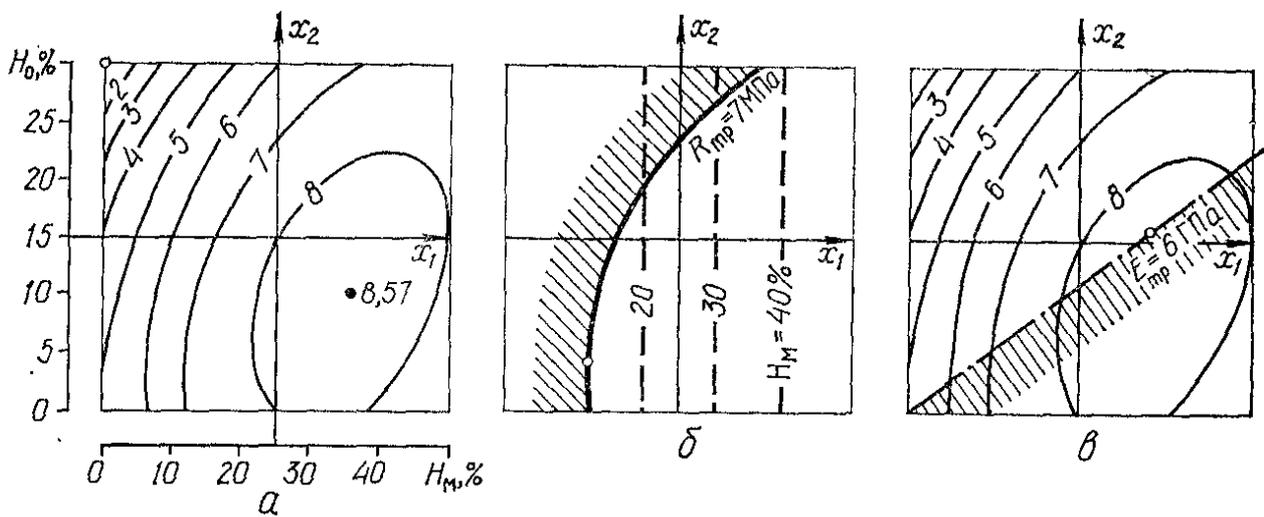


Рис. 11.2. Геометрическая интерпретация задач оптимизации:

a — изолинии прочности с точкой $R_{\max} = 8,57$ МПа; *б* — изолинии содержания минерального наполнителя H_M с ограничением $R \geq 7$ МПа; *в* — изолинии R с ограничением $E \leq 6$ ГПа

представления условий задач оптимизации оказываются полезными при анализе методов их решения и самих решений.

При наличии в постановке задачи ограничения-равенства точки, соответствующие допустимым решениям, находятся на линии $h(x_1, x_2) = 0$, а при нескольких h_i — в точках пересечения таких линий. Ограничению-неравенству удовлетворяют точки по одну сторону от линии $g(x_1, x_2) = 0$. Прямые ограничения отображаются прямыми, параллельными координатным осям. Таким образом, совокупность линий, соответствующих всем ограничениям задачи, выделяет на плоскости область допустимых решений Ω — пересечение областей, разрешенных отдельными ограничениями. Если и прямые, и функциональные ограничения отсутствуют, область Ω совпадает со всей плоскостью R^2 .

На рис. 11.2, *a*, *б* представлена геометрическая интерпретация задач оптимизации, сформулированных в пр. 11.2. В плоскости координат независимых переменных x_1 и x_2 показан квадрат $|x_1| \leq 1, |x_2| \leq 1$ — область допустимых решений задачи А (концентраций наполнителей); на квадрате нанесены изолинии критерия оптимальности — прочности $R(x_1, x_2)$ на рис. 11.2, *a* (соответствует поверхности отклика на рис. В. 4, *в*).

В задаче Б критерий оптимальности $f(\mathbf{x}) = x_1$ (содержание минерального наполнителя) отображается набором прямых, перпендикулярных этой оси (на рис. 11.2, *б* указаны соответствующие концентрации наполнителя — уровни фактора X_1); область решений «уменьшилась» по сравнению с задачей А — линия $R(x_1, x_2) = 7$ отсекает от квадрата зону, запрещенную требованием $R \geq 7$ МПа. Если «ужесточить» условия задачи, потребовав, чтобы прочность была равна 7, то только точки на этой линии будут соответствовать допустимым решениям.

Пример 11.5. К условиям пр. 11.2, А добавлено требование — модуль упругости композиций должен быть не более 6 ГПа. В математической формулировке это выражено неравенством

$$E(x_1, x_2) = 5,5 + 1,5 x_1 - 2x_2 \leq 6.$$

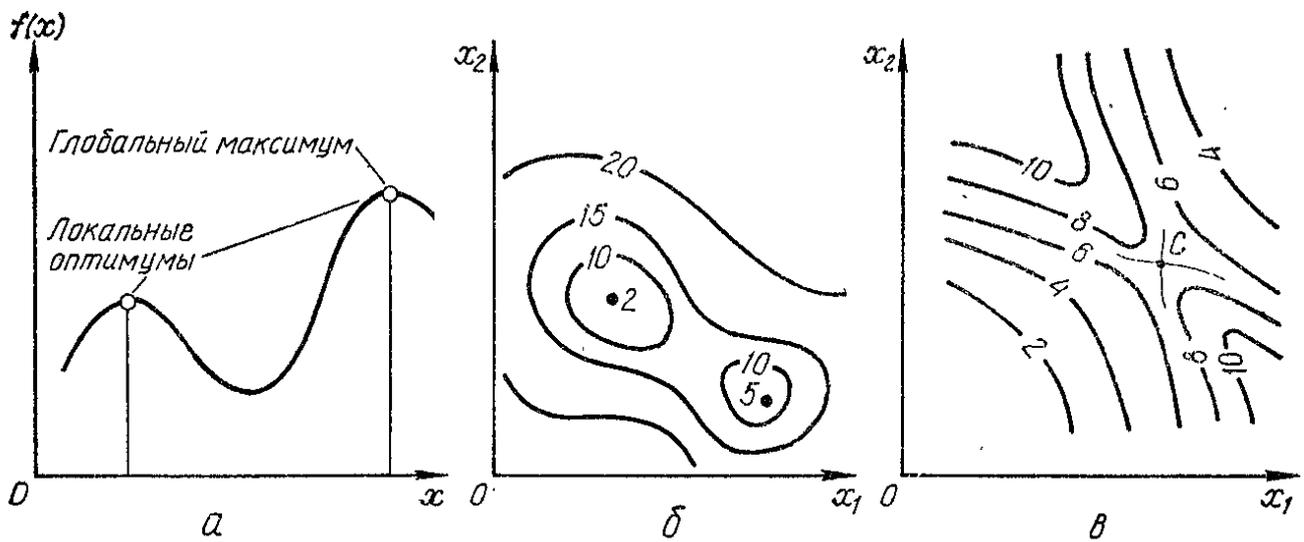


Рис. 11.3. Функции однофакторная двухэкстремальная (а), двухфакторная двухэкстремальная (б) и седловая гиперболоидная (в)

Геометрическая интерпретация этой задачи оптимизации:

$$\begin{aligned} R(x_1, x_2) &\rightarrow \max; \\ -1 \leq x_1 \leq 1, \quad -1 \leq x_2 \leq 1; \\ E(x_1, x_2) - 6 &\leq 0 \end{aligned}$$

показана на рис. 11.2, в. От квадрата, образованного прямыми ограничениями концентраций компонентов, отсечен изолинией $E(x_1, x_2) = 6$ правый нижний угол. Оставшаяся часть соответствует области допустимых решений, в которой и должно быть определено оптимальное решение задачи.

Геометрическая интерпретация задачи оптимизации, выявляя характер изменения целевой функции в области ее определения, дает представление и о зоне оптимального решения — о положении точки экстремума в допустимой области Ω , которое связано со следующими известными [21] математическими понятиями и условиями.

Точка x^* называется *точкой локального минимума* функции $f(x)$, если неравенство $f(x^*) \leq f(x)$ (для максимума $f(x^*) \geq f(x)$) справедливо для всех тех точек x из области Ω , которые находятся в окрестности точки x^* (см. понятие окрестности в п.1.2). Если неравенство выполняется для любых x из Ω , то точку x^* называют *глобальным экстремумом*. Функция $f(x)$ *одноэкстремальна* (унимодальна), если она имеет один экстремум (см. рис. 11.1), и *многоэкстремальна* (рис. 11.3), если она имеет более одного минимума (максимума). Глобальным экстремумом является точка (рис. 11.3), в которой целевая функция имеет наименьшее (наибольшее) значение среди всех локальных экстремумов.

Если функция $f(x)$, по крайней мере, дважды дифференцируема и область ее определения не ограничена (все пространство R^k), то для того чтобы точка x^* была точкой локального экстремума функции $f(x)$, необходимо, чтобы все ее частные производные по компонентам x_i вектора x в этой точке были равны нулю:

$$\left. \frac{\partial f(x)}{\partial x_i} \right|_{x^*} = 0 \quad (i = \overline{1, k}). \quad (11.8)$$

Условие (11.8) есть *условие стационарности* точки x^*

Градиент функции $f(\mathbf{x})$ в точке (x_1, \dots, x_k) — k -мерный вектор

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_k} \right), \quad (\text{II.9})$$

элементы которого равны частным производным функции $f(\mathbf{x})$, вычисленным по k ее аргументам (см. прил. I.1). Длина этого вектора определяет скорость изменения функции, а направление совпадает с направлением наибольшего ее возрастания. Условие того, что точка \mathbf{x}^* является стационарной точкой функции $f(\mathbf{x})$, записывается с помощью градиента

$$\nabla f(\mathbf{x}^*) = 0. \quad (\text{II.10})$$

Равенство нулю градиента является необходимым условием экстремума в точке \mathbf{x}^* . Характер стационарной точки и достаточное условие экстремума функции одной переменной определяются знаком второй производной. Характер стационарной точки многомерной функции определяется матрицей вторых производных — *гессианом* [5, 64]. Если в точке \mathbf{x}^* матрица определена положительно (см. прил. I.1), то здесь находится локальный минимум функции, если отрицательно — максимум; в остальных случаях точка \mathbf{x}^* — «седловая» (точка С на рис. II.3, в).

Если область Ω допустимых значений управляемых параметров \mathbf{x} ограничена (что характерно для большинства практических ситуаций), то в такой оптимизационной (экстремальной) задаче определяется *условный оптимум* (условный экстремум). Он находится либо в стационарной точке внутри области, либо в точке на границе (она может оказаться стационарной лишь в исключительных случаях).

Так, в пр. II.2, Б линейная целевая функция $f(\mathbf{x}) = x_1$ не имеет стационарных точек. Как видно на рис. II.2, б, условный экстремум находится на границе допустимой области, сформированной функциональным ограничением по прочности композиций. Такие ограничения, обуславливающие оптимум, называются *активными*. Если в задаче II.2, Б снять требование к прочности, то активным станет прямое ограничение $x_1 \geq -1$, т. е. минимальное содержание наполнителя x_1 будет обусловлено нижней границей его концентрации $x_1 = -1$ ($X_1 = X_{1н} = 0$).

В задаче II.5 $\max R(\mathbf{x} \in \Omega)$ будет, очевидно (см. рис. II. 2, в), находиться на линии $E(\mathbf{x}) = 6$. Если это ограничение по модулю упругости снято (пр. II.2, А), то, как показано на рис. II, 2, а, точка максимума R находится внутри области, ограниченной предельными концентрациями, и совпадает с точкой *безусловного, абсолютного экстремума* функции R (II.4). Ее координаты могут быть определены из условия стационарности (II.8) как решение системы уравнений

$$\left. \begin{aligned} \partial R(x_1, x_2)/\partial x_1 &= 2 - 4x_1 + x_2 = 0; \\ \partial R(x_1, x_2)/\partial x_2 &= x_1 - 1 - 2x_2 = 0; \end{aligned} \right\}$$

$$x_1^* \approx 0,43; \quad x_2^* \approx -0,29.$$

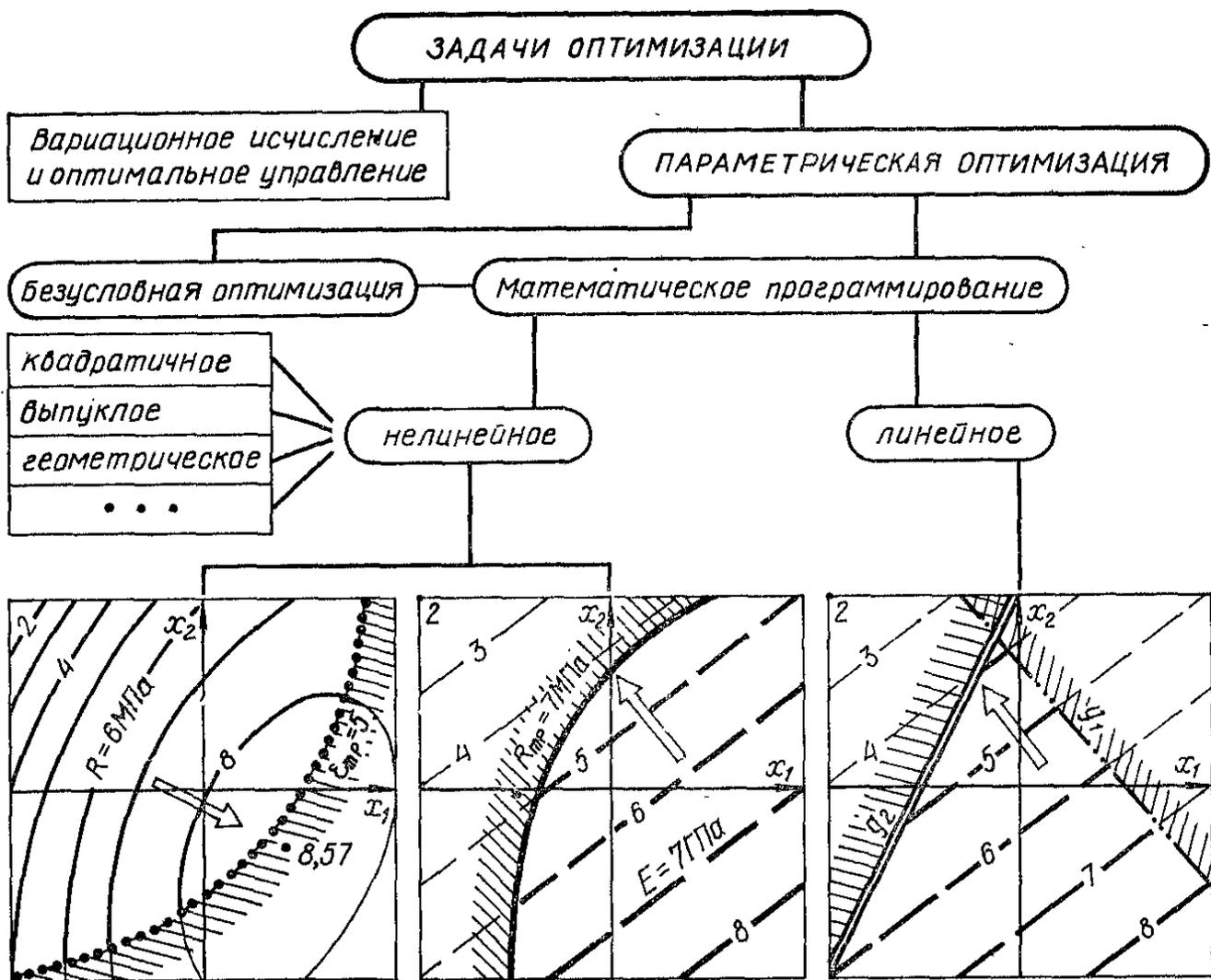


Рис II.4. Классификация задач оптимизации и три диаграммы, иллюстрирующие особенности задач нелинейного и линейного программирования

Очевидно, что подобный способ решения оптимизационной задачи, заключающийся в определении стационарных точек и сопоставлении значений целевой функции в этих точках с ее значениями на границах области определения, применим лишь в простейших случаях (несложная целевая функция небольшого числа аргументов без ограничений или при наличии самых простых ограничений). Так, в пр. II.4 безусловный минимум функции одной переменной $S(r)$ легко определяется при анализе производной

$$S'(r) = 2(\pi r^3 - V)/r^2.$$

Она обращается в нуль при $r^* = \sqrt[3]{V/\pi}$, отрицательна в интервале $0 < r < r^*$ (где функция S убывает), положительна при $r^* < r < \infty$ (где $S(r)$ возрастает). Следовательно, точка r^* есть решение задачи (пр. II.4).

Рассматриваемая классификация задач оптимизации в целом представлена на рис. II.4. В зависимости от характера области допустимых решений Ω и вида целевой функции $f(\mathbf{x})$ все многообразие задач параметрической оптимизации делится на следующие (показанные на схеме) классы.

Если в выражении (II.1) область Ω совпадает с пространством R^k ($\Omega = R^k$), то это задача безусловной оптимизации — на аргументы

функции $f(x_1, \dots, x_k)$ не накладывается никаких дополнительных условий (ограничений). Несмотря на то, что такая ситуация не характерна для реальных систем, в том числе в строительном материаловедении и технологии, безусловный оптимум приходится находить в процессе решения экстремальных задач с ограничениями. Кроме того, для последних существуют методы (см. п. II.4), которые сводят задачи с ограничениями к задачам безусловной оптимизации.

Если в формулировке (II.1) область Ω пространства R^k ограничена прямыми и (или) функциональными ограничениями (II.5) — (II.7) в виде равенств и неравенств, то имеет место *задача математического программирования*, которая в общем случае записывается так:

$$f(\mathbf{x}) \rightarrow \min; \mathbf{x} = (x_1, \dots, x_k)^T \in \Omega \subset R^k,$$

$$\text{где } \Omega = \{ \mathbf{x} : h_j(\mathbf{x}) = 0, j = \overline{1, p}; \\ g_j(\mathbf{x}) \leq 0, j = \overline{1, q} \}. \quad (\text{II.11})$$

Среди задач этого класса различают задачи с ограничениями типа неравенств, когда $p = 0$, с ограничениями типа равенств, когда $q = 0$, задачи со смешанными ограничениями — $p \neq 0, q \neq 0$; функции h и g в (II.11) обобщают прямые и функциональные ограничения значений управляемых параметров. Так, в пр. II.1 область Ω допустимых значений параметров $x_1 = R$ и $x_2 = r$ задается одним ограничением-равенством, которое может быть записано в виде (II.6)

$$h(x_1, x_2) = \pi(x_1^2 - x_2^2) - S = 0,$$

и четырьмя неравенствами вида (II.7)

$$g_1(x_1, x_2) = R_H - x_1 \leq 0, g_2(x_1, x_2) = x_1 - R_B \leq 0, \\ g_3(x_1, x_2) = r_H - x_2 \leq 0, g_4(x_1, x_2) = x_2 - r_B \leq 0,$$

соответствующими прямым ограничениям. В пр. II.2 — II.5 сформулированы задачи с ограничениями типа неравенств.

К формулировке (II.11) сводится большинство практических задач оптимизации технических объектов, включая оптимизацию технологии и свойств строительных материалов и изделий.

Задачи математического программирования далее классифицируются в зависимости от свойств функций, участвующих в их формулировке. Такое деление обусловлено тем, что учет особенностей задачи позволяет решить ее наиболее эффективным способом. Этот принцип разделил задачи математического программирования прежде всего на два класса: задача (II.11) является *задачей нелинейного программирования*, если хотя бы одна из функций $f(\mathbf{x}), h(\mathbf{x}), g(\mathbf{x})$ нелинейная; если же каждая из этих функций линейна, то имеет место *задача линейного программирования* (см. рис. II.4).

Методы решения задач линейного программирования наиболее разработаны. Они используются при оптимизации стоимостных и других аддитивных показателей. На рис. II.4 этот класс задач

иллюстрируется геометрическим отображением: линейной целевой функции $E(\mathbf{x})$ — модуля упругости из пр. II.5 в качестве минимизируемого критерия оптимальности; ограничения на общее количество наполнителя $g_1 = 9x_1 + 8x_2 - 5 \leq 0$; ограничения на соотношение фракций $g_2 = -2x_1 + x_2 - 1 \leq 0$ (органический наполнитель нецелесообразно использовать без минерального). Оптимальные решения в таких задачах с линейной целевой функцией всегда находятся на границе областей допустимых решений, представляющих собой многогранники в k -мерном пространстве. Постановки, методы и примеры решения задач линейного программирования рассмотрены в п. II.5 — II.7.

Большинство задач оптимизации в материаловедении и технологии формулируется как задачи нелинейного программирования. Методы их решения рассматриваются в п. II.2 — II.4. Для примера на рис. II.4 показаны условия: а) максимизации нелинейной функции (II.4) — прочности композиции (см. пр. II.2 и рис. В.4, в), нелинейно ограниченной требованием к их относительным деформациям $g(\mathbf{x}) = \varepsilon(\mathbf{x}) \geq \varepsilon_{\text{тр}}$; б) минимизации линейной функции — модуля упругости из пр. II.5 и рис. В.4, б) — при нелинейном ограничении по прочности $R(\mathbf{x}) \geq R_{\text{тр}}$. Случай нелинейной целевой функции, но линейных ограничений, представлен непосредственно в пр. II.5 (см. рис. II.2, в).

Во всем многообразии задач нелинейного программирования с разными сочетаниями разных видов целевых и ограничивающих функций выделяют еще ряд специальных классов задач, для решения которых созданы специальные методы, учитывающие особые свойства целевых функций и допустимых областей: квадратичное программирование, выпуклое программирование и др. [64].

После того как содержательная задача оптимизации сформулирована математически в конкретизированной форме (II.11), необходимо решить ее тем или иным методом. Выбор метода зависит от конкретных особенностей постановки задачи.

II.2. Основные этапы и классификация методов поиска экстремума

В ряде случаев математическая постановка экстремальной задачи позволяет решить ее аналитически. *Аналитические методы* решения задач оптимизации заключаются в определении экстремума функции $f(\mathbf{x})$ путем нахождения тех значений \mathbf{x} , которые обращают в нуль производные $f'(\mathbf{x})$ по x_i (см. условие стационарности (II.8) в п. II.1). Эти методы применимы лишь тогда, когда: а) все участвующие в задаче функции заданы в явном виде; б) они дважды дифференцируемы; в) свойства функций и их производных можно определить по известным формулам за приемлемое время. Так, расчет координат максимума или минимума квадратичной параболы лежит в основе диссоциативно-шагового метода оптимизации по полиномиальным моделям второго порядка на единичном гиперкубе (см. п. III.15).

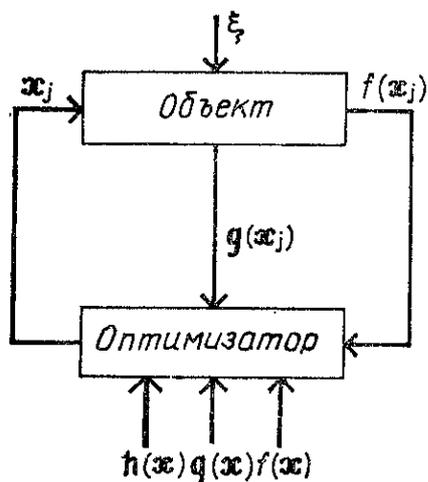


Рис. II.5. Структурная схема системы поисковой оптимизации

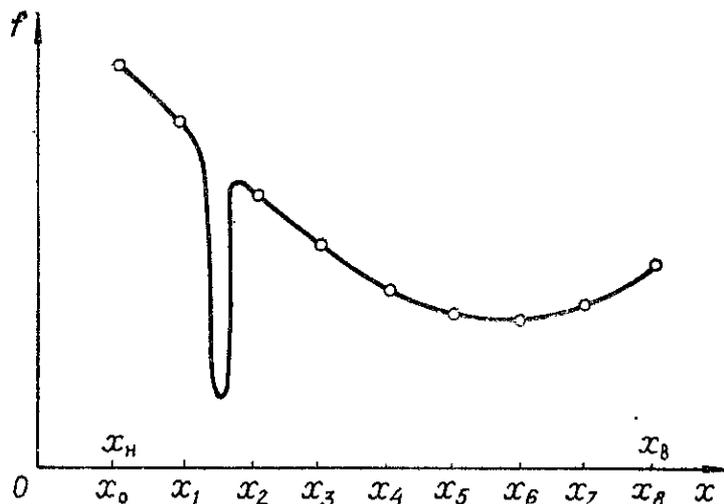


Рис. II.6. Пропуск «языка» функции при поиске минимума методом сканирования

Для решения большинства практических задач оптимизации аналитические методы оказываются непригодными, а используются *численные методы* поиска экстремума (методы поисковой оптимизации), основанные на последовательном (итерационном) получении и анализе информации о численных значениях целевой функции в отдельных точках области решений. Оценки значений $f(x)$ получают расчетным путем или экспериментально, в зависимости от степени определенности целевой функции (см. п. II.1).

Если на рис. В.1 конкретизировать цель управления как достижение минимума критерия функционирования объекта управления, то можно дать следующее определение. *Поиск* оптимума — это последовательность действий (шагов), выполняемых оптимизатором (блоком управления на рис. В.1) по сбору информации об объекте управления и принятию решений, направленных на достижение цели. На рис. II.5 показана структурная схема системы поисковой оптимизации.

Блок «объект» может представлять и физические объекты, и их математические модели. Если в результате действия случайных помех ξ выходы объекта принимают случайные значения, то в качестве соответствующих функций g и f используются условные математические ожидания. Оптимизатор реализует алгоритм поиска — правило выбора последовательности действий. Он вырабатывает последовательность решений (воздействий на объект, состояний объекта) $x_1, x_2, \dots, x_j, \dots$, сходящуюся к x^* (II.2), в общем случае, на основе результатов экспериментов — измерений откликов $f(x_j)$ и $g(x_j)$ на объекте или вычислений по математическим моделям.

Поиск экстремума можно сравнить [67] с поиском самого глубокого места в озере. При каждом замере лотом получают новую информацию. Если в текущем замере глубина оказалась больше, чем в предыдущем, то полученная информация полезна. Если же глубина в последующем замере меньше, чем в предыдущем, то можно считать, что результат не несет полезной информации и затра-

ценные на его получение средства были напрасными. Разрабатывая методы поиска или выбирая метод, подходящий для решения конкретной задачи, стремятся найти экстремум как можно быстрее (с наименьшими затратами), сделав как можно меньше, «бесполезных» попыток.

Методы поиска оптимума классифицируют по нескольким признакам. В зависимости от типа разыскиваемого экстремума различают методы локального и глобального поиска, условной и безусловной оптимизации.

Большинство используемых на практике методов является *методами локального поиска*. Надежные и в то же время достаточно экономичные *методы глобального поиска* для решения реальных задач оптимизации не разработаны. Глобальный экстремум обычно определяется лучшей из нескольких оценок локальных экстремумов.

Как один из самых простых и естественных методов заслуживает упоминания *метод сканирования* (перебора). Метод заключается в последовательном переборе N решений (состояний) \mathbf{x}_j из множества допустимых, для которых определяются (вычисляются или измеряются) значения целевой функции; запоминается наименьшее значение

$$f^* = \min f(\mathbf{x}_j) \quad (j = \overline{1, N}) \quad (\text{II.12})$$

и соответствующее решение \mathbf{x}^* .

Если область Ω непрерывна, то она разбивается на N подобластей, и значения f определяются в фиксированных точках каждой подобласти. Чтобы получить достоверную картину поведения функции, например не пропустить «язык» функции (рис. II.6), необходимо сканировать допустимую область с достаточно малым шагом по каждой переменной x_i . Если для функции одной переменной необходимо $N=n$ ее определений, то в k -мерном случае число N порядка k^n . Это число резко растет с ростом k и n . Поэтому метод чаще всего применяют для оптимизации дискретных систем с небольшим числом допустимых состояний и в однопараметрических экстремальных задачах.

Необходимость полного перебора фиксированных состояний системы — недостаток такого «слепого» поиска. Однако он имеет и достоинства — простоту алгоритма поиска и его независимость от характера и вида целевой функции, а также принципиальную возможность всегда найти глобальный экстремум.

Методы безусловной оптимизации используются в том случае, когда в постановке задачи отсутствуют прямые и функциональные ограничения или когда к такой постановке может быть сведена (см. п. II.4) исходная задача нелинейного программирования (II.11). При поиске оптимума (и безусловного, и условного) строится последовательность шагов от некоторой начальной точки \mathbf{x}_0 в допустимой области через промежуточные точки \mathbf{x}_s к экстремуму \mathbf{x}^* , как правило, локальному. Глобальный оптимум оценивается за счет многократного поиска из разных начальных приближений \mathbf{x}_0 . На рис. II.7 показаны схема (основные этапы) такого обоб-

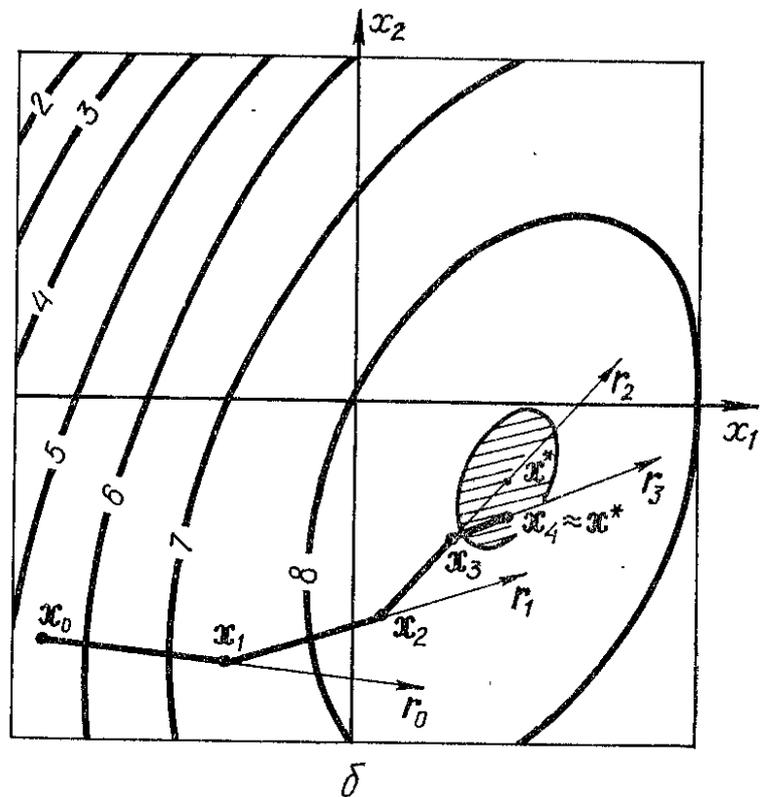
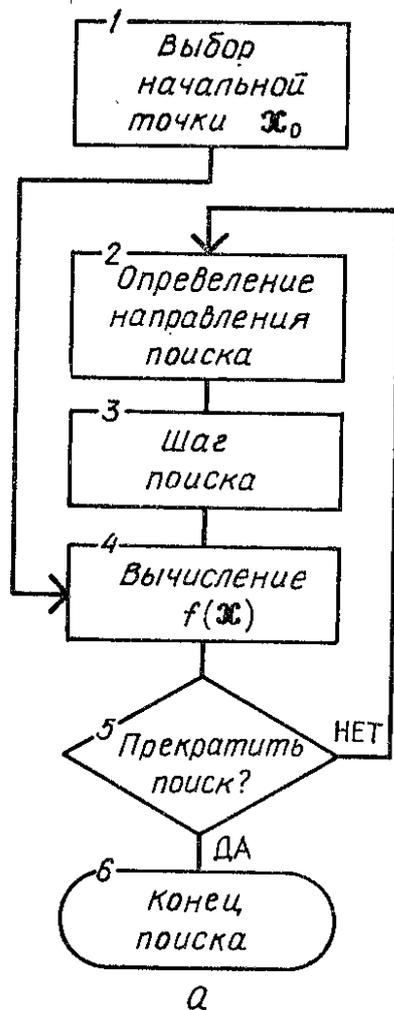


Рис. II.7. Схема (а) и траектория (б) обобщенного метода поиска оптимума

щенного метода [39] и графическое отображение поиска в двумерном случае, когда в качестве целевой функции $f(\mathbf{x})$ взята функция (II.4) со знаком минус.

Задача поисковой оптимизации заключается в отыскании такого пути, двигаясь по которому, точка \mathbf{x} пришла бы от \mathbf{x}_0 в достаточно малую ε -окрестность цели \mathbf{x}^* , в которой выполняется условие

$$f(\mathbf{x}) < f(\mathbf{x}^*) + \varepsilon. \quad (\text{II.13})$$

Величина ε — малое число, заданное, исходя из инженерных условий или точности поиска оптимума (на рис. II.7, б заштрихована область $f(\mathbf{x}) < -8,5$ или $R(\mathbf{x}) > 8,5$ МПа, т. е. $\varepsilon < 0,1$ МПа). Выбор начального приближения к оптимуму (этап 1) в значительной степени определяет скорость и вероятность его достижения. Очередной s -й шаг $\Delta \mathbf{x}_s$ из некоторой точки \mathbf{x}_s на траектории поиска (начиная от \mathbf{x}_0) осуществляется в направлении вектора $\mathbf{r}_s = (r_{1s}, \dots, r_{is}, \dots, r_{ks})$ так, что каждая координата x_{is} получает приращение

$$\Delta x_{is} = \alpha_s r_{is}. \quad (\text{II.14})$$

Этапы 2 и 3 алгоритма поиска — выбор направления \mathbf{r}_s движения к оптимуму и скалярной величины α_s для вычисления координат очередной точки

$$\mathbf{x}_{s+1} = \mathbf{x}_s + \alpha_s \mathbf{r}_s \quad (\text{II.15})$$

определяют сущность того или иного метода поиска.

При выборе направления движения и величины шага используется та или иная норма, т. е. расстояние между двумя точками (см. прил. I.1 и п. I.1). В связи с этим все управляемые параметры x должны иметь одинаковую размерность или быть безразмерными. Возможны разные способы нормировки. Один из наиболее распространенных способов — переход (B.2) — (B.4) от натуральных переменных X_i к нормализованным x_i .

В процессе поиска собирается информация об успехе движения к оптимуму. Значения целевой функции $f(x_s)$ и функций-ограничений $g(x_s)$, вычисляемые на этапе 4, обуславливают следующий шаг. Количество вычислений целевой функции n , выполненных за s шагов, служит характеристикой ресурсов, затраченных на сбор информации о движении к цели ($n \geq s$).

Поиск продолжается или прекращается в точке x_s в зависимости от выполнения проверяемых на этапе 5 назначенных условий остановки. В частности, поиск заканчивается при достижении заданной точности, когда для x_s выполняется условие (II.13). В реальных задачах истинное значение $f(x^*) = f^*$ неизвестно и, следовательно, трудно оценить сходимость алгоритма (см. п. I.2) и истинную, достигнутую на s -м шаге, точность

$$\varepsilon_s = |f(x_s) - f(x^*)|. \quad (\text{II.16})$$

Поэтому о достижении заданной точности судят по косвенным признакам, в частности по выполнению требования

$$|f(x_{s-1}) - f(x_s)| \leq \varepsilon. \quad (\text{II.17})$$

Число n_ε выполненных для этого вычислений f служит оценкой скорости сходимости алгоритма.

Условие остановки может определяться также имеющимися в наличии вычислительными ресурсами. В частности, может быть ограничено число вычислений функции f (например, $n \leq 1000$), при этом оценивается достигнутая точность ε_n .

Оценки скорости сходимости (n_ε) и точности (ε_n) характеризуют эффективность тех или иных методов поисковой оптимизации. Эффективность поиска связана с видом целевой функции и допустимой области, с размерностью задачи, а также с разрядностью ЭВМ, на которой реализуются вычисления, и с другими факторами. Поэтому однозначные рекомендации по выбору метода затруднительны. Однако очевидно, что выбор метода решения задачи нелинейного программирования должен осуществляться с учетом ее конкретной постановки (характера функций f и g , их числа, размерности пространства поиска, требований точности и «стоимости» решения и других условий — [64, 67 и др.]).

Наиболее эффективным инструментом для решения сложных оптимизационных задач представляется *программная система*, которая содержит библиотеку модулей, реализующих разные методы оптимизации, что позволяет инженеру-технологу адаптировать (приспосабливать) алгоритм поиска к текущим условиям

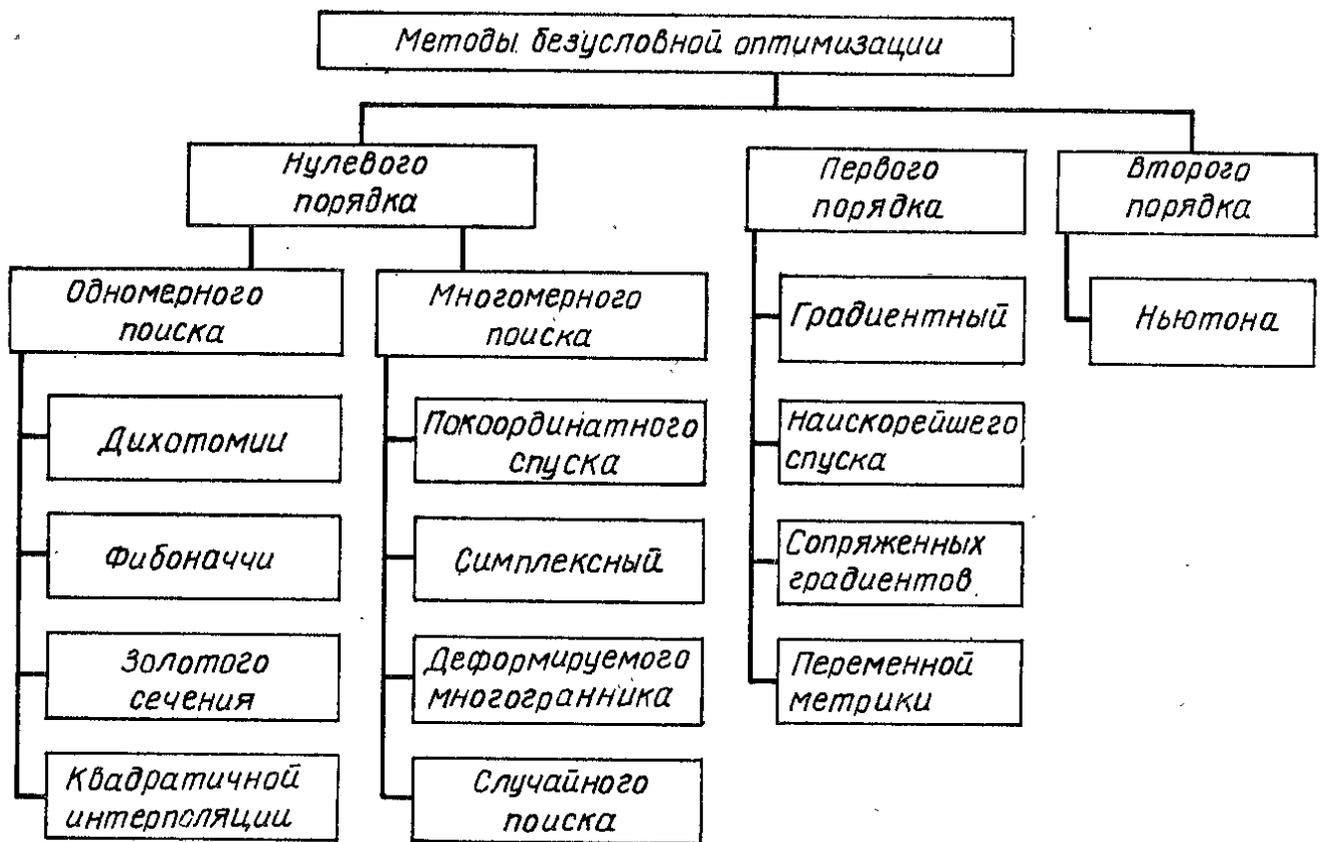


Рис. II.8. Классификация методов безусловной оптимизации

задачи в процессе ее решения: либо с использованием автоматизированного пакета программ, либо в режиме диалога с ЭВМ (см. п. IV.4).

Класс методов безусловной оптимизации наиболее многочислен (рис. II.8). Эти методы делят на методы нулевого, первого и второго порядков, в зависимости от порядка производных целевой функции, используемых при выборе направления движения к экстремуму. В методах *нулевого порядка*, или прямых методах, используется информация только о значениях целевой функции. Методы *первого порядка*, или градиентные, используют и значения целевой функции, и оценки ее первых производных (II.9) по управляемым параметрам. В методах *второго порядка* поиск экстремума ведется с учетом значений целевой функции, а также оценок ее первых и вторых производных.

Выделяют методы *одномерного* и *многомерного поиска*. Методы одномерного поиска помимо своего очевидного назначения для решения задач однопараметрической оптимизации используются как часть процедур многомерного поиска при определении оптимальной величины α_s для шага (II.14) в выбранном направлении r_s .

Некоторые из методов поиска безусловного экстремума рассмотрены в п. II.3, методам условной оптимизации посвящен п. II.4.

II.3. Поиск безусловного экстремума

В *методах одномерного поиска*, в отличие от аналитического метода определения точки экстремума, предполагается, что о целевой функции $f(x)$ неизвестно ничего, кроме того, что она унимодальна — имеет один оптимум в интервале $a \leq x^* \leq b$, который на-

зывается *интервалом неопределенности*. Требования дифференцируемости или непрерывности к функции не предъявляются, но предполагается, что значение $f(x)$ может быть вычислено (получено в эксперименте) для любого x в интервале $a \leq x \leq b$. Суть методов заключается в последовательном уменьшении интервала неопределенности таким образом, чтобы не исключить из него точку x^* . Такой подход не позволяет определить точный оптимум, но дает возможность получить сколь угодно малый интервал, содержащий искомую точку x^* .

Наиболее естественный способ сужения интервала неопределенности для унимодальной функции — деление его на n равных частей (шагов) с последующим вычислением значений $f_s = f(x_s)$ в узлах полученной сетки: $x_0 = a, x_1, \dots, x_s, \dots, x_n = b$. В результате интервал сужается до двух шагов сетки (рис. II.9, а). Затем выполняется следующее дробление суженного интервала и так далее до достижения требуемой точности, заданной, в частности, в виде (II.17) или

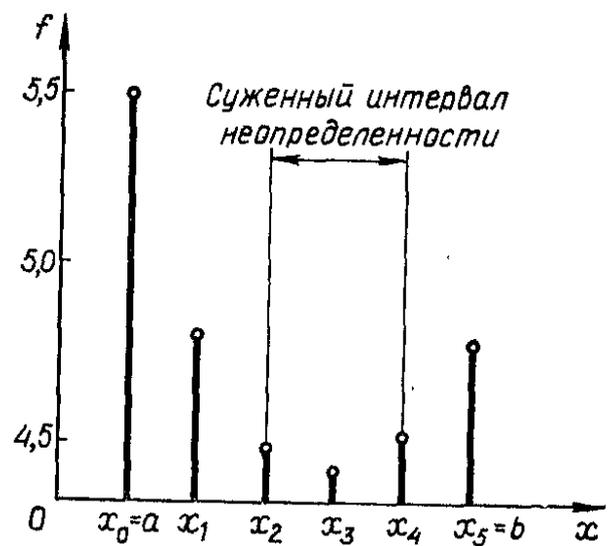
$$x_s - x_{s-1} \leq \varepsilon. \quad (\text{II.18})$$

Этот метод называют *равномерным*, или *общим*, поиском.

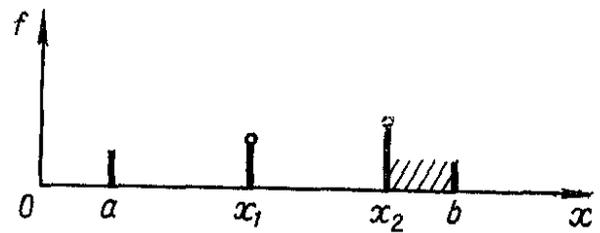
Пример II. 6. В условиях пр. II.4 задано $V = 1 \text{ м}^3$. Требуется, используя равномерный поиск, определить с точностью до 0,05 м радиус цилиндра в пределах $0,4 \leq r \leq 0,9 \text{ м}$, при котором площадь его поверхности (без крышки) минимальна:

$$S(r) = \pi r^2 + 2V/r = \pi r^2 + 2/r \rightarrow \min.$$

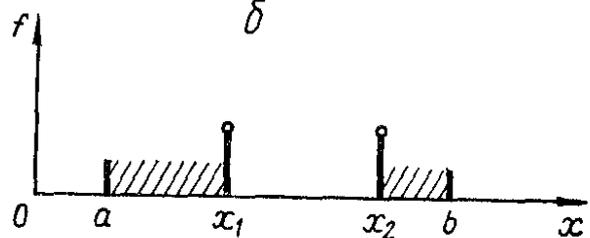
При делении исходного интервала на $n = 5$ равных частей получается равномерная сетка с шагом $\delta = 0,1$. Значения целевой функции $f(x) = S(r)$ в узлах сетки показаны на рис. II.9, а. Очевидно, минимум находится



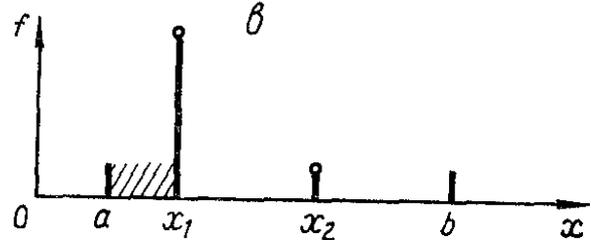
а



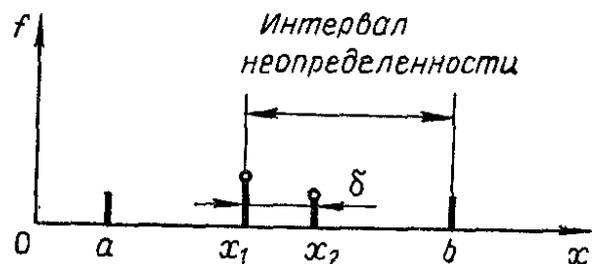
б



в



г



д

Рис. II.9. Сужение интервала неопределенности при равномерном (а) и при последовательном (б, в, г) поиске, а также в методе дихотомии (д)

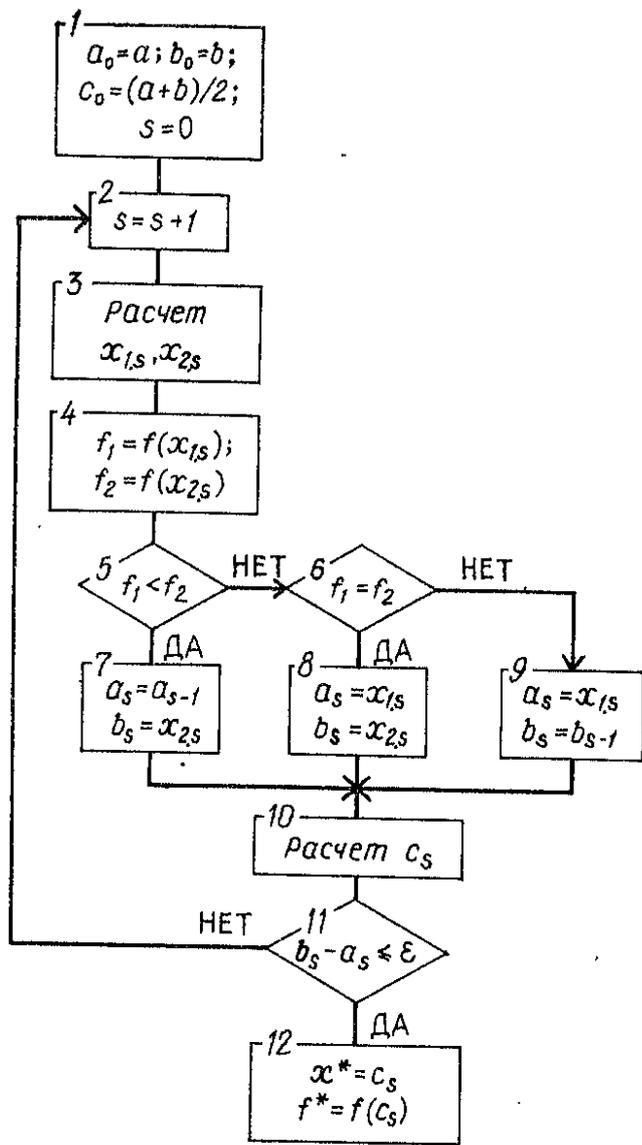


Рис. II.10. Блок-схема алгоритма поиска по методу дихотомии

между узлами x_2 и x_4 , т. е. суженный интервал неопределенности $0,6 < r^* < 0,8$. После его деления на 4 части с шагом 0,05 м и вычисления S при $r = 0,65$ и $r = 0,75$ определяется интервал $0,65 < r^* < 0,75$. В качестве оценки r^* принимается середина интервала $\hat{r}^* = 0,7$. Рассчитанное аналитически из условия стационарности (II.8) значение $r^* = \sqrt[3]{1/\pi} = 0,705$ м.

Разработаны более эффективные, с меньшим числом определений f по сравнению с изложенной схемой, стратегии одномерного поиска (сужения интервала неопределенности). Методы одномерного поиска делят (см. рис. II.8) на методы последовательного поиска (дихотомии, Фибоначчи, золотого сечения) и методы, использующие аппроксимацию функций (в частности, квадратичной интерполяции), с которыми можно познакомиться в книгах [26, 39, 44] и др. Для этих методов разработаны стандартные программы, которые есть в большинстве существующих пакетов программ оптимизации.

Последовательный поиск основывается на том, что для унимодальной функции любая пара опытов (вычислений $f(x)$) позволяет сузить область поиска — интервал неопределенности. Получая и сравнивая значения $f(x)$ в таких точках x_1 и x_2 , где $a < x_1 < x_2 < b$, можно локализовать интервал неопределенности:

- если $f(x_1) < f(x_2)$, то $a \leq x^* < x_2$;
- если $f(x_1) = f(x_2)$, то $x_1 < x^* < x_2$;
- если $f(x_1) > f(x_2)$, то $x_1 < x^* \leq b$.

Эти три типа возможных ситуаций показаны на рис. II.9, б, в, г для условий пр. II.6.

Стратегия выбора x_1 и x_2 с учетом предыдущих соотношений определяет суть того или иного метода последовательного поиска. В частности, рассматривается простейший метод *дихотомии* (половинного деления), в котором на каждом s -м шаге x_{1s} и x_{2s} выбираются (см. рис. II.9, д) на расстоянии $\delta/2$ справа и слева от середины c_{s-1} интервала $[a_{s-1}, b_{s-1}]$:

$$c_{s-1} = (a_{s-1} + b_{s-1})/2; \quad x_{1s} = c_{s-1} - \delta/2; \quad x_{2s} = c_{s-1} + \delta/2.$$

На рис. II.9, δ для условий пр. II.6 показаны: середина исходного интервала неопределенности, выбор точек x_{11} и x_{21} на первом шаге решения задачи методом дихотомии, значения $f(x)$ в этих точках, а также полученный при этом новый, суженный интервал неопределенности, для которого снова следует рассчитать центр c_1 , и т. д. На рис. II.10 показана блок-схема алгоритма поиска по методу дихотомии.

Простейший из методов многомерного поиска — метод покоординатного спуска, или метод Гаусса — Зейделя. Он представляет собой распространение методики одномерного поиска на многомерный случай. В основе его лежит последовательное изменение каждого параметра x_i из точки $x_{0i} = (x_{1.0i}, \dots, x_{i.0i}, \dots, x_{k.0i})$, при сохранении остальных параметров постоянными, до тех пор, пока на шаге s_i не будет достигнут минимум $f(x^*)$ в точке $x_{si} = (x_{1.0i}, \dots, x_{i.si}, \dots, x_{k.0i})$. Начальной точкой x_{0i} для поиска по i -й переменной является точка минимума, достигнутая при поиске по $(i-1)$ -й переменной (за исключением исходной точки x_0). После завершения такой процедуры для всех k

переменных за $(s = \sum_{l=1}^k s_l \text{ шагов})$ начинается новый цикл поиска

с изменения первой переменной из новой начальной точки $x_0 = (x_{1.s1}, \dots, x_{l.sl}, \dots, x_{k.sk})$ с тем, чтобы попытаться еще улучшить решение. Для минимизации по каждой переменной могут использоваться любые методы одномерного поиска. В частности, простейший прием использован для показа алгоритма покоординатного спуска в блок-схеме на рис. II.11.

Параметр L в этом описании алгоритма регистрирует попытки

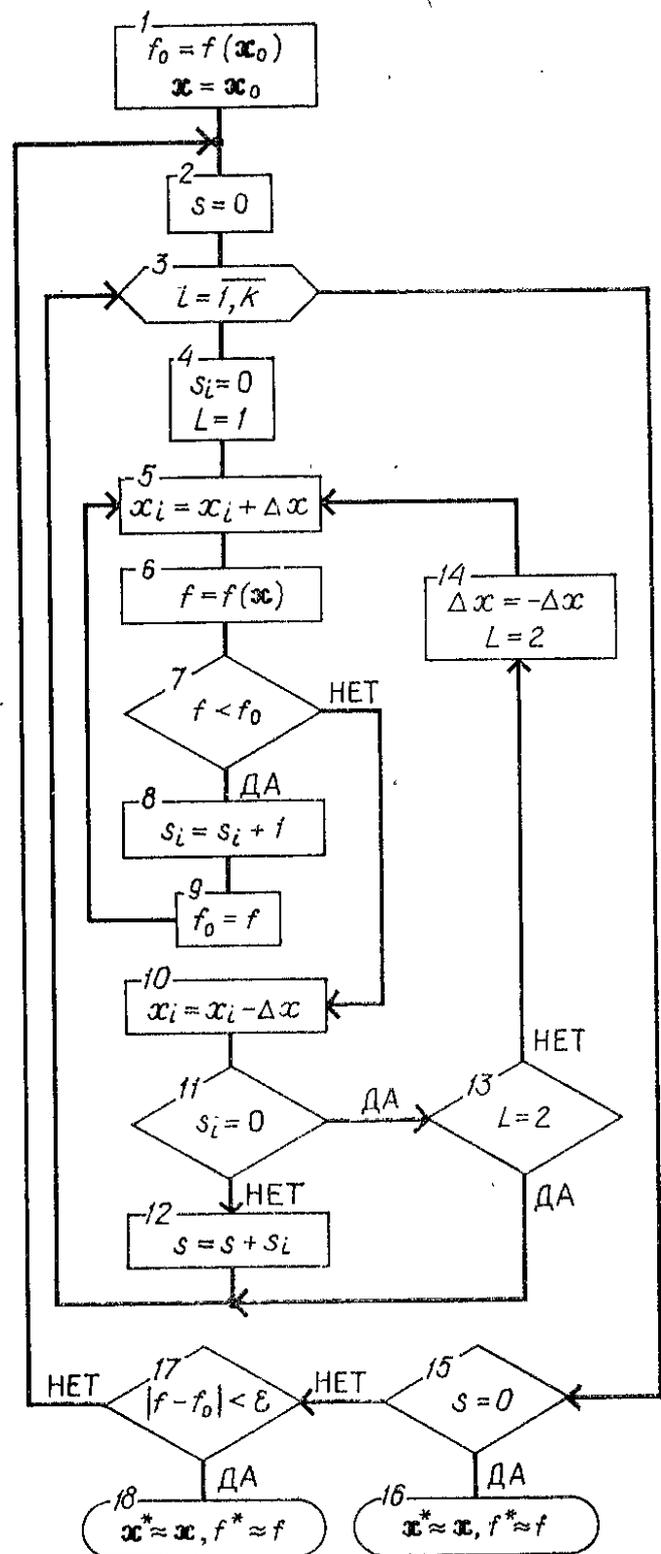


Рис. II.11. Блок-схема алгоритма покоординатного спуска (метод Гаусса-Зейделя)

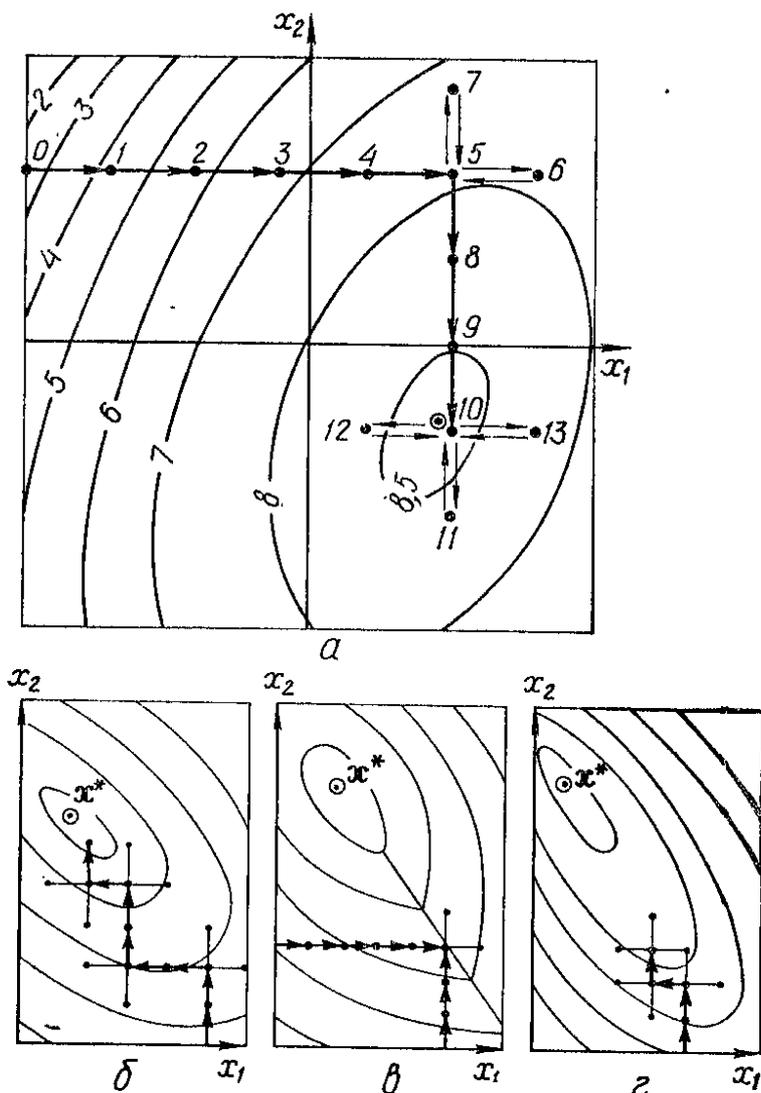


Рис. 11.12. Поиск оптимума по методу Гаусса—Зейделя:

a — поиск максимума прочности композита; *б* — *в* — поиск в особых условиях (*б* — наклон главных осей эллипса к осям координат близок к 45° , *в* — поверхность с линией излома, *г* — «овражная» функция)

двигаться по одной координате в прямом ($L = 1$; блоки 4,5) или обратном ($L = 2$; блоки 14,5) направлениях. Если после s_i успешных шагов (подсчитываемых в блоке 8) в одном из этих направлений очередное приращение Δx_i оказалось неудачным (проверка в блоке 7), то происходит возврат в предыдущую точку (блок 10) и осуществляется переход к движению по следующей $(i + 1)$ -й координате (изменение параметра цикла в блоке 3). Этот же переход выполняется и в том случае, когда движение по i -й переменной не принесло успеха ($s_i = 0$, что проверяется в блоке 11) ни в одном из двух направлений ($L = 2$, что проверяется в блоке 13). Если же продвинуться к оптимуму не удалось последовательно ни по одной из k

переменных (накопленное в блоке 12 число шагов $s = \sum_{i=1}^k s_i = 0$, что проверяется в блоке 15), то поиск прекращается (блок 16), поскольку дальнейшие попытки движения приводят к повтору шагов — «заикливание». Достигнутая при этом лучшая точка рассматривается в качестве приближенной оценки экстремума.

При возможности продолжить движение из этой точки ($s \neq 0$) новый цикл поиска (блок 2 — 15) осуществляется в том случае, если не достигнута заданная точность ε (что проверяется в блоке 17). Если же условия остановки выполняются, то полученный результат f можно считать равным с точностью ε экстремуму f^* (блок 18).

Пример 11.7. Работа алгоритма Гаусса — Зейделя иллюстрируется при решении задачи оптимизации, сформулированной в пр. 11.2, А. Возможность использовать алгоритм безусловной оптимизации в этой задаче с прямыми ограничениями обеспечена (в учебных целях) тем, что поиск максимума прочности по модели (11.4) ведется в пределах допустимой области. В качестве начального приближения взята точка: $x_1 = -1,0$; $x_2 = 0,6$. Шаг по каждой переменной $\Delta x_i = 0,3$; $\varepsilon = 0,1$ МПа. Поскольку в описанном алгоритме

предусмотрен поиск минимума, принимается $f(x) = -R(x)$. Траектория поиска показана на рис. II.12, а. Координаты точек 0 — 13 траектории и значения целевой функции в них приведены также в табл. II.1.

Таблица II.1. Изменение параметров и целевой функции при поиске методом Гаусса—Зейделя

Точки	x_1	x_2	R	Точки	x_1	x_2	R
0	—1,0	0,6	2,44	7	0,5	0,9	7,24
1	—0,7	0,6	4,24	8	0,5	0,3	8,26
2	—0,4	0,6	5,68	9	0,5	0,0	8,50
3	—0,1	0,6	6,76	10	0,5	—0,3	8,56
4	0,2	0,6	7,48	11	0,5	—0,6	8,44
5	0,5	0,6	7,84	12	0,2	—0,3	8,47
6	0,8	0,6	7,84	13	0,8	—0,3	8,29

Поиск прекращается в точке 10 по условию $|f(x_{10}) - f(x_9)| = 8,56 - 8,50 < \varepsilon$. Если задать большую точность, в частности $\varepsilon = 0,05$, то улучшить результат все равно не удастся, поскольку после точки 13 начнется повтор точек и алгоритм прекратит работу по условию защиты от заикливания. Очевидно, что при лучшем начальном приближении $x_0 = (0,5; 0,6)$ для получения того же результата необходимо только 6 определений целевой функции.

В этом примере может быть определена и истинная, достигнутая численным методом, точность по выражению (II.16), поскольку из условия (II.8) легко определить аналитическое решение задачи:

$$R^* = R(x_1^*, x_2^*) = R(0,43; -0,29) = 8,57 \text{ МПа};$$

$$\varepsilon_{10} = 8,56 - 8,57 = 0,01 \text{ МПа}.$$

Эффективность и возможность применения метода покоординатного спуска определяются характером целевой функции. Он прекрасно работает в тех случаях, когда ее изоповерхности по форме близки k -мерной сфере или эллипсу, оси которого параллельны осям координат. Близкий к такому идеальному случаю пример показан на рис. II.12, а. При большем наклоне осей, когда целевая функция содержит существенные взаимодействия переменных, алгоритм становится неэффективным (для нахождения оптимума приходится определять гораздо больше значений целевой функции, чаще меняя направление движения — рис. II.12, б). Метод не всегда позволяет найти оптимальное решение, если линии имеют точки излома (рис. II.12, в) или если поверхность функции носит «овражный» характер (рис. II.12, г). Создан ряд модификаций метода [39, 44, 64], избавляющих идею покоординатного спуска от указанных ограничений, в частности с поворотом системы координат по мере продвижения к оптимуму. Однако за подобные улучшения приходится расплачиваться усложнением алгоритма и добавочными вычислениями $f(x)$ для сбора информации при выборе направлений поиска.

Отказ от одномерных представлений при оптимизации в многомерном пространстве позволяет увеличить число возможных направ-

лений для движения к оптимуму. Так, вместо перемещения точки вдоль k координат для обследования k -мерного пространства в нем может перемещаться, ориентируясь в зависимости от значений целевой функции, множество точек некоторой k -мерной конфигурации. Простейшей из них является симплекс — правильный многогранник со стороной l и с $k + 1$ вершинами (см. также п. III. 13). Координаты вершин равностороннего треугольника в пространстве двух переменных, тетраэдра в трехмерном пространстве и т. д. в общем случае определяются с помощью матрицы Q , в которой столбцы соответствуют $k + 1$ вершинам, а строки — k координатам:

$$Q = \begin{pmatrix} 0 & q_1 & q_2 & \dots & q_2 \\ 0 & q_2 & q_1 & \dots & q_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & q_2 & q_2 & \dots & q_1 \end{pmatrix}, \quad (\text{II.19})$$

где

$$q_1 = \frac{l}{k\sqrt{2}}(\sqrt{k+1} + k - 1); \quad q_2 = \frac{l}{k\sqrt{2}}(\sqrt{k+1} - 1).$$

Симплексный метод поиска состоит в следующем.

Устанавливаются координаты вершин исходного (начального) симплекса на базе точки x_0 (точки начального приближения к оптимуму). Пусть x_0 соответствует вершине с номером 1, тогда остальные k вершин определяются при сложении соответствующих столбцов матрицы Q с вектором x_0 .

Вычисляются значения целевой функции в каждой из вершин и определяется худшая из них — с наибольшим значением f . Пусть для треугольника на рис. II.13,а это вершина 1. На основе такой информации направление наискорейшего улучшения целевой функции определяется от худшей точки (1) по прямой, проходящей через центр симплекса c_S или центр противоположной грани c — показано стрелкой на рис. II.13, а. На этой прямой определяется вместо худшей, исключаемой, вершины новая, *отраженная вершина* 4, координаты которой $x_{i4} = c_i + (c_i - x_{i1}) = 2c_i - x_{i1} = 2(x_{i2} + x_{i3})/2 - x_{i1} = x_{i2} + x_{i3} - x_{i1}$. Отраженная вершина вместе с оставшимися вершинами образует новый, *отраженный симплекс*.

Продолжение этой процедуры, в которой каждый раз вычеркивается вершина, где целевая функция максимальна; обеспечивает продвижение центра симплекса к минимуму шагами постоянной величины, но в изменяемом направлении.

Пример II.8. На рис. II.13, б показана работа симплексного метода в той же задаче оптимизации, которая использована для демонстрации алгоритма покоординатного спуска в пр. II.7. Задана та же точность поиска, и та же начальная точка принимается в качестве базовой вершины исходного симплекса со стороной $l = 0,3$. Координаты двух других вершин определяются с помощью (II.19):

$$\begin{aligned} x_{11} &= x_{10} + q_1; & x_{21} &= x_{20} + q_2; \\ x_{12} &= x_{10} + q_2; & x_{22} &= x_{20} + q_1, \end{aligned}$$

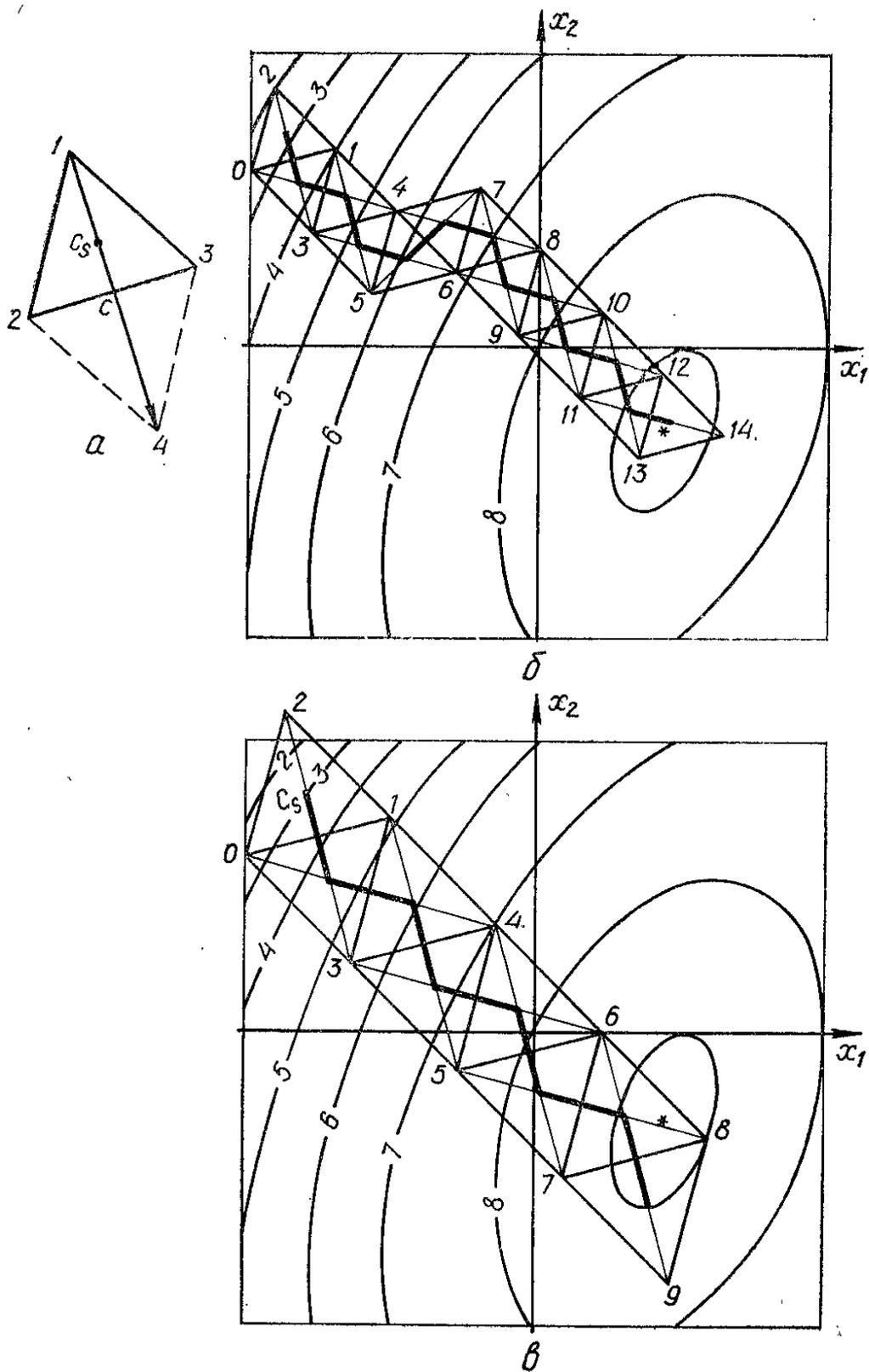


Рис. 11.13. Симплексный метод поиска:

a — построение отраженного симплекса; *б* — траектория поиска максимума прочности композита; *в* — то же, при увеличении стороны симплекса

где

$$q_1 = 0,3(\sqrt{2+1} + 2 - 1)/2\sqrt{2} = 0,290;$$

$$q_2 = 0,3(\sqrt{2+1} - 1)/2\sqrt{2} = 0,078.$$

В табл. 11.2 приведены координаты этих и следующих точек, а также полученные для них по модели (11.4) значения максимизируемой функции $R(x)$.

Остановка поиска по условию (11.17) происходит в точке 13: $|f(x_{13}) - f(x_{12})| < 0,1$. Лучшее достигнутое значение R , полученное в этой точке, практически совпадает с рассчитанным аналитически R^* .

Таблица 11.2. Изменение параметров и целевой функции при поиске симплексным методом

Точки	x_1	x_2	R	Точки	x_1	x_2	R
0	-1,00	0,60	2,44	8	0,00	0,33	7,56
1	-0,71	0,68	3,95	9	-0,07	0,04	7,79
2	-0,92	0,89	1,95	10	0,22	0,12	8,09
3	-0,79	0,39	4,34	11	0,14	-0,17	8,35
4	-0,50	0,47	6,06	12	0,43	-0,09	8,53
5	-0,58	0,18	5,88	13	0,35	-0,38	8,56
6	-0,29	0,25	6,87	14	0,64	-0,30	8,48
7	-0,21	0,55	6,54				

Если для (II.17) задать $\varepsilon = 0,01$ (по содержанию решаемой задачи такая точность не имеет смысла), то после точки 14 вершины начнут повторяться; требуемая точность не может быть достигнута с помощью симплекса выбранного размера. Поэтому в программах, реализующих симплексный метод, должна быть предусмотрена остановка в случае повтора точек. Для предотвращения циклического движения на подступах к экстремуму в алгоритм могут быть также введены правила уменьшения размера симплекса.

В данном примере при $l = 0,3$ шаг смещения центра симплекса меньше, чем Δx в пр. II.7, и число n_e вычислений $f(x)$ оказалось большим.

Поиск с шагом $\|\Delta x\| = 0,3$ при $l = \|\Delta x\| \sqrt{3} = 0,52$ показан для сравнения на рис. II.13, в, где видно, что симплекс такого размера быстрее приходит в зону $R > 8,3$. Однако остановка по условию $|f(x_7) - f(x_8)| < 0,1$ не обеспечивает желаемую точность. Если же условие остановки — начало повтора вершин, то после испытания точки 9 лучшее решение будет зафиксировано в точке 8, где $|R(x_8) - R^*| < 0,1$.

Простота симплексного поиска, с одной стороны, а с другой — такие его недостатки, как заикливание, низкая эффективность при минимизации «овражных» функций (см. рис. II.12, г), привели к созданию ряда его модификаций, в частности *метода деформируемого многогранника* (Нелдера и Мида). В этом методе [39, 64] исходный симплекс на первом же шаге может изменить свою конфигурацию, стать нерегулярным. Пусть A — худшая, заменяемая вершина многогранника, B — лучшая вершина с наименьшим значением $f(x)$, A' — отраженная (рис. II.14, а). Деформация многогранника происходит в результате выполнения одной из трех операций в зависимости от значения целевой функции в отраженной вершине.

1. Если $f(A') \leq f(B)$, выполняется *растяжение* многогранника в этом, оказавшемся удачным, направлении, и новая вершина определяется в точке D на расстоянии, в γ раз большем от центра грани s , чем расстояние от s до A , если при этом $f(D) < f(A')$.

2. Если значение функции в точке A' оказалось больше ее значений во всех вершинах многогранника, кроме A , то выполняется его *сжатие* и новая вершина D располагается вдвое ближе к s , чем A .

3. Если же $f(A') > f(A)$, выполняется *редукция* — уменьшение размеров многогранника в два раза по отношению к вершине B .

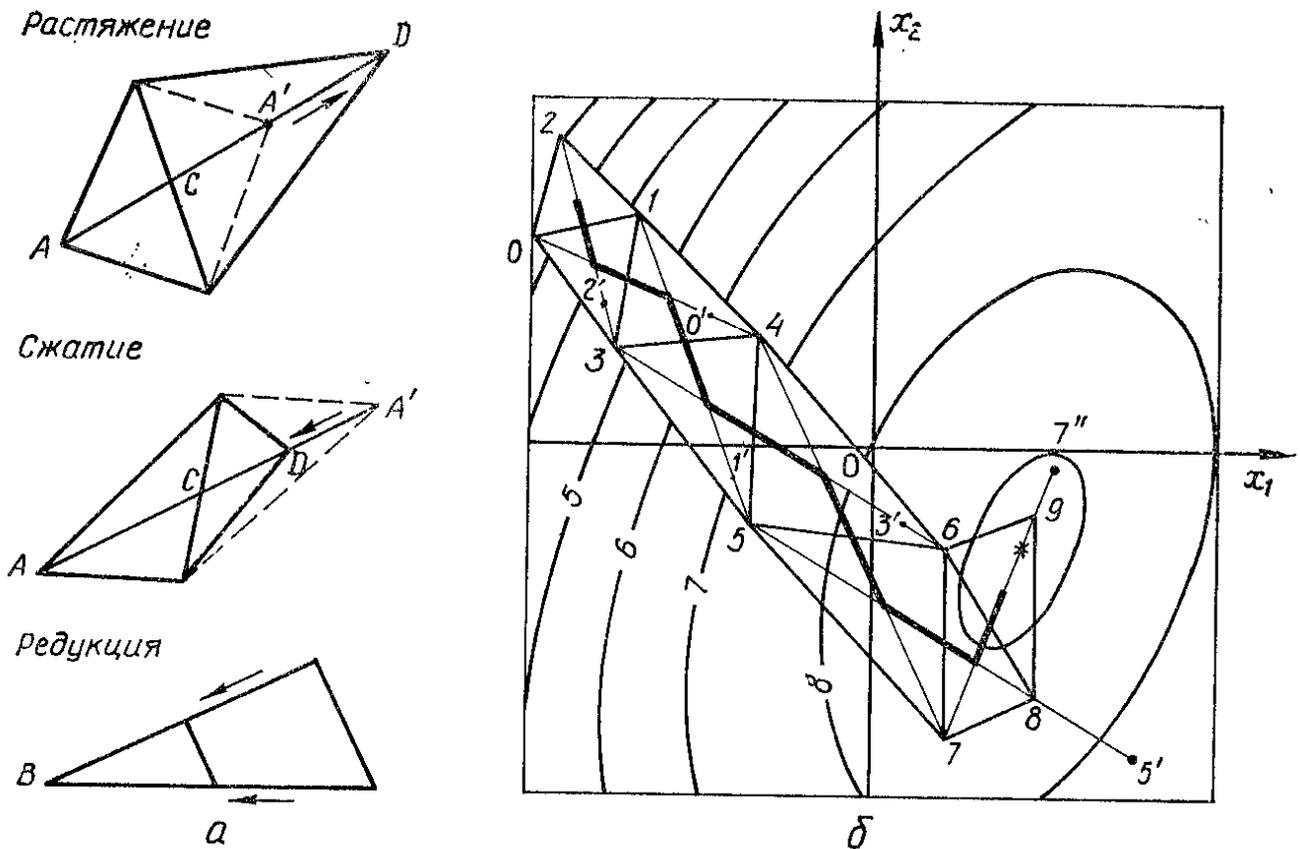


Рис. 11.14. Поиск оптимума методом деформируемого многогранника:
a — типы деформаций; *b* — траектория поиска максимума прочности композита

В остальных случаях новой вершиной D становится точка A' и сохраняются размеры предыдущего многогранника.

Таким образом, в этом методе, по сравнению с симплексным поиском, движение к оптимуму становится менее жестким, получает то или иное ускорение в зависимости от текущей информации об уровне целевой функции; в процессе поиска изменяются не только направление, но и величина шага.

Пример 11.9. Поиск деформируемым многогранником максимума функции (11.4) показан на рис. 11.14, б. Значение коэффициента растяжения $\gamma = 1,4$ подобрано так, чтобы при показе безусловного поиска не парушались условия $x_1 \geq -1$, $x_2 \geq -1$ (пр. 11.2, А), поскольку при $x_i < -1$ оптимизируемая функция $R(x)$ не имеет содержательного смысла; для практических

Таблица 11.3. Изменение параметров и целевой функции при поиске методом деформируемого многогранника

Точки	x_1	x_2	R	Точки	x_1	x_2	R
0	-1,00	0,60	2,44	5	-0,36	-0,22	7,27
1	-0,71	0,68	3,95	3'	0,04	-0,19	8,24
2	-0,92	0,89	1,95	6	0,20	-0,28	8,47
2'	-0,79	0,39	4,34	7 (4')	0,21	-0,82	8,30
3	-0,76	0,29	4,73	5'	0,76	-0,89	7,79
0'	-0,47	0,37	5,94	8	0,48	-0,72	8,35
4	-0,36	0,32	6,48	9 (7')	0,47	-0,18	8,56
1'	-0,42	-0,07	6,90	7''	0,53	-0,05	8,52

задач рекомендуется $2 \leq \gamma \leq 3$ [64]. Условия начала и окончания поиска взяты те же, что в пр. II. 7, II.8. Координаты точек и уровни функции $R(x)$ при движении к максимуму приведены в табл. II.3. Оптимальное решение получается в точке 9: $\hat{x}^* = x_9 = (0,47; -0,18)$; $\hat{R}^* = R(x_9) = 8,56$ МПа.

В методах случайного поиска [64] основная идея одного из детерминированных методов реализуется с использованием тех или иных случайных элементов. Так, «слепой» поиск с выбором для определения значений целевой функции случайных точек из области решений родственен детерминированному сканированию, последовательность случайных направлений для поочередной минимизации вдоль направления аналогична покоординатному спуску, случайный комплекс вершин многогранника используется как аналог симплексного метода и т. п. Развитие таких методов в СССР связано с работами проф. Л. А. Растригина. Случайный поиск применяют при решении сложных задач большой размерности, где он оказывается эффективнее детерминированных процедур.

Класс поисковых методов первого порядка (см. классификацию в п. II.2), использующих свойства градиента целевой функции при определении шага поиска (II.14), называют *градиентными методами*. Градиент $\nabla f(x)$ (II.9) в точке x ортогонален поверхности уровня $f(x)$, проходящей через эту точку, и направлен в сторону наискорейшего увеличения функции — *наискорейшего подъема*.

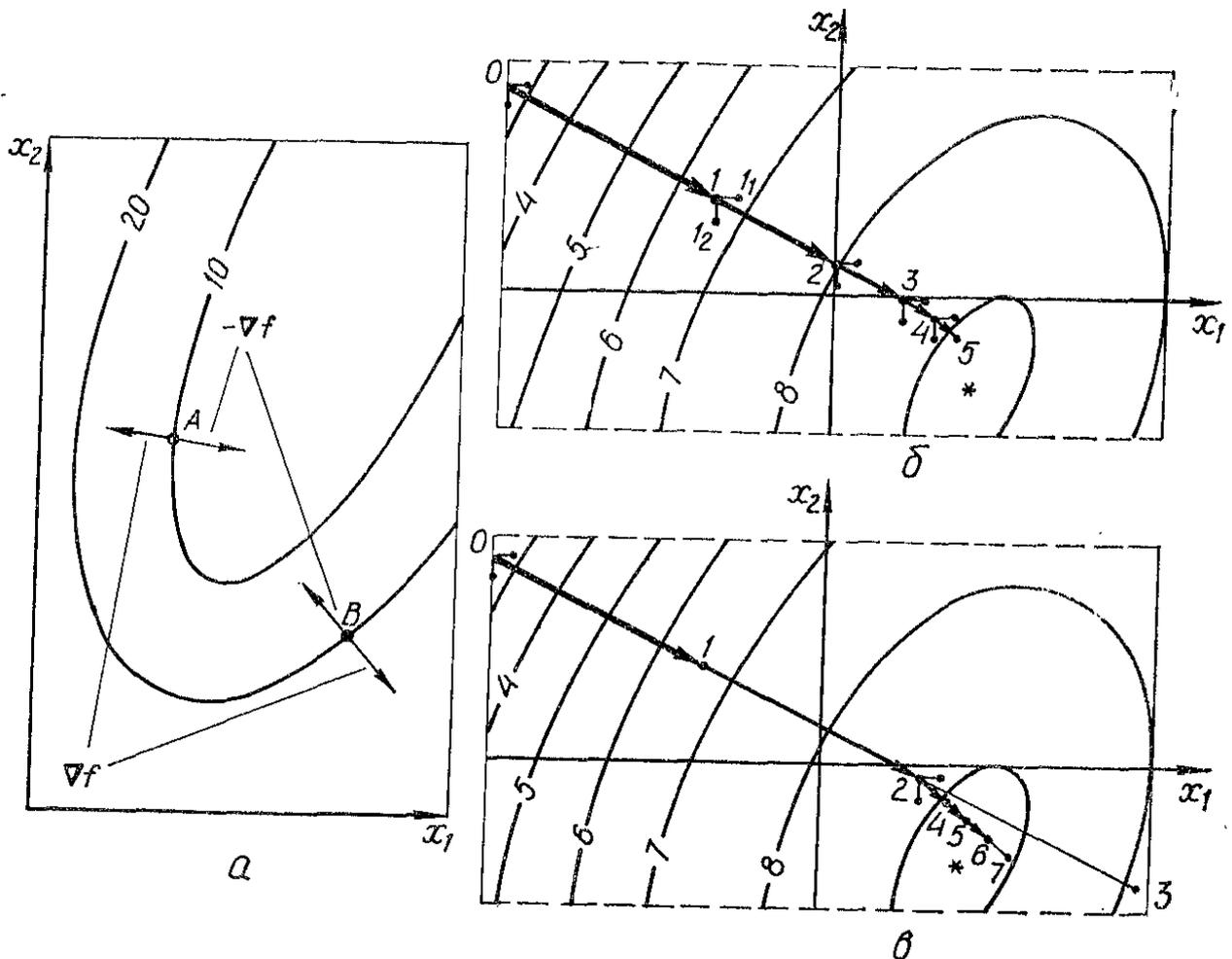


Рис. II.15. Поиск оптимума градиентными методами:

a — изменение направления градиента при переходе от точки *A* к точке *B*; *б* — траектория поиска R_{\max} методом градиента; *в* — то же, методом крутого восхождения

ема. Вектор $-\nabla f$, противоположный градиенту (*антиградиент*, или отрицательный градиент) направлен в сторону наискорейшего убывания функции — *наискорейшего спуска*. Оба эти направления показаны на рис. II.15, а для двух точек на изолиниях $f(\mathbf{x})$. Градиент в точке экстремума равен нулю (II.10).

Собственно *метод градиента* состоит в многократном повторении двух операций: операции определения направления градиента целевой функции в достигнутой ранее точке \mathbf{x}_s и операции перемещения из этой точки в направлении, противоположном найденному, на вычисляемый шаг

$$\Delta \mathbf{x}_s = -a \nabla f(\mathbf{x}), \quad (\text{II.20})$$

где a — задаваемый параметр шага поиска (если $a = \text{const}$, то величина шага пропорциональна модулю градиента).

Для определения направления градиента вычисляются различными способами оценки частных производных целевой функции по оптимизируемым параметрам, в частности

$$\frac{\partial f(\mathbf{x})}{\partial x_i} = \frac{\Delta f(\mathbf{x})}{\Delta x_i} = \frac{1}{g_i} \{f[x_1, \dots, (x_i + g_i), \dots, x_k] - f[\mathbf{x}]\}, \quad (\text{II.21})$$

где g_i — величина так называемого *пробного* шага вдоль i -й координаты.

Окончание поиска происходит при уменьшении величины *рабочего* шага (II.20) или модуля градиента до задаваемых значений.

Пример II.10. Для пояснения метода градиента используются, так же как в пр. II.7 — II.9, условия оптимизационной задачи из пр. II.2, А. Поиск максимума $R(\mathbf{x})$ на основе оценок градиента вдоль траектории движения к цели показан на рис. II.15, б. Для получения оценок составляющих вектора градиента по (II.21) размеры пробных шагов равны $g_1 = 0,05$ и $g_2 = -0,05$; параметр шага поиска $a = 0,1$. Координаты «пробных» и «рабочих» точек с соответствующими значениями R , а также покоординатные размеры рабочих шагов $\Delta x_1, \Delta x_2$ приведены в табл. II. 4. Поиск прекращается в точке 5 по условию $\Delta x_i < 0,05$ ($i = \overline{1, 2}$).

Таблица II.4. Изменение параметров и целевой функции при поиске методом градиента

Точки	x_1	Δx_1	x_2	Δx_2	R	Точки	x_1	Δx_1	x_2	Δx_2	R
0	-1,00		0,60		2,44	3	0,21		-0,02		8,35
0 ₁	-0,95	0,65	0,60	-0,32	2,76	3 ₁	0,26	0,10	-0,02	-0,07	8,40
0 ₂	-1,00		0,55		2,60	3 ₂	0,21		-0,07		8,38
1	-0,35		0,28		6,60	4	0,31		-0,09		8,48
1 ₁	-0,30	0,36	0,28	-0,19	6,78	4 ₁	0,36	0,06	-0,09	-0,05	8,51
1 ₂	-0,35		0,23		6,69	4 ₂	0,31		-0,14		8,50
2	0,01		0,09		7,92	5	0,37		-0,14		8,53
2 ₁	0,06	0,20	0,09	-0,11	8,02						
2 ₂	0,01		0,04		7,98						

Высокая точность и стремление к движению на каждом рабочем шаге в наилучшем направлении обеспечили широкое применение градиентного метода и многих его модификаций. Эти модификации, как правило, направлены против основного недостатка метода градиента — необходимости выполнить как минимум k пробных шагов перед каждым рабочим шагом.

Так *метод наискорейшего спуска* предполагает выполнение следующих операций:

1) определение наилучшего направления движения (оценка направления градиента) в некоторой достигнутой точке;

2) движение в антиградиентном направлении с некоторым шагом до тех пор, пока не будет зарегистрировано увеличение минимизируемой функции, после чего снова выполняется операция 1, и т. д.

Аналогично в *методе крутого восхождения* подобная стратегия предусматривает поиск максимума.

Метод обладает повышенной скоростью сходимости по сравнению с методом градиента, поскольку $k + 1$ пробных вычислений (измерений) проводятся здесь не на каждом шаге и часто заменяются однократным замером приращения функции за рабочий шаг. Поэтому метод наискорейшего спуска эффективен при оптимизации многомерных объектов, особенно в тех случаях, когда направление градиента вдоль траектории движения к цели изменяется незначительно (см. рис. II.15, б).

Пример II.11. На рис. II.15, в и в табл. II.5 показан поиск максимума функции $R(x)$ (II.4) методом крутого восхождения (или поиск минимума $f(x) = -R(x)$ методом наискорейшего спуска). Значения g_1, g_2 и a те же, что и в пр. II.10.

Таблица II.5. Изменение параметров и целевой функции при поиске методом крутого восхождения

Точки	x_1	x_2	R	Точки	x_1	x_2	R
0	-1,00	0,60	2,44	2 ₁	0,35	-0,04	8,48
0 ₁	-0,45	0,60	2,76	2 ₂	0,30	-0,09	8,47
0 ₂	-1,00	0,55	2,60	4	0,37	-0,10	8,52
1	-0,35	0,28	6,60	5	0,44	-0,16	8,55
2	0,30	-0,04	8,45	6	0,51	-0,22	8,56
3	0,45	-0,36	7,98	7	0,58	-0,28	8,52

Следует обратить внимание, что в точках 2 и 6 регистрируются максимальные значения вдоль соответствующих направлений восхождения к оптимуму; это обнаруживается соответственно в точках 3 и 7, из которых осуществляется возврат в предыдущую точку. Поиск прекращается в точке 6, где следующие вычисленные размеры рабочего шага оказываются меньше 0,05. Для получения оптимального решения потребовалось только дважды оценить направление движения с помощью пробных шагов — ср. в пр. II.10.

В рассмотренных численных методах безусловной оптимизации не имеет значения, каким образом при поиске добывается информа-

ция об уровне целевой функции: в физическом эксперименте (на объекте или его физической модели) или в результате вычислений (по точной формуле, по экспериментально-статистической модели, при выполнении программы на ЭВМ). Для оптимизации объектов в реальном времени на основе прямых или косвенных измерений критерия оптимальности используются, в частности, симплексный поиск [1, 40] и процедура крутого восхождения Бокса — Уилсона [9, 48 и др.] (см. п. III. 15). При этом важно, чтобы было приемлемым время, необходимое для выполнения каждого шага поиска и определения значения f .

II.4. Методы условной оптимизации

В большинстве практических задач оптимизации требуется определить условный экстремум критерия оптимальности. Для этой цели богатый арсенал средств безусловной оптимизации может быть непосредственно применен лишь в тех исключительных случаях, когда заранее известно, что оптимум x^* находится внутри области решений, и путь из точки начального приближения x_0 к точке $x_s \approx x^*$ проходит также внутри этой области, не пересекая ее границ.

Поэтому существуют следующие два подхода к условной оптимизации. Первый предполагает использование методов поиска безусловного экстремума после того, как поставленная задача нелинейного программирования (II.11) сведена с помощью тех или иных преобразований и «расширений» к задаче безусловной оптимизации (в пространстве R^k без ограничений). При втором подходе используются методы, специально предназначенные для оптимизации с ограничениями.

Свести задачу условной оптимизации к безусловной означает устранить прямые и функциональные ограничения, но таким образом, чтобы определяемый безусловный экстремум однозначно соответствовал искомому условному экстремуму.

Прямые ограничения (II.5) можно устранить соответствующим нормированием оптимизируемых параметров. Например [39], если $a < x_i < b$, то формула (II.22) преобразует ограниченный параметр x_i в неограниченный z_i :

$$z_i = \operatorname{tg} \{ \pi [x_i - (b + a)/2] / (b - a) \}. \quad (\text{II.22})$$

Функциональные ограничения (II.6), (II.7) в общем случае устраняются из постановки оптимизационной задачи за счет введения новой, расширенной целевой функции, которая конструируется с учетом типа ограничений.

Каждое ограничение-равенство (II.6) в задаче нелинейного программирования уменьшает на единицу число независимых переменных, т. е. число степеней свободы оптимизируемой системы (см. п. III.3). Так, в пр. II.4 заданный объем цилиндра обуславливает зависимость его высоты от радиуса. Очевидно, что число p зависимостей h_j в (II.11) должно быть меньше общего числа k оптимизируемых

параметров. Тогда число независимых переменных $k - p$ не равно нулю. Если же $p = k$, то оптимизировать нечего — значения всех переменных можно определить из совместного решения системы уравнений $h_j = 0$ ($j = \overline{1, p}$).

Если в задаче (II.11) все $p < k$ зависимых переменных удастся на основе (II.6) выразить через $k - p$ остальных, независимых переменных, то, таким образом, ограничения-равенства из условий задачи исключаются: при этом уменьшается размерность пространства поиска (см. пр. II.4).

В задачах (II.11) с ограничениями типа равенств (когда $q = 0$, а равенства исключить не удастся, $p \neq 0$) в качестве расширенной целевой функции используется *функция Лагранжа* — линейная комбинация целевой функции и функций h_j :

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \lambda_1 h_1(\mathbf{x}) + \lambda_2 h_2(\mathbf{x}) + \dots + \lambda_p h_p(\mathbf{x}), \quad (\text{II.23})$$

где λ_j — *множители Лагранжа* [36, 44].

Функция (II.23) является функцией $k + p$ переменных. Ее стационарные точки (см. (II.8) в п. II.1) получают, приравнявая нулю частные производные по λ_j ($j = \overline{1, p}$) и по x_i ($i = \overline{1, k}$):

$$\frac{\partial L(\mathbf{x}, \boldsymbol{\lambda})}{\partial \lambda_j} = h_j(\mathbf{x}) = 0 \quad (j = \overline{1, p}); \quad (\text{II.24})$$

$$\frac{\partial L(\mathbf{x}, \boldsymbol{\lambda})}{\partial x_i} = 0 \quad (i = \overline{1, k}). \quad (\text{II.25})$$

Уравнения (II.24) совпадают с ограничениями (II.6) и, как следует из (II.23), при их выполнении $L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x})$. Поэтому, если в стационарной точке $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ функция $L(\mathbf{x}, \boldsymbol{\lambda})$ достигает минимума, то \mathbf{x}^* обеспечивает и минимум функции $f(\mathbf{x})$.

Таким образом, поиск безусловного оптимума функции Лагранжа оказывается эквивалентным поиску условного оптимума целевой функции.

Пример II.12. Решая задачу, сформулированную в пр. II.4, с помощью изложенного метода множителей Лагранжа получим:

$$L(r, H, \lambda) = \pi(r^2 + 2rH) + \lambda(V - \pi r^2 H).$$

Дифференцируя по r , H и λ , находим систему уравнений:

$$\frac{\partial L}{\partial r} = 2\pi(r + H - \lambda r H) = 0;$$

$$\frac{\partial L}{\partial H} = 2\pi r - \pi r^2 \lambda = 0;$$

$$\frac{\partial L}{\partial \lambda} = V - \pi r^2 H = 0.$$

Решением системы являются $r^* = H^* = \sqrt[3]{V/\pi}$, $\lambda^* = 2\sqrt[3]{\pi/V}$; полученные значения r^* и H^* обеспечивают минимальную площадь поверхности цилиндра заданного объема V .

Дополнение $\sum_{j=1}^p \lambda_j h_j(\mathbf{x})$ к исходной, основной целевой функции $f(\mathbf{x})$ в (II.23) можно рассматривать как *штраф*, накладываемый на

те значения \mathbf{x} , которые не удовлетворяют условиям задачи — системе ограничений-равенств $h_j(\mathbf{x}) = 0$ ($j = \overline{1, p}$). Введение штрафа приводит к тому, что все точки вне допустимой области Ω оказываются по уровню расширенного вспомогательного критерия оптимальности хуже любой точки, принадлежащей Ω . Этот принцип используется при записи расширенной целевой функции и в тех случаях, когда к эквивалентной задаче безусловной оптимизации требуется свести исходную задачу нелинейного программирования (II.11) с ограничениями типа неравенств (когда $p = 0$, $q \neq 0$). Соответствующую группу методов называют *методами штрафных функций* [5, 39, 44]. Основная идея этих методов состоит в замене критерия оптимальности $f(\mathbf{x})$ исходной задачи вспомогательным критерием $F(\mathbf{x})$, значения которого совпадают со значениями исходного основного критерия внутри допустимой области; при выходе из нее вспомогательный критерий резко возрастает за счет зависящих от $g_j(\mathbf{x})$ ($j = \overline{1, q}$) таких функций $G[g(\mathbf{x})]$, для которых $G(\mathbf{x}) \rightarrow 0$ внутри и на границе допустимой области, но $G(\mathbf{x}) \gg 0$, если не выполнено хотя бы одно из ограничений $g_j(\mathbf{x}) \leq 0$. Задача (II.11) сводится при этом к последовательности задач безусловной минимизации штрафных функций

$$F_n(\mathbf{x}) = f(\mathbf{x}) + r_n G(\mathbf{x}) \rightarrow \min, \quad \mathbf{x} \in R^k, \quad (\text{II.26})$$

где $r_n \geq 0$ — параметр штрафа.

Пределом решений задач (II.26) при $n \rightarrow \infty$ будет искомая точка условного минимума.

Примером функции $G(\mathbf{x})$ является

$$G_1(\mathbf{x}) = \sum_{j=1}^q \left[\frac{1}{2} (g_j(\mathbf{x}) + |g_j(\mathbf{x})|) \right]^2. \quad (\text{II.27})$$

Подобные выражения обеспечивают при поиске $\mathbf{x}^* = \arg \min F_n(\mathbf{x})$ быстрое возвращение точки \mathbf{x}_s в допустимую область. В ряде методов точка \mathbf{x}_s удерживается внутри области Ω (не может пройти сквозь воздвигаемую на ее границе «стену», преодолеть «барьер»). Это достигается с помощью так называемых *барьерных функций*, в частности

$$G_2(\mathbf{x}) = \sum_{j=1}^q [1/g_j(\mathbf{x})]. \quad (\text{II.28})$$

Успешное применение методов штрафных функций требует глубоких знаний математического анализа и теории численных методов для формулировки и решения задач (II.26). Поэтому инженеру-технологу предпочтительнее решить непосредственно исходную задачу (II.11) условной оптимизации с заданными ограничениями, используя так называемые *методы спуска*, в которых контролируется соблюдение ограничений и обеспечивается движение к оптимуму по последовательности допустимых точек \mathbf{x}_s .

В методах спуска [5, 26, 44, 64] основа того или иного метода поиска безусловного экстремума дополнена правилами действий в случае нарушения ограничений типа неравенств. В результате этих действий точка *перемещается* в пределы допустимой области на то или иное расстояние от границы (в частности, так, что траектория поиска будет проходить вдоль границы и вблизи от нее).

Одним из методов спуска является *метод Бокса* [8, 64], который использует идею симплексного метода безусловной оптимизации, но разработан специально для условной оптимизации при большом числе ограничений типа неравенств. В этом методе в качестве набора исходных точек для начала движения к цели используются не $t = k + 1$ вершин правильного выпуклого многогранника в k -мерном пространстве, а *комплекс* из $t > k + 1$ точек (вершин комплекса), случайно выбираемых из области, заданной ограничениями.

Как и в методе деформируемого многогранника, в процессе движения к оптимуму комплекс деформируется: при замене худшей вершины на отраженную он растягивается в $\gamma = 1,3$ раз. Если новая точка оказалась хуже исключенной, она подтягивается к лучшей вершине комплекса на половину расстояния между ними; если же новая точка нарушает функциональные ограничения (II.7), она подтягивается вдвое ближе к центру тяжести остальных вершин комплекса; при нарушении новой точкой i -го факторного ограничения соответствующей координате присваивается значение, близкое к граничному ($x_i = x_{hi} + \varepsilon$ или $x_i = x_{vi} - \varepsilon$).

Таким образом, комплекс точек (отсюда другое название метода Бокса — *комплекс-метод*) при спуске к оптимуму изменяет размеры и конфигурацию, как бы адаптируясь к характеру поверхности целевой функции и границам допустимой области.

Пример II.13. К условиям пр. II.2, А добавлено требование к величине относительных деформаций материала: $\varepsilon_{тр} \geq 5$ мм/м. Зависимость этого показателя от содержания в композите минерального и органического наполнителей описывается моделью:

$$\ln \hat{\varepsilon} = Y(x) = 1,97 - 0,59x_1 - 0,23x_1^2 + 0,52x_2 - 0,23x_2^2. \quad (\text{II.29})$$

Таким образом, с учетом $R(x)$ (II.4) формулируется задача

$$R(x) \rightarrow \max; \quad |x_1| \leq 1; \quad |x_2| \leq 1; \quad 5 - e^{Y(x)} \leq 0. \quad (\text{II.30})$$

На рис. II.4 задача оптимизации (II.29) — (II.30) представлена геометрически как пример задачи нелинейного программирования с нелинейными целевой и ограничивающей функциями; здесь она используется для иллюстрации работы комплекс-метода.

В качестве одной из вершин исходного комплекса взята общая для пр. II. 7 — II.11 начальная точка. Перемещение деформируемого комплекса в допустимой области в направлении оптимума показано на рис. II.16. Координаты вершин и промежуточных точек с вычисленными по (II.4) значениями целевой функции приведены в табл. II.6.

Вершина 5, в частности, является отражением 2' с коэффициентом $\gamma = 1,3$ худшей вершины 2 исходного комплекса относительно центра тяжести c_{124} остальных его вершин. Она получена следующим образом:

$$\begin{aligned} c_{134} &= (x_1 + x_2 + x_4)/3; \\ x_5 &= c_{134} + 1,3(c_{134} - x_2) = 2,3c_{134} - 1,3x_2. \end{aligned}$$

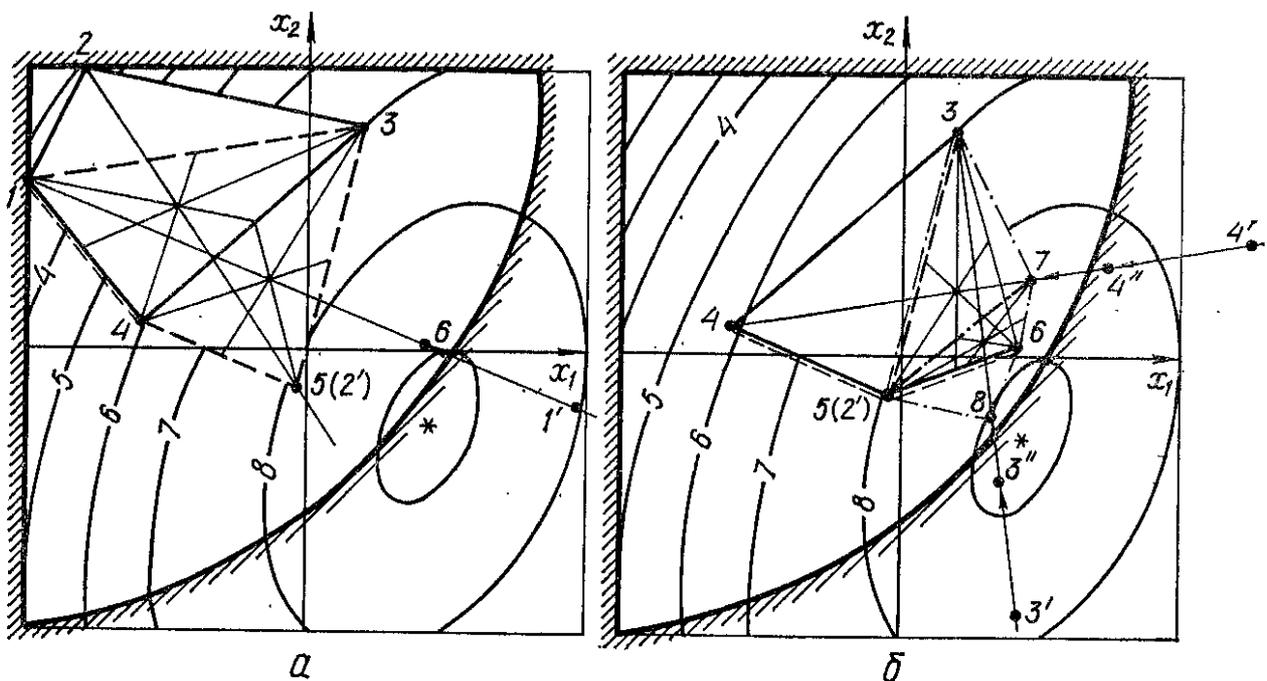


Рис. 11.16. Поиск условного оптимума комплекс-методом (Бокса) на первой (а) и заключительной (б) стадиях

В каждой анализируемой точке, прежде чем определить уровень R , проверяется с помощью модели (11.29) выполнение ограничения. Оно нарушается (что видно на рис. 11.16) в точках $1'$, $4'$, $4''$, $3'$, $3''$, для которых целевая функция не вычисляется; в точке $3''$, в частности, $\varepsilon = 2,85 < \varepsilon_{\text{тр}} = 5$. Дальнейшее движение комплекса до выполнения одного из рассмотренных ранее критериев остановки не показано, однако полученное уже в точке 8 решение (как видно на рис. 11.16, б) близко к оптимальному.

Таблица 11.6. Изменение параметров и целевой функции при поиске комплекс-методом

Точки	x_1	x_2	R	Точки	x_1	x_2	R
1	-1,00	0,60	2,44	4'	1,26	0,41	—
2	-0,80	1,00	2,32	4''	0,74	0,32	—
3	0,20	0,80	7,04	7	0,47	0,28	8,27
4	-0,60	0,10	5,91	3'	0,43	-0,91	—
5(2')	-0,03	-0,14	8,06	3''	0,36	-0,44	—
1'	0,97	-0,19	—	8	0,33	-0,20	8,54
6	0,42	0,03	8,47				

Методы спуска применимы для широкого класса задач нелинейного программирования. В задачах условной оптимизации с линейными ограничениями и целевой функцией (см. п. 11.5 — 11.7) используются специфические эффективные методы, построенные с учетом свойств таких линейных моделей.

11.5. Постановка задач линейного программирования

Одним из частных случаев постановки и решения задач оптимизации методами математического программирования являются построение, анализ и численное решение линейных оптимизационных

моделей. Как уже отмечалось (см. рис. II.4), задача линейного программирования (ЛП) имеет место лишь тогда, когда ограничения и целевая функция объекта управления (см. п. II.1) линейны относительно факторов X . В то же время это один из широко распространенных на практике методов математического программирования [2, 5, 34, 36, 59]. Он используется при решении производственных и управленческих задач в различных областях народного хозяйства, если параметры объекта моделирования отвечают требованиям аксиом делимости и аддитивности.

Принцип делимости (пропорциональности), в частности производственного процесса, означает, что общий (суммарный) объем потребления каждого вида ресурсов и соответствующая получаемая при этом прибыль предприятия *прямо пропорциональны* объему выпускаемой продукции. Так, если увеличить вдвое объем продукции, то потребление всех ресурсов и прибыль при этом также возрастут вдвое.

Аддитивность в этих же условиях заключается в том, что прибыль предприятия от реализации продукции есть *сумма* вкладов от реализации различных ее видов. В то же время полное количество каждого из потребляемых ресурсов равно сумме одноименных ресурсов, затраченных на каждом производстве всех видов продукции (например, общее потребление на заводе цемента за день складывается из его расходов на производство всей дневной номенклатуры изделий).

Построение математической модели для решения задачи ЛП обусловлено содержательной постановкой задачи и заключается в идентификации факторов X (которые в ЛП, как правило, называются *переменными*) и представлении цели и ограничений задачи в виде математических функций от этих переменных (см. п. II.1).

Пример II.14. Завод ЖБИ выпускает (рис. II. 17, а) декоративные плиты двух типов — P_1 и P_2 , затрачивая (табл. II.7) цветные минеральные наполнители (ЦМН) и цветные минеральные заполнители (ЦМЗ), единовременные ресурсы которых ограничены; доход от реализации продукции P_1 и P_2 различен (табл. II.7).

Переменные X , т. е. величины, подлежащие определению в результате решения задачи ЛП, в данном случае — объемы производства плит X_1 , шт., типа P_1 и X_2 , шт., типа P_2 . Ясно, что доход от реализации всех плит P_1 составит $3X_1$, руб (выполняется требование аксиомы пропорциональности), а от всех плит P_2 составит $2X_2$, руб, т. е. общий доход в соответствии с принципом аддитивности

$$C = 3X_1 + 2X_2. \quad (II.31)$$

Эта функция, отражающая результат деятельности предприятия, и является целевой — необходимо найти такие значения X_1 и X_2 , чтобы доход C был максимален. Однако при этом нельзя превысить ресурсы ни по одному из материалов:

$$70X_1 + 40X_2 \leq 168\ 000; \quad (II.32)$$

$$75X_1 + 100X_2 \leq 240\ 000. \quad (II.33)$$

Эти ограничения также отражают принцип аддитивности и следуют из анализа данных в табл. II.7 и схемы на рис. II.17, а. Неравенство (а не точное равенство) появляется в связи с тем, что максимальный доход может быть достигнут и в том случае, если запасы декоративного сырья исчерпаны не пол-

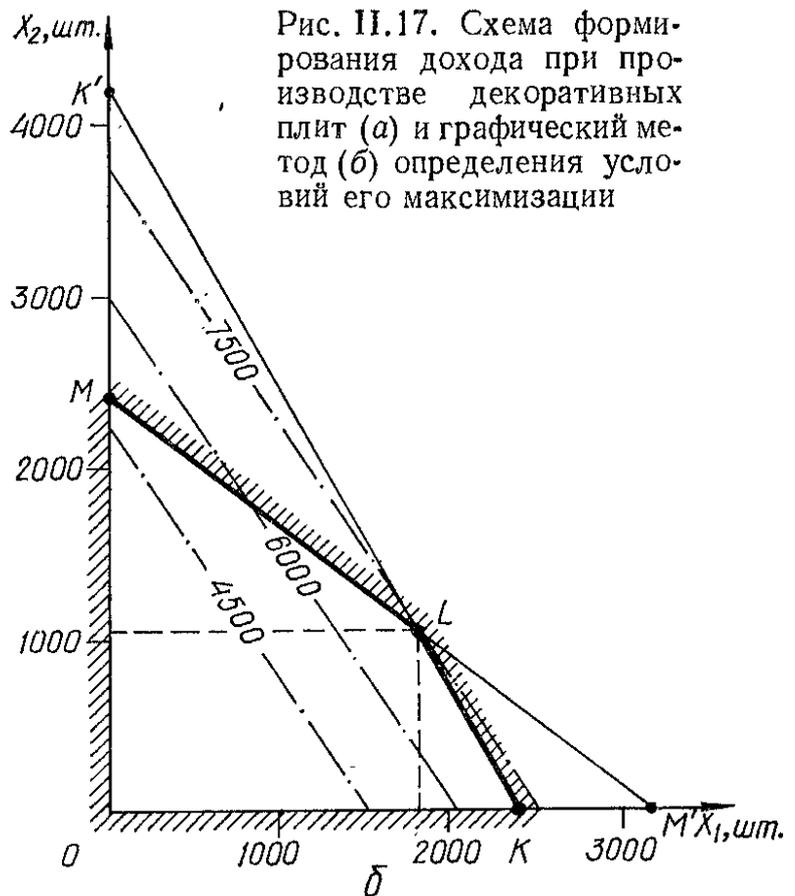
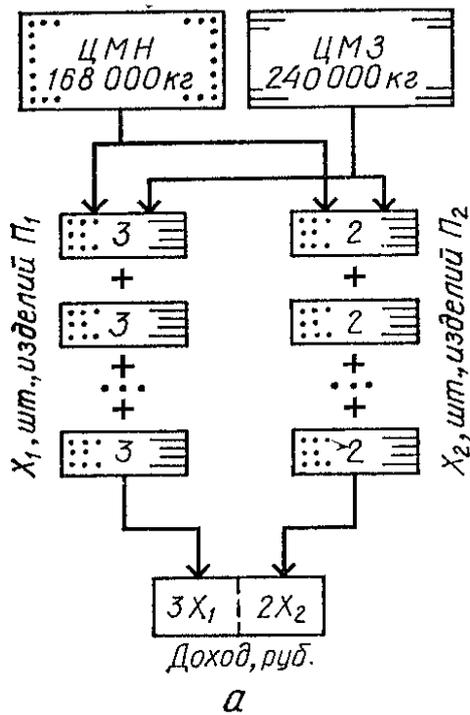


Рис. II.17. Схема формирования дохода при производстве декоративных плит (а) и графический метод (б) определения условий его максимизации

ностью (что само по себе также положительно скажется на хозяйственной деятельности предприятия). К (II.32), (II.33) следует добавить также *условие неотрицательности переменных* как натуральных результатов производства: $X_1 \geq 0, X_2 \geq 0$.

Таблица II.7. Условия производства декоративных плит на заводе ЖБИ*

Показатель	Тип плит		Ресурс материала, кг
	П ₁	П ₂	
Расход цветных материалов, кг			
наполнителей ЦМН	70	40	168 000
заполнителей ЦМЗ	75	100	240 000
Доход от выпуска одной плиты, руб.	3	2	—

* Здесь и далее в главе II технико-экономические данные условные.

Задача с двумя переменными может быть представлена и решена геометрически (рис. II.17, б) в координатах X_1 и X_2 . При этом ясно, что любое линейное ограничение типа (II.32) делит все множество $\{X_1, X_2\}$ на две полуплоскости: с одной стороны прямой (в данном случае нижней) лежат допустимые значения переменных, а с другой (верхней) — запрещенные.

Область допустимых значений $\Omega\{X_1, X_2\}$ определяется на первом этапе решения задачи. На рис. II.17, б вследствие неотрицательности переменных она находится в первом квадранте и ограничена координатными осями ($X_1 = X_2 = 0$) и двумя прямыми: $75X_1 + 100X_2 = 240\,000$ (проходит через точки $M\{3\,200; 0\}$ и $M'\{0; 2400\}$) и $70X_1 + 100X_2 = 168\,000$ (проходит через

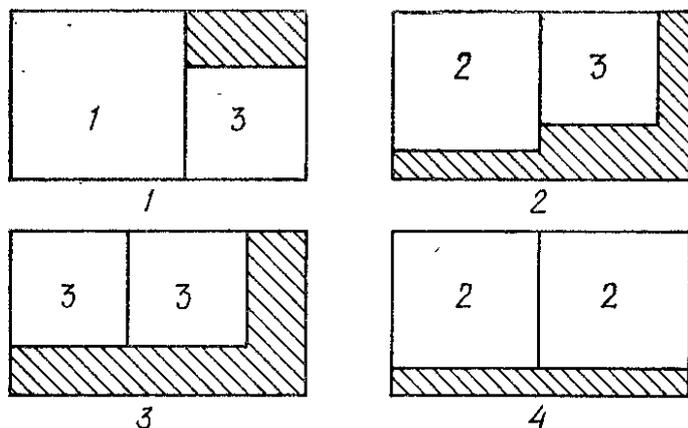


Рис. II.18. Четыре варианта раскроя панелей

точки $K\{2400; 0\}$ и $K\{0; 1680\}$, образующими выпуклый многоугольник. В каждой точке внутри него (незаштрихованная область Ω) выполняются все ограничения задачи.

Второй этап в решении задачи основан на анализе целевой функции (II.31) — на диаграмму наносятся ее изолинии при последовательно возрастающих на $\Delta C = \text{const}$ значениях общего дохода. Очевидно, что максимально возможное значение C_{max} в допустимой области $\Omega\{X_1, X_2\}$ достигается в вершине L многоугольника — на пересечении прямых (II.32) и (II.33). Координаты ее определяются из уравнения, полученного при приравнивании одной из переменных (X_1 , без потери общности) после ее выражения из равенств (II.32) и (II.33) в явном виде:

$$\begin{aligned} X_1 &= (168\,000 - 40X_2)/70; \\ X_1 &= (240\,000 - 100X_2)/75; \\ 70(240\,000 - 100X_2) &= 75(168\,000 - 40X_2); \\ X_2\{L\} &= 1050 \text{ и } X_1\{L\} = 1800. \end{aligned} \quad (\text{II.34})$$

Значение целевой функции в точке максимума $C\{L\} = 3 \cdot 1800 + 2 \cdot 1050 = 7500$ руб.

Большой класс задач, решаемых методами линейного программирования, образуют задачи о комплексном использовании сырья. Исходное сырье или материал может перерабатываться с помощью различных технологических приемов. При реализации каждого i -го приема получается k сочетаний видов продукции, причем для каждого j -го из этих k видов известны как выход продукции на единицу сырья, так и объем выпуска. Требуется найти такой план переработки, интенсивности применения способов $X_1, \dots, X_j, \dots, X_k$, при котором заданные объемы конечной продукции получались бы с наибольшей экономией исходного сырья (общие затраты ресурсов минимальны).

К числу наиболее важных отраслевых примеров задач этого типа относится задача рационального раскроя материала.

Пример II.35. При производстве ячеистого бетона по резательной технологии из однородного массива получают готовые мелкие блоки трех типоразмеров, причем возможны 4 варианта раскроя этого массива (рис. II.18). Необходимо получить не менее 10 блоков 1-го типа, 10 блоков 2-го типа и 16

Таблица II.8. Число блоков и отходы резания для задачи раскроя материала

Типоразмеры блоков ($l = 3$)	Способы раскроя ($k = 4$)			
	1	2	3	4
1	1	0	0	0
2	0	1	0	2
3	1	1	2	0
Отходы резания p_j , % по массе	13	32	41	17
Число реализаций j -го способа	X_1	X_2	X_3	X_4

блоков 3-го типа в каждой партии. Число блоков j -го типа при i -м способе раскрыя задается матрицей в табл. II.8, в которой в процентах от массы массива представлен объем отходов p_j (пропорционально заштрихованной площади на рис. II.18).

Формулировка задачи ЛП: необходимо соблюсти комплектность блоков и минимизировать отходы ячеисто-бетонной массы в среднем на один комплект изделий. Переменные задачи — числа реализаций X_1, X_2, X_3, X_4 каждого из способов. Отходы сырья складываются из потерь на обрезки пропорционально числу реализаций j -го способа резания, т. е. целевая функция формулируется как

$$C = \sum_{j=1}^4 p_j X_j = 13X_1 + 32X_2 + 41X_3 + 17X_4 \rightarrow \min$$

при ограничениях на комплектность

$$\begin{aligned} X_1 & \geq 10; \\ X_2 + 2X_4 & \geq 12; \\ X_1 + X_2 + 3X_3 & \geq 16 \end{aligned}$$

с учетом $X_j \geq 0$.

Другой класс образуют задачи о смесях и рецептурные задачи, когда требуется минимизировать общие затраты C , руб., на выпуск k смесей, состоящих из l компонентов, если известны:

для каждой j -й из k смесей отпускная цена d_j за единицу продукции и заданный объем выпуска (спрос) t_j единиц;

для каждого i -го из l компонентов его расход (содержание) q_{ij} в каждой j -й смеси и его общий ресурс Q_i .

Пример II.16. На бетоносмесительный завод прибывает щебень трех месторождений, причем для каждого j -го заполнителя известны его фракционный (гранулометрический) состав q_{ij} и стоимость d_j , руб/м³ (табл. II.9). В этой же таблице указано еженедельное потребное количество каждой

Таблица II.9. Фракционный состав q_{ij} и себестоимость щебня j -го месторождения

Фракция с размерами зерен, мм	Содержание i -й фракции в щебне j -го карьера, %			Потребное количество i -й фракции, м ³
	1	2	3	
1—5	20	10	30	50
5—10	40	20	20	300
10—20	35	40	15	200
20—40	5	30	35	100
Себестоимость щебня, руб/м ³	5	7	4	
Потребительный объем щебня j -го карьера, м ³	X_1	X_2	X_3	

фракции щебня Q_i , м³, заданное лабораторией, исходя из некоторых критериев оптимальности составов бетонов.

Необходимо определить поставки щебня X_j , м³, каждого j -го карьера, минимизируя общие затраты на заполнитель

$$C = \sum_{j=1}^3 d_j X_j = 5X_1 + 7X_2 + 4X_3 \rightarrow \min \quad (\text{II.35})$$

и выполняя заданные лабораторией условия-равенства

$$\begin{aligned} 0,2X_1 + 0,1X_2 + 0,3X_3 &= 50; \\ 0,4X_1 + 0,2X_2 + 0,2X_3 &= 300; \\ 0,35X_1 + 0,4X_2 + 0,15X_3 &= 200; \\ 0,05X_1 + 0,3X_2 + 0,35X_3 &= 100 \end{aligned}$$

при ограничении $X_j \geq 0$.

Пример II.17. Для приготовления трех глазурей j -го типа ($j = 1, 2, 3$) на керамическом заводе используются пять компонентов; сведения о рецептуре глазурей q_{ij} , ресурсе сырья Q_i , себестоимости глазури d_j и спросе t_j приведены в табл. II.10. Переменные – количества X_{ij} , кг, i -го компонен-

Таблица II.10. Рецептура глазурей, ресурсы сырья и технико-экономические условия реализации продукции

Наименование компонентов	Тип глазури			Ресурс сырья, кг
	1	2	3	
Перлит	0,5	0,6	—	1000
Датолит	0,5	—	—	600
Бой стекла	—	0,3	0,7	800
Беложгущаяся глина	—	0,1	0,2	400
Песок	—	—	0,1	300
Себестоимость глазури d_j , руб/кг	4	5	8	
Спрос t_j , кг, на j -ю глазурь	800	1000	500	

та, идущего на j -ю глазурь; с учетом принципов пропорциональности и аддитивности они формируют целевую функцию

$$\begin{aligned} C = \sum_{i=1}^5 \sum_{j=1}^3 d_j (q_{ij} X_{ij}) &= 4(0,5X_{11} + 0,5X_{21}) + 5(0,6X_{12} + 0,9X_{32} + 0,1X_{42}) + \\ &+ 8(0,7X_{33} + 0,2X_{43} + 0,1X_{53}). \end{aligned} \quad (\text{II.36})$$

Требуется по (II.36) определить X_{ij} при соблюдении ограничений на объем ресурсов

$$\begin{aligned} X_{11} + X_{12} &\leq 1000; \\ X_{21} &\leq 600; \\ X_{32} + X_{33} &\leq 800; \\ X_{42} + X_{43} &\leq 300; \\ X_{53} &\leq 300 \end{aligned}$$

и на потребность в продукции

$$\begin{aligned} X_{11} + X_{21} &\geq 800; \\ X_{12} + X_{32} + X_{42} &\geq 1000; \\ X_{33} + X_{43} + X_{53} &\geq 500, \end{aligned}$$

а также с учетом $X_{ij} \geq 0$.

Наибольшее распространение получил класс *задач транспортного типа*, к которым относятся собственно транспортная задача, детально анализируемая в п. II.7, а также задачи о назначениях или загрузке оборудования, о целераспределении, распределительная задача и др. [2, 5, 36, 59]. Модели транспортного типа можно применить также к решению таких задач, как прикрепление потребителей к поставщикам, оптимальное распределение автомобильного транспорта по маршрутам, расстановка рабочих по объектам и т. п.

Если имеется несколько технологических линий (агрегатов, комплексов, предприятий), на каждой из которых выполняются работы нескольких типов, то может быть сформулирована и решена *оптимизационная задача загрузки оборудования*. При заданном соотношении между объемами работ и известной производительности по каждому виду работ каждой линии определяется такой план их загрузки, при котором достигается максимальная общая производительность. Другим вариантом постановки задач может быть поиск такого распределения работ по линиям, который минимизирует суммарные затраты при условии, что известны затраты на выполнение каждой работы на каждой линии.

Пример II.18. Информация об удельной себестоимости q_{ij} , руб/шт, и затратах времени τ_{ij} , ч/шт, при изготовлении четырех марок ($j = \overline{1,4}$) железобетонных изделий на трех ($i = \overline{1,3}$) технологических линиях представлена в табл. II.11 (q_{ij} — над чертой, τ_{ij} — под чертой). Потребность строительных организаций в изделиях каждой марки равна p_j , шт.; фонд рабочего времени каждой линии — T_i .

Необходимо определить количество изделий X_{ij} i -й марки, производимых на i -й линии, исходя из минимизации суммарных затрат

$$\begin{aligned} C = \sum_{i=1}^3 \sum_{j=1}^4 q_{ij} X_{ij} = & 20X_{11} + 40X_{12} + 30X_{14} + 60X_{22} + \\ & + 30X_{23} + 20X_{24} + 40X_{31} + 20X_{33} + 50X_{34} \rightarrow \min. \quad (\text{II.37}) \end{aligned}$$

При этом должны выполняться ограничения по ресурсу времени для любой технологической линии ($\sum_{j=1}^4 X_{ij} \leq T_i$):

$$\begin{aligned} 2X_{11} + X_{12} + & 1,5X_{14} \leq 16; \\ & 2X_{22} + 1,5X_{23} + 0,5X_{24} \leq 16; \\ 2X_{31} + & X_{33} + 2X_{34} \leq 24, \end{aligned}$$

а количество произведенных изделий по номенклатуре должно удовлетворять потребности в них ($\sum_{i=1}^3 X_{ij} \geq p_i$):

Таблица II.11. Техничко-экономическая информация к задаче о загрузке оборудования

Номер технологической линии	Затраты q_{ij} и τ_{ij} на изготовление j -го изделия				Ресурс времени работы линии, T_i , ч
	1	2	3	4	
1	$\frac{20}{2}$	$\frac{40}{1}$	—	$\frac{30}{1,5}$	16
2	—	$\frac{60}{2}$	$\frac{30}{1,5}$	$\frac{20}{0,5}$	16
3	$\frac{40}{2}$	—	$\frac{20}{1}$	$\frac{50}{2}$	24
Потребность в изделиях p_j , шт.	10	6	4	8	

$$\begin{aligned} X_{11} + X_{31} &\geq 10; \\ X_{12} + X_{22} &\geq 6; \\ X_{23} + X_{33} &\geq 4; \\ X_{14} + X_{24} + X_{34} &\geq 8 \end{aligned}$$

при условии $X_{ij} \geq 0$.

В задаче оптимального размещения производства продукции одного вида (рис. II.19) анализируется информация:

о производстве (X_j — искомый объем производства в j -ом пункте ($j = \overline{1, k}$); s_j — себестоимость производства единицы продукции в j -м пункте);

о потреблении (b_i — известный объем потребления в пункте, ($i = \overline{1, l}$));

о перевозках (U_{ij} — искомый объем перевозок продукции из j -го пункта ее производства в i -й пункт потребления; a_{ij} — стоимость перевозки единицы продукции из j -го пункта производства в i -й пункт потребления).

Анализ этой информации позволяет сформулировать основные соотношения модели линейного программирования. Так, основным требованием сбалансированности производства и потребления является равенство количеств продукции, выпускаемой в j -м пункте и отправляемой из этого пункта во все пункты потребления:

$$X_j = \sum_{i=1}^l U_{ij}. \quad (\text{II.38})$$

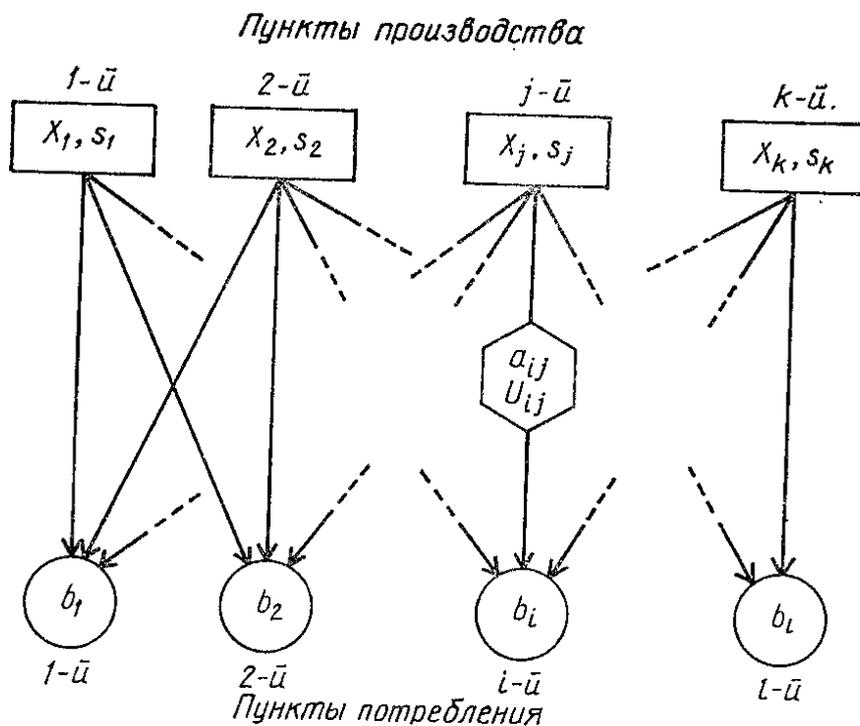


Рис. II.19. Схема взаимосвязи пунктов производства и потребления

Потребность в продукции в i -ом пункте потребления должна покрываться доставкой из пунктов производства, причем не должно возникать избытка продукции как в пунктах производства, так и потребления:

$$\sum_{j=1}^k U_{ij} = b_i. \quad (\text{II.39})$$

Целевая функция — сумма стоимостей производства и перевозки продукта — должна достигать минимума

$$C = \sum_{j=1}^k s_j X_j + \sum_{i=1}^l \sum_{j=1}^k a_{ij} U_{ij}. \quad (\text{II.40})$$

Особенностью рассматриваемой задачи ЛП является то, что ограничения представлены в виде равенств, а для определения $k + l$ неизвестных переменных (X_j и U_{ij}) имеется лишь $k + l$ уравнений (II.38) и (II.39).

Рассмотренные выше отраслевые технико-экономические примеры позволяют выделить ряд общих признаков математических моделей линейных систем:

а) целевая функция — линейная форма (см. прил. I.1)

$$C = p_1 X_1 + p_2 X_2 + \dots + p_k X_k = \sum_{j=1}^k p_j X_j \quad (\text{II.41})$$

в зависимости от цели управления должна иметь в заданной области максимальное или минимальное значение;

б) значения k подлежащих определению переменных (параметров, факторов) неотрицательны:

$$X_j \geq 0 \quad (j = \overline{1, l_0}, l_0 \leq k); \quad (\text{II.42})$$

в) на переменные наложены $m = l_1 + l_2$ ограничений в виде l_1 неравенств

$$\sum_{j=1}^k a_j X_j \leq b; \quad (\text{II.43})$$

$$\sum_{j=1}^k a_j X_j \geq b \quad (\text{II.44})$$

и в виде l_2 равенств

$$\sum_{j=1}^k a_j X_j = b. \quad (\text{II.45})$$

Общность признаков математических моделей обуславливает и общность методов поиска решения в задачах ЛП (см. п. II.5, II.6). Совокупность чисел $X = (X_1, X_2, \dots, X_k)$, удовлетворяющих ограничениям задачи (II.42 — II.45), называется *допустимым решением* или *планом** (см. любую из точек в незаштрихованной области Ω на рис. II.17). План $X = (X_1^*, X_2^*, \dots, X_k^*)$ оптимален, если при нем целевая функция (II.41) достигает минимума $C^* \leq C(X)$ или максимума $C^* \geq C(X)$. Наиболее удобные алгоритмы поиска оптимума построены для так называемой *канонической*, или *основной*, задачи линейного программирования (ОЗЛП).

Цель решения ОЗЛП состоит в определении по (II.41) C при условиях неотрицательности переменных (II.42) и ограничениях-равенствах (II.45) с неотрицательной правой частью ($b \geq 0$). В канонической задаче не должно быть ограничений-неравенств ($l_1 = 0, l_2 = m$), поэтому условия (II.43), (II.44) преобразуются за счет введения в левую часть дополнительных неотрицательных переменных \tilde{X}_j :

$$(a_1 X_1 + \dots + a_k X_k) + \tilde{X}_{k+1} = b^{(1)}; \quad (\text{II.46})$$

$$(a_1 X_1 + \dots + a_k X_k) - \tilde{X}_{k+2} = b^{(2)}. \quad (\text{II.47})$$

Эти переменные \tilde{X}_j иногда называются *логическими*, в отличие от исходных X_j (отражающих существо объекта) *структурных* переменных. Однако и логическая переменная не лишена смысла — ее можно интерпретировать: а) при знаке « \leq » в исходном уравнении, например, как экономию средств \tilde{X}_{k+1} на выпуск $X_1 + \dots + X_k$ изделий по сравнению с имеющимся ресурсом $b^{(1)}$; б) при знаке « \geq » как, например, выигрыш в доходе \tilde{X}_{k+2} при выпуске $X_1 + \dots + X_k$ изделий по сравнению с заданным нормативом $b^{(2)}$. Очевидно, что дополнительные переменные можно ввести и в уравнение целевой функции (II.41), где они будут иметь нулевые коэффициенты. Важно подчеркнуть, что за переход от начальной за-

* Этот термин отражает историзм создания такого раздела математики, как линейное программирование, которое первоначально охватывало решение технико-экономических задач о планировании перевозок и т. п. Термин «план» в гл. III—IV будет по смыслу отличен от обсуждаемого.

дачи к канонической нужно расплачиваться увеличением числа переменных от k до $n = k + l_1$. При этом следует отметить, что возможен и обратный переход от ОЗЛП к задаче с ограничениями-неравенствами [5, 59 и др.]. Для обеспечения же неотрицательности правой части ограничения (II.45) обе его части можно умножить на -1 .

Пример II.19. При выпуске декоративных плит (см. пр. II. 14) ограничения на ресурс из неравенств (II.32), (II.33) преобразуются за счет введения дополнительно двух логических переменных в равенства

$$70X_1 + 40X_2 + \tilde{X}_3 = 168\ 000; \quad (\text{II.48})$$

$$75X_1 + 100X_2 + \tilde{X}_4 = 240\ 000. \quad (\text{II.49})$$

В этой системе из двух уравнений теперь четыре неизвестных, они же, вообще говоря, входят и в целевую функцию

$$C = 3X_1 + 2X_2 + 0\tilde{X}_3 + 0\tilde{X}_4. \quad (\text{II.50})$$

В общем виде ОЗЛП, таким образом, сводится к тому, чтобы минимизировать (максимизировать) в n -мерном пространстве (здесь и далее учитываются без различия и структурные, и логические переменные с общим обозначением X_j) линейную форму

$$\begin{aligned} C &= p_1X_1 + \dots + p_jX_j + \dots + p_nX_n = \\ &= \sum_{j=1}^n p_jX_j \rightarrow \min(\max), \end{aligned} \quad (\text{II.51})$$

удовлетворяя m условиям-равенствам (здесь и далее подразумевается, что все ограничения-неравенства преобразованы и в особом выделении \tilde{X} нет необходимости):

$$\left. \begin{aligned} a_{11}X_1 + \dots + a_{1j}X_j + \dots + a_{1n}X_n &= b_1, \\ \dots & \dots \\ a_{i1}X_1 + \dots + a_{ij}X_j + \dots + a_{in}X_n &= b_i, \\ \dots & \dots \\ a_{m1}X_1 + \dots + a_{mj}X_j + \dots + a_{mn}X_n &= b_m \end{aligned} \right\} \quad (\text{II.52})$$

при n условиях неотрицательности переменных $X_j \geq 0$.

В матричной форме вышесказанное записывается с помощью вектор-строки p размером $1 \times n$ — вектора доходов или оценок ОЗЛП; вектор-столбца X размером $n \times 1$, называемого вектором переменных; матрицы a размером $m \times n$ — матрицы коэффициентов; вектор-столбца b размером $m \times 1$ — вектора ресурсов:

$$C = pX \rightarrow \min(\max); \quad (\text{II.53})$$

$$aX = b \quad (b \geq 0); \quad (\text{II.54})$$

$$X \geq 0. \quad (\text{II.55})$$

Из самой постановки ОЗЛП ясно, что число уравнений m меньше, чем число неизвестных $n \geq m$, поэтому система неопределенна, и если она совместна, то имеет бесчисленное множество решений.

Все неотрицательные решения (II.55) систем (II.53) — (II.54) являются допустимыми — это неотрицательные векторы размера $n \times 1$. *Оптимальным решением* X^* является такой допустимый вектор, который обращает линейную целевую функцию (II.51), (II.53) в минимум (максимум) C^* . Оптимальное решение *неединственно*, когда существует более одного допустимого значения X со значением целевой функции, равным C^* . Если C^* бесконечно, то задача ЛП имеет *неограниченный оптимум*.

Геометрическим аналогом множества допустимых решений является бесконечное множество точек, принадлежащих k -мерной области $\Omega \subseteq R^k$, образованной в виде k -мерного *выпуклого многогранника* t плоскостями-ограничениями (II.51), (II.54) и n плоскостями (II.55), ограничивающими неотрицательные решения. Число вершин такого многогранника $N_{mn} \leq n!/[m!(n - m)!]$.

Поскольку целевая функция линейная, то ее градиент постоянен ($\nabla C(X) = \text{const}$) в любой точке многогранника допустимых решений. Движение от любой внутренней точки в направлении градиента ∇C приведет в лучшую точку. Неулучшаемая точка может лежать только на границе многогранника, следовательно, *оптимальное решение* всегда находится в *одной из вершин* или на одной из граней (т. е., в частности, в соседних вершинах). Поэтому из всех допустимых решений (или планов) наибольший практический интерес для ускорения поиска C^* представляют те $X = (X_1, \dots, X_j, \dots, X_n)$, которые совпадают с *вершинами*, так как только им может *соответствовать минимум* (максимум) целевой функции. В процессе такого поиска C^* вектор X движется от вершины к вершине *по границам* допустимой области Ω , поэтому часть его элементов X_j обязательно должна быть на нулевом уровне; эти переменные $X_j = 0$ получают временное название *небазисных*, в отличие от *базисных*, которые в данный момент поиска не равны нулю. Вершины многогранника играют при этом особую роль — каждая из них определяет так называемый *опорный план* [12], или *базисное решение* [44].

Естественно, что оптимальное решение проще всего было бы найти перебором вершин, но при большой размерности реальных задач это практически неосуществимо (так, при числе факторов $k = 5$ и ограничений $l_1 = m = 8$ и $n = k + l_1 = 13$, а $N_{nm} \leq \leq 13!/[8!5!] = 1287$). В то же время необходима информация о *координатах хотя бы одной из вершин*, так как от нее должен начаться шаговый переход к другим, «лучшим», вершинам.

Пример II.20. Ограничения на ресурс (II.48), (II.49) в задаче о выпуске декоративных плит (см. пр. II.14, II.19) представлены двумя уравнениями с четырьмя неотрицательными неизвестными X_1, \dots, X_4 :

$$\left. \begin{aligned} 70X_1 + 40X_2 + X_3 &= 168\,000; \\ 75X_1 + 100X_2 + X_4 &= 240\,000; \end{aligned} \right\} \quad (\text{II.56})$$

$$X_1 \geq 0, X_2 \geq 0, X_3 \geq 0, X_4 \geq 0. \quad (\text{II.57})$$

Одно из возможных решений неопределенной системы (II.56) можно найти, приравнявая какие-либо $n - m$ переменных нулю; в частности, если $X_1 = 0$ и $X_2 = 0$, то

$$X_3 = 168\,000 \quad \text{и} \quad X_4 = 240\,000,$$

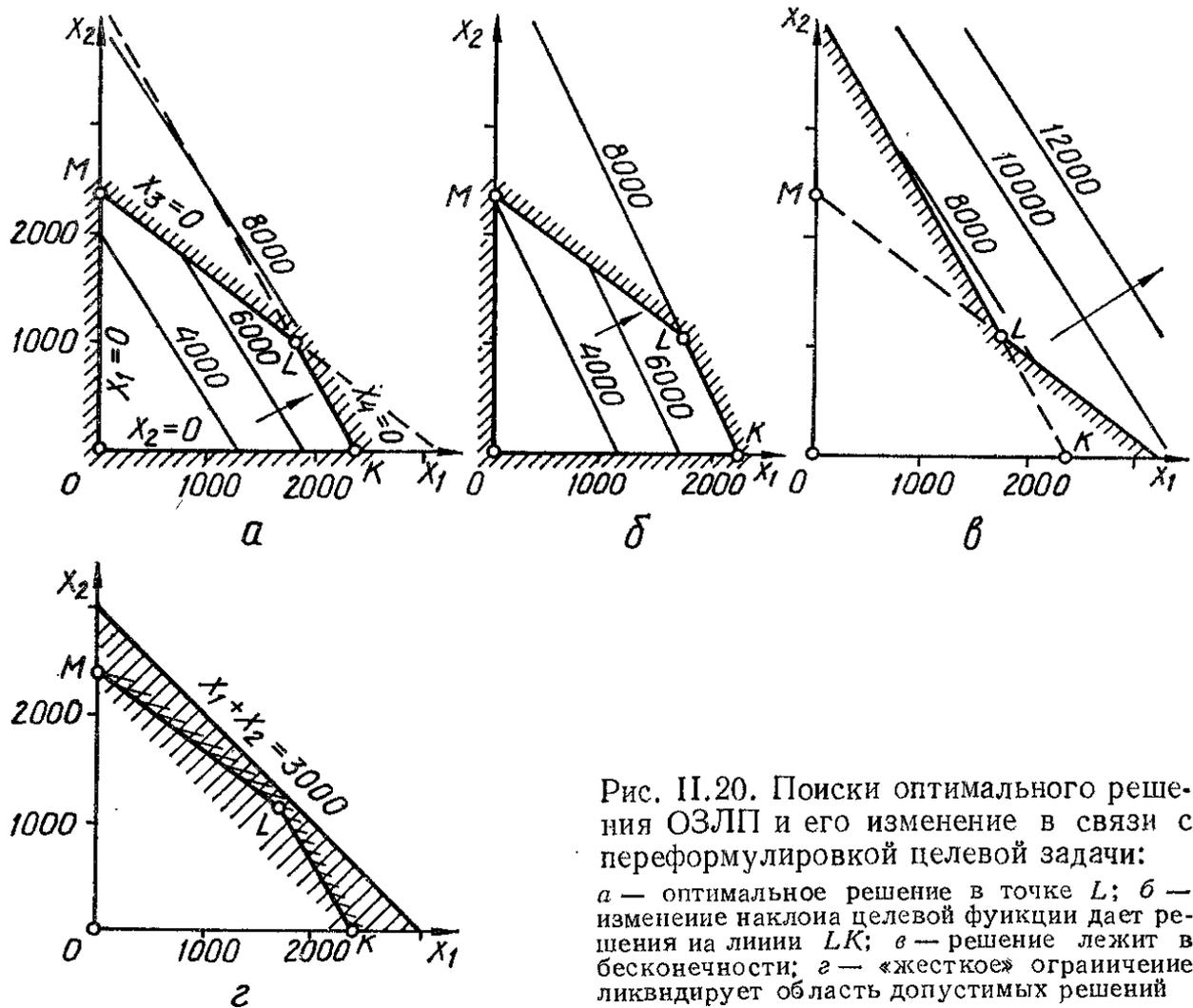


Рис. 11.20. Поиски оптимального решения ОЗЛП и его изменение в связи с переформулировкой целевой задачи:

a — оптимальное решение в точке *L*; *б* — изменение наклона целевой функции дает решения на линии *LK*; *в* — решение лежит в бесконечности; *г* — «жесткое» ограничение ликвидирует область допустимых решений

причем целевая функция (11.50) в этой точке $C = 0$, так как приравнены нулю структурные переменные.

Всего решений, полученных таким способом по (11.56), может быть $n!/[m!(n-m)!] = 4!/(2!2!) = 6$; они соответствуют точкам попарного пересечения четырех прямых, описываемых в двумерном пространстве структурных переменных $X_1 \geq 0, X_2 \geq 0, X_3 = X_4 = 0$ уравнениями (11.56) и (11.57). Результаты решения представлены в табл. 11.12 и на рис. 11.20, *a*.

Таблица 11.12. Возможные решения системы (11.56) — координаты вершин многогранника и соответствующие им значения целевой функции C

Номер варианта	Приравненные нулю переменные X_j	Решение системы (11.56) или координаты вершины				Точка	Значение целевой функции C (11.50)
		X_1	X_2	X_3	X_4		
1	$X_1 = X_2 = 0$	0	0	168 000	240 000	<i>O</i>	0
2	$X_1 = X_3 = 0$	0	4200	0	-180 000	Недопустимое решение ($X_4 < 0$)	—
3	$X_1 = X_4 = 0$	0	2400	72 000	0	<i>M</i>	4800
4	$X_2 = X_3 = 0$	2400	0	0	60 000	<i>K</i>	7200
5	$X_2 = X_4 = 0$	3200	0	-56 000	0	Недопустимое решение ($X_3 < 0$)	—
6	$X_3 = X_4 = 0$	1800	1050	0	0	<i>L</i>	7500

На рис. II.20, а в двумерном пространстве $\{X_1, X_2\}$ показана область допустимых решений Ω в виде четырехугольника $OKLM$: внутри незаштрихованной области ни одна из переменных X_j не равна нулю, на ребрах (гранницах) многогранника равна нулю одна переменная; каждой вершине соответствует вектор, у которого две переменные имеют нулевое значение; вершинам соответствуют 4 опорных, или базисных, решения, используя которые, можно легко найти (см. п. II.6) оптимальное решение C^* в вершине L (см. рис. II.17).

Если первый опорный план будет соответствовать вершине O с координатами $X_1 = X_2 = 0; X_3 \neq 0$, то следующей (в направлении градиента ∇C) окажется вершина $K \{X_1 \neq 0, X_2 = X_3 = 0; X_4 \neq 0\}$, а потом и наилучшая $L \{X_1 \neq 0; X_2 \neq 0; X_3 = X_4 = 0\}$. Каждый вектор X , соответствующий вершине, имеет элементы X_j , отличные от нуля (общим числом $n - k = 2$), и $X_j = 0$. Переменные $X_j \neq 0$ — базисные, а $X_j = 0$ — небазисные. При переходе от вершины к вершине часть из них меняет свою роль (свое назначение, название): например, в точке O небазисными были X_1 и X_2 , но при движении по горизонтальному ребру OK переменная X_1 начала возрастать, а переменная X_4 — стремиться к нулю и в результате в точке K небазисными стали $X_2 = X_4 = 0$. Этот процесс используется далее при построении методов поиска C^* .

Если изменится технико-экономическая ситуация при производстве плит, то изменяется и конкретная ОЗЛП, и результаты ее решения. Так, если соотношение доходов для плит P_1 и P_2 изменится от $3 : 2$ (II.50) до $3,5 : 2$, то изолинии целевой функции $C = 3,5X_1 + 2X_2$ пойдут (рис. II.20, б) параллельно стороне LK многоугольника, а следовательно, оптимальное решение охватит множество точек на этом ребре от вершины L до вершины K . Если же ограничения переформулированы так, что не разрешается использовать сырья меньше, чем имеющийся ресурс b_1 и b_2 , причем верхний уровень расхода ничем не ограничен, то в такой ситуации (достаточно нелепой, но возможной) оптимальное решение лежит в бесконечности (рис. II.20, в). И наконец, может быть введено слишком жесткое добавочное ограничение (в частности, на суммарный выпуск плит $X_1 + X_2 \geq 3000$ без увеличения ресурсов), которое «закроет» область допустимых решений. Такое «пустое» из-за несовместимости ограничений множество показано на рис. II.20, г.

Упорядоченный перебор опорных планов при поиске оптимальности ведется различными методами решения задач ЛП [12, 36, 43 и др.], однако большинство из них используют переход от одного из базисных решений к близлежащему лучшему.

II.6. Симплекс-метод решения задач линейного программирования

При большом числе ограничений решение задачи ЛП прямым перебором всех вершин многогранника области допустимых решений чрезвычайно громоздко. Систематический вычислительный метод, позволяющий находить численные решения для линейных оптимизационных моделей за меньшее число шагов, разработан Дж. Данцигом и называется *симплекс-методом** [12, 43 и др.]. Помимо упрощения чисто вычислительной процедуры симплекс-метод позволяет анализировать и правильно интерпретировать получаемые решения на каждом шаге выполнения алгоритма.

* В данном методе термин «симплекс» не имеет того геометрического смысла, который был вложен в него в п. I.2 — I.4.

Процесс решения начинается с определения начальной вершины многогранника решений, координаты которой образуют *начальное допустимое базисное решение*, или *начальный опорный план*. Оптимальное решение находится как результат последовательного улучшения текущего базисного решения в направлении градиента целевой функции (II.53). Переход от одного допустимого базисного решения к другому осуществляется по тому ребру многогранника решений, которое имеет наибольшую по абсолютному значению проекцию градиента ∇C .

В выражении целевой функции C ненулевые коэффициенты соответствуют небазисным переменным. Для получения оптимального решения C выражают через другие (базисные) переменные, которые, таким образом, приобретают отрицательные коэффициенты. Так как все переменные неотрицательные, то C достигает максимума, когда эти базисные переменные равны нулю. Эта процедура — существо *операции замены базиса*.

Координаты каждой последующей вершины многогранника определяются заменой одной из нулевых переменных текущего решения текущей базисной переменной. Так как число вершин многогранника конечно, то оптимальное решение находится за конечное число шагов, причем меньшее, чем число вершин.

Эти положения обобщаются *итерационным алгоритмом* (для случая поиска максимума целевой функции), который называется по традиции симплекс-методом.

Этап 1. На основании формулировки ОЗЛП определяется начальное допустимое базисное решение, которое находится как решение системы линейных уравнений (II.52), (II.54), получаемой приравнением нулю $n - m$ переменных, которые приобретают статус небазисных. Это обеспечивает единственность и однозначность решения.

Пример II.21. Допустимое начальное решение задачи в пр. II.19, II.20 можно получить, полагая обе небазисные переменные равными нулю. Тогда начальная вершина т. 0 на рис. II.20, а имеет координаты $x_1 = 0$; $X_2 = 0$; $X_3 = 168000$; $X_4 = 240000$; значение целевой функции в этой вершине равно нулю (см. табл. II.12).

Этап 2. Производится проверка возможности улучшения значения целевой функции за счет выбора и увеличения одной из переменных, приравненных вначале нулю. Если такой переменной в задаче не существует, то вычисления прекращаются, поскольку имеющееся базисное решение оказалось оптимальным. В противном случае выбранная переменная включается в новый базис. Является ли полученный опорный план наилучшим, определяется по уравнению целевой функции C . Так, в частности, анализируя уравнение (II.50), можно заметить, что обе небазисные переменные $X_1 = X_2 = 0$ имеют положительные коэффициенты в исходной целевой функции, следовательно, с ростом X_1 и X_2 от нуля будет увеличиваться и целевая функция ($C \rightarrow C_{\max}^*$). Так как целевая функция линейная, то ее увеличения, очевидно, можно добиться быстрее,

используя для включения в базис ту переменную, которая имеет наибольший коэффициент; при этом оптимум достигается быстрее, за меньшее число шагов. Это свойство позволяет сформулировать условие оптимальности, или так называемый симплекс-критерий I (максимизация): *если все небазисные переменные в S -уравнении имеют отрицательные коэффициенты, полученное допустимое базисное решение является оптимальным; в противном случае в качестве включаемой в базис переменной следует взять ту, коэффициент при которой наибольший.*

Этап 3. Задается предельное (максимальное) значение включаемой переменной, при котором одна из базисных переменных — исключаемая переменная — не обратится в нуль. Ее исключают из выражения для целевой функции.

Процедура выбора исключаемой переменной в общем случае также должна корректироваться условиями задачи. Очевидно, что с увеличением включаемой в базис переменной сильнее возрастает значение S . Однако увеличение X_j возможно лишь за счет уменьшения значений базисных переменных в каждой строке, содержащей X_j с положительным коэффициентом. Поэтому решение о выборе исключаемой переменной должно согласовываться с проверкой условия допустимости, требующего, чтобы в качестве исключаемой бралась та переменная, которая первая обращается в нуль при увеличении X_j до значения, соответствующего смежной экстремальной точке. Алгебраическое обобщение этого положения следует из геометрических представлений (см. рис. II.17). Максимально допустимое значение переменной X_1 определяется ближайшей к нулю точкой оси X_1 , в которой одна из прямых-ограничений пересекает эту ось. Координата точки пересечения соответствует значению отношения правой части ограничения к положительному коэффициенту при включаемой переменной. Очевидно, что линии-ограничения, у которых коэффициенты при X_1 отрицательны или равны нулю, не пересекают ось X_1 в области положительных значений, а следовательно, не могут быть исключены на данном шаге. *Исключаемой переменной будет та переменная текущего базиса, для которой отношение правой части ограничения к коэффициенту при включаемой переменной минимально* — это так называемый симплекс-критерий II, или условие допустимости.

Этап 4. Методом Гаусса—Жордана (см. п. I.4) решается система $n - k$ уравнений относительно новых базисных переменных и получается новое текущее базисное решение. Анализ нового опорного плана начинается с этапа 2.

Изложенный алгоритм приводит к оптимальному решению для любой правильно сформулированной задачи ЛП за конечное число итераций. Количество итерационных процедур (т. е. возвратов с четвертого на второй этап) симплекс-метода всегда меньше числа вершин N_{mn} . Последовательные итерации симплекс-алгоритма удобно представлять в компактном виде, используя симплекс-таблицы (табл. II.13), где $X_{k+1}, X_{k+2}, \dots, X_n$ — базисные переменные. Самая нижняя строка, называемая *индикаторной*, содержит коэффици-

Таблица II.13. Общий вид начальной симплекс-таблицы

Номер базисного решения	Базис	Коэффициенты при переменных								Правая часть	$\frac{b_j}{a_{ij}}$
		X_1	X_2	\dots	X_k	X_{k+1}	X_{k+2}	\dots	X_n		
0	X_{k+1}	a_{11}	a_{12}	\dots	a_{1k}	1	0	\dots	0	b_1	
	X_{k+2}	a_{21}	a_{22}	\dots	a_{2k}	0	1	\dots	0	b_2	
	X_n	a_{m1}	a_{m2}	\dots	a_{mk}	0	0	0	1	b_m	
Индикаторная строка		p_1	p_2	\dots	p_k	0	0	\dots	0	C	

енты при переменных в уравнении целевой функции

$$p_1X_1 + p_2X_2 + \dots + p_kX_k + 0X_{k+1} + \dots + 0X_n = C, \quad (\text{II.58})$$

или *индикаторы*. Как следует из симплекс-алгоритма, цель преобразований базиса состоит в том, чтобы при помощи составления линейных комбинаций строк табл. II.13 достичь на некоторой l -й итерации исключения всех положительных индикаторов. Последовательность операций с использованием симплекс-таблиц представлена на блок-схеме (рис. II.21). Из нее следует, что все элементы столбца правых частей b_i не меняют знак при преобразованиях, а выражение в правой части индикаторной строки только убывает (что приводит, соответственно, только к росту целевой функции).

Пример II.22. На основании данных пр. II.19, II.20 составляется начальная симплекс-таблица II.14, из которой видно, что переменные X_1 и X_2 в индикаторной строке имеют положительные коэффициенты (а решается задача максимизации), следовательно, данное допустимое базисное решение не является оптимальным. Согласно симплекс-критерию I, в качестве включаемой на следующем этапе в базис переменной следует выбрать переменную X_1 , имеющую наибольший положительный коэффициент в индикаторной строке.

Условие допустимости позволяет определить исключаемую из базиса переменную. Эта переменная X_3 соответствует *ведущей* строке, содержащей *центральный элемент* — элемент столбца при X_1 , имеющий наименьшее *положительное значение* отношения правой части к нему самому (обведено рамкой): отношение $168\ 000/70 = 2400$. Из базиса исключается X_3 .

Алгебраический переход от текущего к последующему допустимому базисному решению (этап 4) удобно осуществлять в две операции (табл. II.15).

Первая операция — деление текущей строки 1 в табл. II.14, содержащей X_3 , на центральный элемент ($a_{11} = 70$) — приводит к новой ведущей строке (строка 1 в табл. II.15). Вторая операция дает возможность получить остальные строки симплекс-таблицы. Для этого из каждой строки предыдущего решения (строка 2 в табл. II.14) поэлементно вычитают новую ведущую строку (строка 1), умноженную (строка 3) на соответствующий ей коэффициент ведущего столбца текущего решения. Для строки 2 табл. II.15, содержащей базисную переменную X_4 , ведущий коэффициент, соответствующий X_1 , равен 75.

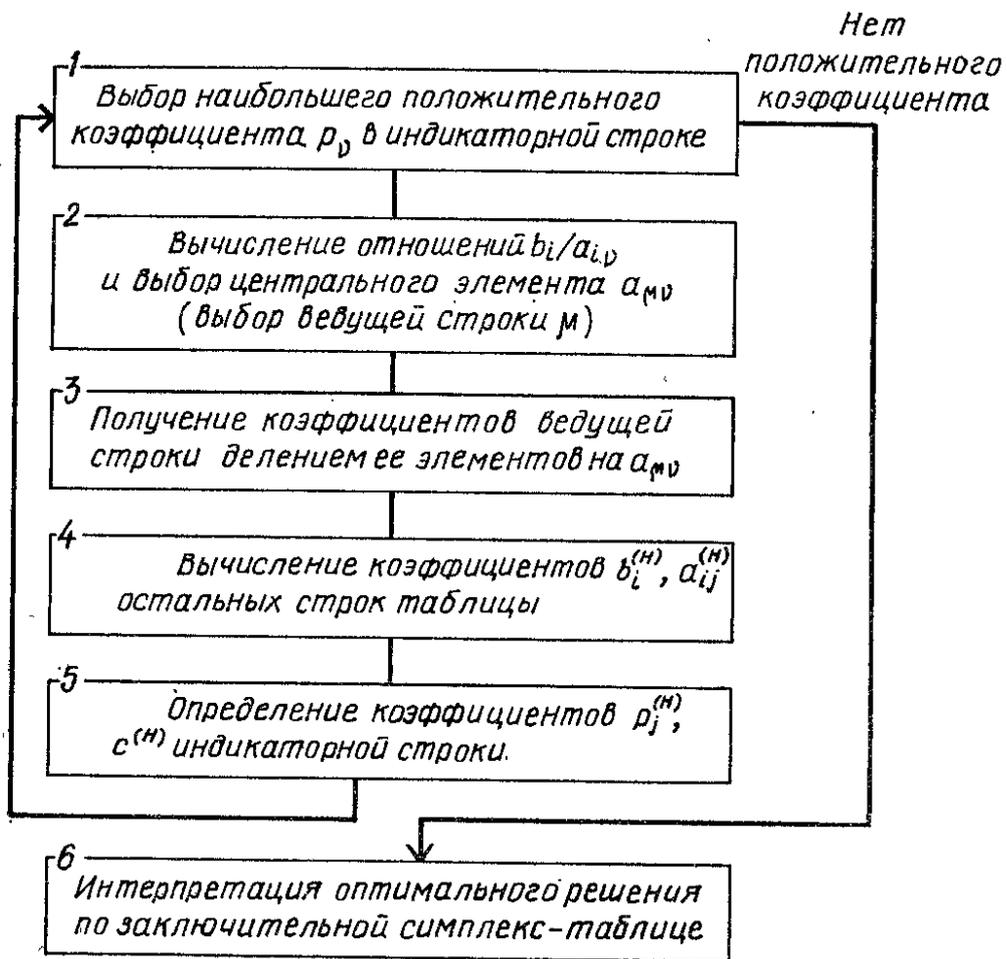


Рис. II.21. Укрупненная блок-схема алгоритма симплекс-метода решения ОЗЛП

Таблица II.14. Начальная таблица симплекс-алгоритма решения задачи планирования производства декоративных плит

Номер базисного решения	Базисная переменная	Коэффициенты при переменных				Правая часть	b_j / a_{ij}
		X_1	X_2	X_3	X_4		
0	X_3	70	40	1	0	168 000	2400
	X_4	75	100	0	1	240 000	3200
Индикаторная строка		3	2	0	0	C	

Аналогичные операции проводятся с индикаторной строкой при ведущем коэффициенте 3. Результаты алгебраического перехода приведены в табл. II.16. Она содержит положительный индикатор (2/7) при X_2 , поэтому новое базисное решение еще не является оптимальным. Однако на его основании выполняется следующий шаг. Включаемой в базис переменной, очевидно, будет X_2 ; минимальное из отношений b_i / a_{ij} имеет строка с базисной переменной X_4 . Следовательно, элемент в рамке будет центральным, а переменная X_4 удаляется из базиса. Промежуточные преобразования, выполненные на этапе 4, приводят к табл. II.17, которая описывает новое базисное допустимое решение.

Таблица 11.15. Алгебраические переходы от текущего базисного решения (табл. 11.14) к последующему (табл. 11.16)

Номер строки	Операция	Коэффициенты при переменных				Правая часть
		X_1	X_2	X_3	X_4	
1	Образование базиса X_1 из X_3 — новая ведущая строка	1	4/7	1/70	0	2400
2	Строка X_4 из табл. 11.14	75	100	0	1	240000
3	Произведение новой ведущей строки 1 на ведущий коэффициент 75 из строки 2	75	300/7	15/14	0	180000
4	Разность строк 2—3 — строка X_4 нового решения	0	400/7	-15/14	1	60000

Таблица 11.16. Промежуточное допустимое базисное решение задачи планирования производства

Номер текущего базисного решения	Базисные переменные	Коэффициенты при переменных				Правая часть	b_i/a_{ij}
		X_1	X_2	X_3	X_4		
1	X_1	1	4/7	1/70	0	2400	4200
	X_4	0	$\boxed{400/7}$	-15/14	1	60 000	1050
Индикаторная строка		0	2/7	-3/70	0	$C - 7200$	

Здесь в индикаторной строке нет положительных индикаторов при переменных, значит табл. 11.17 является окончательной и выражает оптимальное допустимое решение задачи.

Таблица 11.17. Оптимальное решение задачи планирования производства

Номер текущего базисного решения	Базисные переменные	Коэффициенты при переменных				Правая часть	b_i/a_{ij}
		X_1	X_2	X_3	X_4		
2	X_1	1	0	1/40	-1/100	1800	—
	X_2	0	1	-3/160	7/400	1050	—
Индикаторная строка		0	0	-3/80	-1/200	$C - 7500$	

Применение симплекс-алгоритма к решению задачи пр. II.22 основывалось на сравнительно легко полученном начальном базисном решении, которое последовательно улучшалось вплоть до достижения оптимума. Однако существуют задачи, в которых отыскать начальный опорный план непросто. К ним относятся задачи с моделями, содержащими ограничения в виде неравенства « \geq » (приводимые к каноническому виду с помощью введения избыточных логических переменных, см. п. II.5) и строгие равенства.

Пример II.23. К условиям пр. II.14 добавлено условие выполнения производственного задания по общему количеству производимых изделий: суммарное количество панелей типов Π_1 и Π_2 равно 2500. Тогда ограничения (II.32), (II.33) примут вид

$$X_1 + X_2 = 2500, \quad (\text{II.59})$$

$$70X_1 + 40X_2 + \tilde{X}_3 = 168\,000, \quad (\text{II.60})$$

$$70X_1 + 100X_2 + \tilde{X}_4 = 240\,000. \quad (\text{II.61})$$

Эта система содержит четыре неизвестные $X_1, X_2, \tilde{X}_3, \tilde{X}_4$ и три уравнения, т. е. $m = 3, n = 4$. Поэтому каждое базисное решение включает одну небазисную переменную ($X_j = 0$). Если это переменная X_1 , то из (II.59) следует, что $X_2 = 2500$. Подстановка этих значений в (II.61) приводит к тому, что переменная $X_4 = 240\,000 - 100 \cdot 2500 = -100$, т. е. приобретает отрицательное значение, что противоречит и условиям ОЗЛП, и технико-экономическому смыслу.

Если условие выполнения производственной программы (II.59) сформулировать в виде неравенства « \geq » (больше или равно) и ввести избыточную логическую переменную $\tilde{X}_5 \geq 0$, то

$$X_1 + X_2 - \tilde{X}_5 = 2500,$$

и допустимое базисное решение должно содержать $n - m = 5 - 3 = 2$ нулевые переменные. Но и в этом случае при приравнивании нулю X_1 и X_2 переменная \tilde{X}_5 оказывается отрицательной.

Ситуация, когда нет возможности непосредственно найти опорный план, встречается во многих задачах, в которые входят ограничения, не содержащие остаточных переменных. Для преодоления этой трудности используются модификации симплекс-метода, основанные на введении добавочных искусственных переменных (W).

Метод использования искусственных переменных, или *метод искусственного базиса*, предполагает включение новых неотрицательных переменных $W \geq 0$ в левую часть всех тех ограничений, которые не содержат *остаточных* логических переменных $\tilde{X}_j \geq 0$. Введение искусственных переменных позволяет получить начальное базисное решение как стартовую точку для симплекс-метода. При этом, в отличие от логических переменных, введенных на этапе приведения модели к канонической форме, искусственные переменные не относятся к постановке задачи и должны быть исключены (приравнены нулю) в конечном оптимальном решении.

Для построения процесса вычислений с искусственными переменными используется методика, по которой за каждую искусственную переменную W взимается штраф, пропорциональный ее зна-

чению. Существуют два способа, использующих эту идею. Первый — это *метод больших штрафов*, или *M-метод*.

Пример II. 24. В условиях пр. II.23 уравнение (II.59) не содержит логических переменных, поэтому в него вводится искусственная переменная:

$$\left. \begin{aligned} 70X_1 + 40X_2 + \tilde{X}_3 &= 168\,000; \\ 75X_1 + 100X_2 + \tilde{X}_4 &= 240\,000; \\ X_1 + X_2 + W &= 3000. \end{aligned} \right\} \quad (\text{II.62})$$

«Взимание» штрафа отражается в записи целевой функции: требуется максимизировать

$$C = 3X_1 + 2X_2 - MW, \quad (\text{II.63})$$

где M — достаточно большое положительное число, например 10^5 .

Три уравнения (II.62) содержат пять неизвестных, и начальное базисное решение должно включать $n - m = 5 - 3 = 2$ нулевые переменные. Если положить $X_1 = X_2 = 0$, то следует допустимое начальное решение: $X_1 = 168\,000$; $X_2 = 240\,000$; $W = 3000$. Теперь можно непосредственно применить симплекс-метод к системе (II.62).

Недостатком M -метода является возможность появления больших ошибок в вычислениях, вызванных округлением, так как метод оперирует с числами, значительно отличающимися по значению. Обойти эти трудности позволяет второй способ — *двухэтапный симплекс-метод*.

На первом его этапе находится начальное базисное решение. Переменные W , фигурирующие в нем, исключаются из задачи при помощи симплекс-метода. При этом формулируется искусственная целевая функция D , равная сумме искусственных переменных и подлежащая минимизации. Так, в условиях пр. II.24 задача первого этапа формулируется как минимизация $D = W$ при ограничениях (II.62).

Исходная целевая функция на первом этапе не рассматривается. Если минимальное значение искусственной целевой функции равно нулю, то все искусственные переменные обращаются в нуль и получается начальное базисное решение задачи ЛП. Если же минимальное значение D больше нуля, то по меньшей мере одна из искусственных переменных не равна нулю и задача не имеет допустимых решений.

На втором этапе оптимальное базисное решение, полученное из условия $D \rightarrow \min$, принимается в качестве начального решения исходной задачи, которое улучшается симплекс-методом в соответствии с исходной целевой функцией.

В процессе непосредственного применения симплекс-алгоритма к решению задач ЛП можно столкнуться с нетривиальными ситуациями, которые, как правило, не связаны с вычислительными аспектами метода, а вызваны постановкой задачи.

1. Все коэффициенты при вводимой в базис переменной меньше или равны нулю. В этом случае правило минимального отношения не дает возможности определить центральный элемент и, соответственно, не указывает, какая из переменных должна быть исключена из базиса. Вводимая переменная, таким образом, может неограни-

ченно возрастать при выполнении всех ограничений, следовательно, неограниченно возрастает и целевая функция. Такая задача не имеет конечного оптимального решения, что свидетельствует об ошибках при постановке задачи или пропуске одного из ограничений.

2. В процессе решения по крайней мере две строки имеют равные минимальные начения отношения b_i/a_{ij} . В этом случае исключаемую переменную можно выбрать произвольно. Однако последующее базисное решение будет обязательно иметь нулевую базисную переменную. Такое решение называется *вырожденным*; его появление вызвано присутствием в системе ограничений одного лишнего условия, которое, вообще говоря, не мешает определению искомого оптимального решения.

3. Если в индикаторной строке итоговой симплекс-таблицы имеется нулевой индикатор при небазисной переменной, то оптимальное решение неединственно. Значит, прямая (в общем случае гиперплоскость), представляющая изолинию целевой функции ($C = \text{const}$), параллельна прямой (гиперплоскости), проходящей через соседние вершины многогранника решений.

4. Не все искусственные переменные W обращаются в нуль в конечном оптимальном решении, это свидетельствует об отсутствии допустимых решений. При такой формулировке задачи ЛП все ограничения одновременно выполнены быть не могут.

Симплекс-алгоритм включен во многие пакеты прикладных программ для ЭВМ. Поэтому технолог обычно участвует только в процессах построения модели ЛП и подготовки ее к решению на ЭВМ, после чего получает машинную распечатку в виде оптимального решения задачи. На заключительном этапе моделирования основное внимание должно быть уделено анализу и интерпретации информации, содержащейся в полученном решении. На ее основании делаются практически важные выводы, в частности, о расходе ресурсов, о возможности улучшения целевой функции при единичном изменении объема ресурсов, о чувствительности оптимального решения к изменению параметров модели ЛП.

Пример II.25. На основании конечной симплекс-таблицы II.17 пр. II.22 можно сделать ряд выводов о полученном решении. Небазисные переменные \bar{X}_3 и \bar{X}_4 (логические) должны быть приравнены нулю. Остальные (базисные) переменные имеют значения, отраженные в столбце «правая часть» (его иногда называют также «решением»). Оптимальное решение задачи читается так: «объем производства плит типа Π_1 — 1800 шт., а плит типа Π_2 — 1050 шт.; максимальный доход (по индикаторной строке) $C^* = 7500$ руб.» Значение C^* можно получить и из уравнения целевой функции подстановкой объемов производства плит:

$$C^* = 3X_1 + 2X_2 = 3 \cdot 1800 + 2 \cdot 1050 = 7500 \text{ руб.}$$

Под ресурсом, фигурирующим в задачах ЛП, понимается максимальный предел его запаса для нужд производства. Для ограничений на ресурсы характерны только такие отношения левых и правых частей, как « \leq » (меньше или равно), поэтому при приведении к ОЗЛП в них вводятся только остаточные логические переменные. В табл. II.17 X_3 и X_4 равны нулю, следовательно, ресурсы ЦМН и ЦМЗ расходуются полностью. Увеличение запасов ЦМН и ЦМЗ приводит, как это видно из рис. II.22, а, к расширению области допустимых решений и, как следствие, к увеличению значения целевой функции C^* .

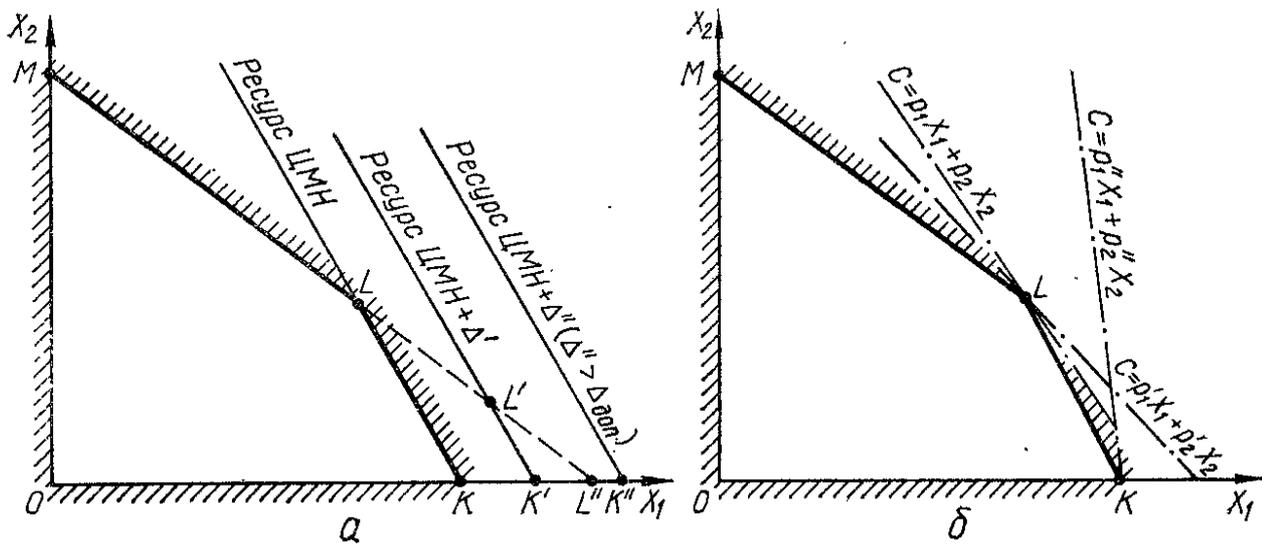


Рис. 11.22. Анализ чувствительности оптимального решения к увеличению ресурса (а) и к изменению целевой стоимостной функции (б) (изменения величин Δ' не приводят к изменению координат оптимальной точки)

Возможность оценить, в какой мере следует увеличивать запас каждого ресурса, дает введение понятия *ценности ресурса* как отношения максимального приращения оптимального значения целевой функции к максимально допустимому приросту объема ресурса. Ценность ресурса ЦМН определяется по абсолютным значениям коэффициентов индикаторной строки при переменных начального базиса в симплекс-таблице оптимального решения.

В табл. 11.17 $X_3 = 3/80$ руб. на 1 кг увеличения запаса ЦМН и $X_4 = 1/200$ руб. на 1 кг увеличения запаса ЦМЗ. Это означает, что увеличение запаса наполнителя (эквивалентное введению $X_3 > 0$) приводит к пропорциональному увеличению S с коэффициентом пропорциональности, равным $3/80$. Очевидно, что значительное увеличение запаса ресурса (сдвиг линии ограничения KI на рис. 11.22, а вправо) может привести к противоречию с другими ограничениями. Существует такое максимально возможное приращение запаса ресурса, при котором не произойдет изменения допустимости решения. Пусть, например, ресурс b_1 (количество наполнителя ЦМН на складе завода) увеличился на Δ . Новое значение запаса ЦМН приведет к изменению правых частей ограничений начальной (табл. 11.14) и всех последующих симплекс-таблиц решения задачи пр. 11.22. При этом изменения произойдут во всех правых частях таблиц. Измененное окончательное решение приведено в табл. 11.18, из которой видно, что правые части оптимального решения изменяются на величины, пропорциональные соответствующим коэффициентам при элементах столбца переменной X_3 (ср. табл. 11.17). Интервал возможного варьирования $\Delta_{\text{доп}}$ определяется из условия допустимости решения, т. е. из условия неотрицательности всех базисных переменных:

$$X_1 = 1800 + 1/40 \cdot \Delta_{\text{доп}} \geq 0, \quad (11.64)$$

$$X_2 = 1050 - 3/160 \cdot \Delta_{\text{доп}} \geq 0. \quad (11.65)$$

Таблица 11.18. Правые части симплекс-таблицы оптимального решения при изменении запаса ЦМН на $\Delta > 0$

Номер базисного решения	Базис	Правая часть
2	X_1 X_2	$1800 + 1/40 \cdot \Delta$ $1050 - 3/160 \cdot \Delta$
Индикаторная строка		$S - 7500 -$ $- 3/80 \cdot \Delta$

Эта система неравенств решается относительно $\Delta_{\text{доп}}$. Если $\Delta_{\text{доп}} > 0$, то из (II.65) следует, что $\Delta_{\text{доп}} \leq 1050 \cdot 160/3 = 56\,000$ кг. При $\Delta_{\text{доп}} < 0$ в качестве ограничения используется (II.64): $\Delta_{\text{доп}} \geq -1800 \cdot 40 = -72\,000$ кг. В общем случае внутри интервала $-72\,000 \leq \Delta_{\text{доп}} \leq 56\,000$ изменение Δ не вызывает изменения допустимости решения. Графическая интерпретация этого вывода дана на рис. II.22, а, откуда следует, что чрезмерное увеличение запаса на $\Delta'' > \Delta_{\text{доп}}$ может привести к изменению оптимальной точки.

Влияние на оптимальный план производства дохода от выпуска каждого изделия показано на рис. II.22, б, на котором изображены различные положения изолинии $C = 7500$ руб. в плоскости $\{X_1, X_2\}$. Из рис. II.22, б следует, что существуют некоторые крайние положения прямой, переход через которые должен изменить полученное оптимальное решение. Так как положение изолинии целевой функции $C = f(X_1, X_2)$ определяется значением и соотношением коэффициентов p_j , то можно установить пределы «безопасного» варьирования p_j , не изменяющего оптимального решения.

Если доход от производства плит Π_1 изменился на Δ_p , т. е. стал равным $3 + \Delta_p$, то уравнение целевой функции (II.31) примет вид

$$C = (3 + \Delta_p)X_1 + 2X_2. \quad (\text{II.66})$$

Индикаторная строка симплекс-таблицы не используется в качестве ведущей в итерациях, поэтому в оптимальном решении видоизменяется только она:

$$C - (7500 + 1800 \Delta_p) = 0X_1 - 0X_2 - (3/80 + 1/40 \cdot \Delta_p)X_3 - (1/200 - 1/100) X_4. \quad (\text{II.67})$$

Очевидно, что коэффициенты в уравнении целевой функции (II.66) изменятся по сравнению с найденным ранее оптимальным решением на величины, пропорциональные коэффициентам при X_1 в строке заключительной симплекс-таблицы (табл. II.17).

Неизменность оптимального решения достигается при соблюдении требования неотрицательности коэффициентов (при небазисных переменных) X_3 и X_4 . Следовательно, должны выполняться неравенства

$$\left. \begin{aligned} 3/80 + 1/40 \cdot \Delta_p &\geq 0; \\ 1/200 - 1/100 \cdot \Delta_p &\geq 0. \end{aligned} \right\} \quad (\text{II.68})$$

Решением системы (II.68) является интервал значений $-1,5 \leq \Delta_p \leq 0,5$.

Таким образом, если доход от реализации плиты Π_1 изменится в пределах от $3 - 0,5 = 2,5$ руб/шт до $3 + 1,5 = 4,5$ руб/шт, то это не повлечет за собой изменения оптимального решения задачи.

Вышеизложенные общие методические рекомендации по решению ОЗЛП могут быть расширены за счет учета специфических особенностей ряда технико-экономических задач.

II.7. Транспортная задача как линейная оптимизационная модель

Транспортная модель соответствует, в частности, следующей типовой задаче строительной индустрии: товарный бетон от m бетоносмесительных заводов определенной производительности нужно перевозить к n строительным объектам; для каждого объекта требуется определенное количество бетона; расстояние между заводами и объектами известно; необходимо определить, сколько бетона с каждого завода и на какой объект надо перевезти, чтобы грузо-перевозки (тоннокилометры и т. п.) были минимальными.

Вводятся обозначения: i — бетоносмесительные заводы ($i = \overline{1, m}$); j — строящиеся объекты ($j = \overline{1, n}$); a_i — количество бетона, производимое i -м заводом, т; b_j — потребность в бетоне j -го объекта, т; p_{ij} — расстояние от i -го завода до j -го объекта строительства, км; X_{ij} — количество бетона, перевозимое с i -го завода на j -й объект, т. Тогда задача ЛП транспортного типа в общем виде формулируется как минимизация грузооборота, т · км, транспорта

$$C = \sum_{i=1}^m \sum_{j=1}^n p_{ij} X_{ij} \quad (\text{II.69})$$

при условиях (ограничениях), что суммарный объем перевозок из i -го бетоносмесительного завода не может превышать количество произведенного бетона

$$\sum_{j=1}^n X_{ij} \leq a_i, \quad (\text{II.70})$$

а суммарные перевозки продукции на j -й объект строительства полностью удовлетворяют спрос на бетон

$$\sum_{i=1}^m X_{ij} \geq b_j, \quad (\text{II.71})$$

при этом $X_{ij} \geq 0$ для всех i и j .

Кроме того, очевидно, что суммарный объем производства на всех m заводах $\sum_{i=1}^m a_i$ не должен быть меньше суммарного количества потребленного бетона (или спроса на него) $\sum_{j=1}^n b_j$. Если объем производства равен спросу

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j, \quad (\text{II.72})$$

то транспортная модель называется *замкнутой* (или *сбалансированной*), если же производство превышает спрос

$$\sum_{i=1}^m a_i > \sum_{j=1}^n b_j, \quad (\text{II.73})$$

то модель называется *открытой* (или *несбалансированной*).

В реальных ситуациях добиться сбалансированности производства и потребления удастся не всегда. Однако транспортную модель можно сбалансировать введением фиктивного пункта производства или фиктивного пункта потребления. Замкнутость транспортной задачи облегчает ее численное решение.

Пример II.26. Мощности трех бетоносмесительных заводов 60, 10, 100 м³ бетона в час. Потребности четырех строительных объектов соответственно 70, 50, 30 и 20 м³/ч. Расстояние от пунктов производства до пунктов потребления приведено в табл. II.19.

Таблица 11.19. Расстояние между заводами и потребителями бетона

Номер бетоно- смесительного завода ($i =$ $= \overline{1, m}$)	Объекты строительства ($j = \overline{1, n}$)			
	1	2	3	4
1	2	3	11	7
2	1	0	6	1
3	5	8	15	9

Требуется сформулировать задачу оптимизации плана перевозок, стремясь к минимуму грузооборота.

Поскольку суммарный объем производства товарного бетона ($60 + 10 + 100$) м³/ч равен суммарному спросу ($70 + 50 + 30 + 20$) м³/ч, данная задача является замкнутой. Целевая функция

$$C = 2X_{11} + 3X_{12} + 11X_{13} + 7X_{14} + 1X_{21} + 0X_{22} + 6X_{23} + 1X_{24} + 5X_{31} + 8X_{32} + 15X_{33} + 9X_{34} \rightarrow \min \quad (\text{II.74})$$

при ограничениях

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j \text{ или } 60 + 10 + 100 = 70 + 50 + 30 + 20; \quad (\text{II.75})$$

$$\left. \begin{aligned} X_{11} + X_{12} + X_{13} + X_{14} &= 60; \\ X_{21} + X_{22} + X_{23} + X_{24} &= 10; \\ X_{31} + X_{32} + X_{33} + X_{34} &= 100; \\ X_{11} + X_{21} + X_{31} &= 70; \\ X_{12} + X_{22} + X_{32} &= 50; \\ X_{13} + X_{23} + X_{33} &= 30; \\ X_{14} + X_{24} + X_{34} &= 20; \\ X_{ij} &\geq 0. \end{aligned} \right\} \quad (\text{II.76})$$

Модель пр. II 26 содержит избыточное ограничение (II.75): если любые $m + n - 1$ ограничений задачи удовлетворяются, то автоматически удовлетворяется и оставшееся ограничение. Таким образом, любое из уравнений спроса и поставок можно опустить. Следовательно, транспортная модель содержит $m + n - 1$ независимых ограничений, вследствие чего в любое допустимое базисное решение входит также $m + n - 1$ базисных переменных.

Исходную модель и процесс численного решения транспортной задачи можно представить в компактном виде при помощи транспортной таблицы (табл. II.20). Она имеет вид матрицы, i -я строка которой соответствует i -му пункту производства, а j -й столбец — j -му пункту спроса; ij -й элемент матрицы представляется прямоугольным «окошком», которое называется ячейкой таблицы.

Расстояния между заводами и объектами p_{ij} даются в правом верхнем углу ij -ячейки. Левый верхний угол заполняется оценками оптимальности d_{ij} , смысл и назначение которых поясняются ниже. Наконец, в большом треугольнике в центре ячейки нахо-

Таблица 11.20. Общий вид транспортной таблицы

Заводы	Объекты строительства				Объем производства
	1	2	3	4	
1	d_{11} x_{11}	d_{12} x_{12}	d_{13} x_{13}	d_{14} x_{14}	a_1
2	d_{21} x_{21}	d_{22} x_{22}	d_{23} x_{23}	d_{24} x_{24}	a_2
3	d_{31} x_{31}	d_{32} x_{32}	d_{33} x_{33}	d_{34} x_{34}	a_3
Спрос	b_1	b_2	b_3	b_4	—

дится значение искомого объема перевозок — переменной x_{ij} (если оно нулевое, то его можно опустить).

Численное решение задачи, или определение оптимального плана перевозок, согласно сформулированному выше критерию оптимальности, принципиально ничем не отличается от решения задачи ЛП, и его можно производить симплекс-методом. Воспроизводятся все четыре этапа алгоритма, приведенного в п. 11.6. Однако такие особенности транспортной модели, как избыточность ограничений и равенство единице всех коэффициентов в уравнениях-ограничениях (11.76), позволяют упростить процедуры поиска включаемой в базис и выводимой из него переменных. При этом вместо проверки условия оптимальности по симплекс-критерию I (минимизация целевой функции) используется определение потенциалов, а построение замкнутого цикла заменяет процедуру применения симплекс-критерия II (условия допустимости). Для этого варианта симплекс-метода иногда используется название «метод потенциалов».

Этап 1. Начальное допустимое базисное решение, или опорный план перевозок транспортной задачи, которая представлена транспортной таблицей 11.21, соответствующей пр. 11.26, находится различными методами: «северо-западного угла», минимального элемента, аппроксимации Фогеля [12, 59].

Таблица 11.21. Исходная транспортная задача

Заводы	Объекты строительства				Объем производства, м ³ /ч
	1	2	3	4	
1	x_{11} 2	x_{12} 3	x_{13} 11	x_{14} 7	60
2	x_{21} 1	x_{22} 0	x_{23} 6	x_{24} 1	10
3	x_{31} 5	x_{32} 8	x_{33} 15	x_{34} 9	100
Спрос	70	50	30	20	—

Процедура поиска начального базисного решения, согласно правилу северо-западного угла, начинается с придания переменной

X_{11} максимально возможного, исходя из спроса и объема производства, значения. Остальные переменные первой строки или первого столбца (т. е. X_{1j} или X_{i1}) полагаются равными нулю. Если одновременно выполняются ограничения и на спрос и на производство, то можно вычеркивать или столбец, или строку (по желанию). В этом случае последующая включаемая в базис переменная имеет обязательно нулевое значение. Последующий этап предусматривает корректировку спроса и предложения во всех остальных строках и столбцах с учетом заданного максимального значения X_{11} . Далее максимально допустимое значение приписывается первому невычеркнутому элементу следующего столбца (строки). Заканчивается эта процедура, когда остается невычеркнутым только один столбец (или строка).

Пример II.27. Определить начальное базисное решение задачи, сформулированной в табл. II-21, методом северо-западного угла.

Переменной X_{11} присваивается значение, равное 60 (это максимально возможное значение, соответствующее объему производства на первом заводе — первой строке). Тем самым исчерпана возможность производства первого завода, поэтому вычеркиваются приравняемые нулю остальные элементы первой строки. На первую стройплощадку нужно доставить еще $10 \text{ м}^3/\text{ч}$ бетона — это весь оставшийся лимит на первый столбец.

Переменной X_{21} приписывается значение 10; вычеркивается оставшийся элемент X_{31} первого столбца; потребность 1-го объекта удовлетворена. Одновременно исчерпывается объем производства во второй строке. Следовательно, в этом случае удовлетворены ограничения как в строке, так и в столбце, поэтому следующая переменная ($X_{22} = 0$) автоматически вычеркивает строку 2. Остается лишь одна третья строка, и процедура закончена. Из табл. II.22 следует искомое начальное базисное решение:

$$X_{11} = 60, X_{21} = 10, X_{22} = 0, X_{32} = 50, X_{33} = 30, X_{34} = 20.$$

Таблица II.22. Начальное допустимое базисное решение транспортной задачи, полученное методом северо-западного угла

Заводы	Объекты строительства				Объем производства, $\text{м}^3/\text{ч}$	
	1	2	3	4		
1	60	2	3	11	7	60
2	10	1	0	6	1	10
3	5	50	8	15	9	100
Спрос	70	50	30	20		

Небазисные переменные (остальные X_{ij}) равны нулю. Число базисных переменных в табл. II.22 равно нужному числу: $m + n - 1 = 4 + 3 - 1 = 6$.

После нахождения начального допустимого базисного решения производится итерационная последовательность вычислений, включающих определение вводимой в базис и выводимой из него переменных и проверку условия оптимальности опорного плана текущей итерации.

Этап 2. Определение вводимой в базис переменной основывается на критерии оптимальности симплекс-метода, модифицированном применительно к решению транспортной задачи.

В *методе потенциалов* строке i и столбцу j транспортной таблицы (см. табл. II.20) ставятся в соответствие некоторые числа, называемые *потенциалами* u_i и v_j . Для каждой *базисной* переменной X_{ij} текущего решения сумма потенциалов должна быть равна соответствующему коэффициенту целевой функции

$$u_i + v_j = p_{ij}. \quad (\text{II.77})$$

Уравнения (II.77), число которых соответствует числу базисных переменных $(m + n - 1)$, составляют систему, в которой $(m + n)$ неизвестных. Значения потенциалов u_i и v_j определяются из этой системы, если придать одному из неизвестных (удобно наиболее часто встречающемуся) произвольное значение, например равное нулю. После этого для каждой *небазисной* X_{pq} переменной определяются *оценки* (критерии) *оптимальности*

$$d_{pq} = u_p + v_q - p_{pq}. \quad (\text{II.78})$$

В базис включается та переменная X_{pq} , которая имеет наибольшую положительную оценку оптимальности. Если среди оценок d_{pq} нет положительных, то данный опорный план — оптимальный.

Пример II.28. По данным пр. II.26 и II.27 отыскивается среди небазисных переменная, включаемая в базис. При этом уравнения, связанные с базисными переменными, имеют вид:

$$\begin{aligned} \text{для } X_{11} : u_1 + v_1 &= p_{11} = 2; \\ \text{—»— } X_{21} : u_2 + v_1 &= p_{21} = 1; \\ \text{—»— } X_{22} : u_2 + v_2 &= p_{22} = 0; \\ \text{—»— } X_{32} : u_3 + v_2 &= p_{32} = 8; \\ \text{—»— } X_{33} : u_3 + v_3 &= p_{33} = 15; \\ \text{—»— } X_{34} : u_3 + v_4 &= p_{34} = 9. \end{aligned}$$

Пусть $u_1 = 0$. Тогда значения потенциалов: $v_1 = 2$; $u_2 = -1$; $v_2 = 2$; $u_3 = 6$; $v_3 = 9$; $v_4 = 3$.

Оценки d_{qp} для небазисных переменных:

$$\begin{aligned} \text{для } X_{12} : d_{12} &= u_1 + v_2 - p_{12} = 0 + 2 - 3 = -1; \\ \text{— » — } X_{13} : d_{13} &= u_1 + v_3 - p_{13} = 0 + 9 - 11 = -2; \\ \text{— » — } X_{14} : d_{14} &= u_1 + v_4 - p_{14} = 0 + 3 - 7 = -4; \\ \text{— » — } X_{23} : d_{23} &= u_2 + v_3 - p_{23} = -1 + 9 - 6 = 2; \\ \text{— » — } X_{24} : d_{24} &= u_2 + v_4 - p_{24} = -1 + 3 - 1 = 1; \\ \text{— » — } X_{31} : d_{31} &= u_3 + v_1 - p_{31} = 6 + 2 - 5 = 3. \end{aligned}$$

Наибольшую положительную оценку имеет X_{31} , которую следует включить в базис.

Простая структура уравнений $u_i + v_j = p_{ij}$ позволяет, не записывая систему уравнений, находить потенциалы непосредственно по транспортной таблице. Вначале в левых верхних углах ячеек, соответствующих каждой *базисной* переменной X_{ij} на текущей итерации, проставляются символы \emptyset . Выбирается любая величина — u_i или v_j . Ей присваивается произвольное значение, про-

ставляемое у строки или столбца, которым принадлежит выбранное u_i или v_j . Строка или столбец анализируется с целью выявления базисных переменных данной итерации. Если выбрана строка i и базисные переменные в ней X_{il} , X_{ir} и X_{is} , то уравнение (II.77) позволяет найти значения потенциалов $v_l = p_{il} - u_i$, $v_r = p_{ir} - u_i$ и $v_s = p_{is} - u_i$, соответствующие столбцам, на пересечении которых с i -й строкой находятся базисные переменные. Полученные значения v_j используются по аналогии для вычисления оставшихся потенциалов.

После этого можно вычислить по (II.78) оценки оптимальности d_{pq} для всех небазисных переменных X_{pq} и проставить их в соответствующих левых углах ячеек табл. II.23.

Таблица II.23. Потенциалы u_i и v_j и оценки оптимальности первой операции, полученные по упрощенной схеме

Заводы	Объекты строительства				u_i	Объем производства, м ³ /ч
	1	2	3	4		
1	Φ 60	2 -2	3 -3	11 -5	7 -7	60
2	Φ 10	1 Φ 0	0 1	6 0	1 -8	10
3	4 5	Φ 50	8 Φ 30	15 Φ 20	9 0	100
v_j	9	8	15	9		—
Спрос	70	50	30	20		

Пример II.29. Требуется найти значения потенциалов u_i и v_j первой итерации решения задачи (II.74) — (II.76), пользуясь упрощенной схемой. Согласно табл. II.22, выбирается $u_3 = 0$. Справа от 3-й строки можно проставить значение u_3 . Строка (столбец), соответствующая u_3 , просматривается для выявления базисных переменных. В строке 3 базису принадлежат X_{32} , X_{33} , X_{34} . С помощью уравнения (II.77) находятся значения потенциалов: $v_2 = p_{32} - u_3 = 8$, $v_3 = p_{33} - u_3 = 15 - 0 = 15$, $v_4 = p_{34} - u_3 = 9 - 0 = 9$. Вычисление остальных потенциалов проводится по вышеизложенной схеме ($u_2 = -8$, $v_1 = -7$, $u_1 = -7$) и заносится в табл. II.23. Теперь можно вычислить оценки d_{pq} для небазисных X_{pq} и занести их в соответствующие ячейки. Как видно из табл. II.23, значения потенциалов u_i и v_j , а также оценки d_{pq} отличаются от значений, полученных при непосредственном решении системы уравнений (II.77), (II.78). Это вызвано выбором в качестве свободного потенциала переменной u_3 вместо u_1 . Однако по критерию оптимальности наибольшее положительное значение d_{pq} по-прежнему имеет небазисная переменная X_{31} , которая и включается в базис.

Этап 3. Согласно условию допустимости симплекс-метода (см. п. II.6), выводимая из базиса переменная определяется как имеющая минимальное отношение правых частей к коэффициентам в j -м столбце транспортной таблицы, соответствующей вводимой в базис переменной X_{ij} . Так как коэффициенты при переменных в ограничениях транспортной модели всегда численно равны нулю или единице, то значения базисных переменных автоматически определяют отношения правых частей к коэффициентам при X_{ij} .

Поэтому на данном этапе вместо симплекс-критерия II используется искусственное построение: в транспортной таблице размечается *замкнутый цикл*, соответствующий вводимой в базис на данном шаге переменной. Цикл представляет собой единственную ломаную, состоящую из горизонтальных и вертикальных прямых, узлы пересечения которых приходятся на ячейки всех переменных текущего базиса и вводимой в базис переменной.

Пример II.30. Построение цикла для текущего решения транспортной табл. II.24 начинается с вводимой в базис переменной. На данной итерации базисные переменные — $X_{11}, X_{21}, X_{22}, X_{32}, X_{33}, X_{34}$.

Таблица II.24. Цикл для включаемой на первой итерации переменной X_{31}

Заводы	Объекты строительства				Объем производства, м ³ /ч
	1	2	3	4	
1	∅ 60	∅ 2	∅ 3	∅ 11	60
2	∅ 10	∅ 0	∅ 6	∅ 1	10
3	∅ 5	∅ 50	∅ 8	∅ 15	100
Спрос	70	50	30	20	—

Единственный возможный для текущего решения замкнутый цикл — $X_{31} \rightarrow X_{21} \rightarrow X_{22} \rightarrow X_{32} \rightarrow X_{31}$ — должен заканчиваться также на переменной X_{31} .

Основная идея выделения из переменных, входящих в цикл, исключаемой из базиса, состоит в перераспределении ресурсов в пределах цикла текущей итерации. Для этого вводимой в базис переменной дается некоторое приращение Δ , причем другие базисные переменные, принадлежащие узловым точкам цикла, также должны претерпеть изменение: часть базисных переменных цикла увеличивается на Δ , а часть уменьшается на ту же величину. Естественно выбрать для исключения из базиса ту переменную, которая быстрее всех других может обратиться в нуль.

Пример II.31. Из переменных цикла (табл. II.24) при $\Delta = 10$ м³/ч становится равной нулю базисная переменная X_{21} , следовательно, она исключается из базиса.

Вычисление потенциалов по упрощенной схеме для нового базисного решения дает результаты, представленные в транспортной таблице (табл. II.25).

Так как среди оценок оптимальности d_{pq} есть положительные, то базисное решение после первой итерации не оптимально. Наибольшее положительное значение d_{pq} имеет переменная X_{12} , которую включают в базис. На основании построения нового цикла $X_{12} \rightarrow X_{32} \rightarrow X_{31} \rightarrow X_{11} \rightarrow X_{12}$ из базиса исключается переменная X_{32} (наименьшая по модулю из уменьшаемых в результате включения X_{12} в базис).

Дальнейший ход итерационного процесса иллюстрируется данными табл. II.26 — II.28.

На основании табл. II.26 в базис включается переменная X_{23} , но выводится из него X_{22} . Те же преобразования выполняются с табл. II.27

Таблица II.25. Вторая итерация

Заводы	Объекты строительства				u_i	Объем производства, $m^3/ч$
	1	2	3	4		
1	Φ 2 2 3 1 11 -1 7 -3	Φ 2 2 3 1 11 -1 7 -3	Φ 2 2 3 1 11 -1 7 -3	Φ 2 2 3 1 11 -1 7 -3	-3	60
2	-4 1 Φ 0 1 6 0 1 -8	-4 1 Φ 0 1 6 0 1 -8	-4 1 Φ 0 1 6 0 1 -8	-4 1 Φ 0 1 6 0 1 -8	-8	10
3	Φ 10 5 Φ 8 Φ 30 15 Φ 9 0	Φ 10 5 Φ 8 Φ 30 15 Φ 9 0	Φ 10 5 Φ 8 Φ 30 15 Φ 9 0	Φ 10 5 Φ 8 Φ 30 15 Φ 9 0	0	100
v_j	5	8	15	9		—
Спрос	70	50	30	20		—

Таблица II.26. Третья итерация

Заводы	Объекты строительства				u_i	Объем производства, $m^3/ч$
	1	2	3	4		
1	Φ 20 2 Φ 3 1 11 -1 7 -3	Φ 20 2 Φ 3 1 11 -1 7 -3	Φ 20 2 Φ 3 1 11 -1 7 -3	Φ 20 2 Φ 3 1 11 -1 7 -3	-3	60
2	-2 1 Φ 0 3 6 2 1 -6	-2 1 Φ 0 3 6 2 1 -6	-2 1 Φ 0 3 6 2 1 -6	-2 1 Φ 0 3 6 2 1 -6	-6	10
3	Φ 50 5 -2 8 Φ 30 15 Φ 20 9 0	Φ 50 5 -2 8 Φ 30 15 Φ 20 9 0	Φ 50 5 -2 8 Φ 30 15 Φ 20 9 0	Φ 50 5 -2 8 Φ 30 15 Φ 20 9 0	0	100
v_j	5	6	15	9		—
Спрос	70	50	30	20		—

Таблица II.27. Четвертая итерация

Заводы	Объекты строительства				u_i	Объем производства, $m^3/ч$
	1	2	3	4		
1	Φ 10 2 Φ 3 1 11 -1 7 -3	Φ 10 2 Φ 3 1 11 -1 7 -3	Φ 10 2 Φ 3 1 11 -1 7 -3	Φ 10 2 Φ 3 1 11 -1 7 -3	-3	60
2	-5 1 -3 0 Φ 6 -1 1 -9	-5 1 -3 0 Φ 6 -1 1 -9	-5 1 -3 0 Φ 6 -1 1 -9	-5 1 -3 0 Φ 6 -1 1 -9	-9	10
3	Φ 60 5 -2 8 Φ 20 15 Φ 20 9 0	Φ 60 5 -2 8 Φ 20 15 Φ 20 9 0	Φ 60 5 -2 8 Φ 20 15 Φ 20 9 0	Φ 60 5 -2 8 Φ 20 15 Φ 20 9 0	0	100
v_j	5	6	15	9		—
Спрос	70	50	30	20		—

(X_{13} вводится, X_{11} исключается). Табл. II.28 является окончательной: все оценки оптимальности d_{pq} неположительны, следовательно, полученное решение является оптимальным и достигнуто за четыре шага.

В настоящее время разработаны различные пакеты прикладных программ для решения задач линейного программирования с помощью различных методических приемов, направленных как на ускорение получения конечного результата, так и на повышение его точности. Инженер-технолог, как правило, должен пользоваться готовыми пакетами программ, сосредоточив свое внимание на содержательной постановке технико-экономических, техноло-

Таблица 11.28. Оптимальное решение транспортной задачи

Заводы	Объекты строительства				u_i	Объем производства, $m^3/ч$
	1	2	3	4		
1	-1 2	Φ 50 3	Φ 10 11	-2 7	-4	60
2	-5 1	-2 0	Φ 10 6	-1 1	-9	10
3	Φ 50 5	-1 8	Φ 30 15	Φ 20 9	0	100
v_j	5	7	15	9		—
Спрос	70	50	30	20		

ических и материаловедческих задач, решаемых как ОЗЛП. В [38] достаточно подробно освещены вопросы прохождения программ на одной из самых распространенных систем ЭВМ — ЕС ЭВМ.

Глава III. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ ТЕОРИИ ЭКСПЕРИМЕНТА

III.1. Экспериментально-статистические модели и информационная таблица для их построения

Математическая модель, заменяющая (см. «Введение») объективно присущее исследуемой технологической системе (веществу, процессу, аппарату, технологической линии и т. п.) уравнение состояния (В. 8), может быть получена как абстрактно-знаковое описание функционирования объекта, т. е. как отражение взаимосвязи между воздействующими на систему факторами X_i и ее выходами Y_j , являющимися откликами на эти воздействия. В такой модели на исходном этапе ее построения, как правило, на заданное время τ неизвестны все элементы: а) ни список основных действующих факторов X_i ($i = 1, k$) и границ их изменения $X_{i, \min} \leq X_i \leq X_{i, \max}$; б) ни список основных показателей качества функционирования системы Y_j и возможных границ их изменения $Y_{j, \min} \leq Y_j \leq Y_{j, \max}$; в) ни законы распределения случайных величин ξ ; г) ни вид функции F_m ; д) ни числовые значения параметров Θ_i — коэффициентов в этой функции, т. е.

$$F_m(Y_j, \Theta_i, X_i, \tau, \xi) = 0. \quad (III.1)$$

Однако две первых неопределенности инженер-технолог может разрешать за счет анализа накопленной к началу исследований так называемой *априорной* (лат. *a priori* — изначально, до данного опыта; *a posteriori* — из последующего, после данного опыта) информации, источники которой и личностное знание, и литература, и «информационный блок» ЭВМ и т. п. Конечно, риск ошибки моделирования [6] за счет составления неполного списка X_i и Y_j (см., в частности, вторую, л-теорему подобия [29, 52]) остается, но он тем меньше, чем больше интеллектуальные затраты инженера

на этапах, предшествующих основному эксперименту и моделированию на ЭВМ.

В отношении третьей неопределенности — закона распределения случайных величин ξ — приходится, как правило, в большинстве инженерных задач допускать, что это распределение соответствует нормальному (см. прил. I.4).

В отношении вида функции F_M при заданной формулировке цели моделирования системы как «отражения взаимосвязи между ее факторами X_i и выходами Y_j в области Ω » не может быть, вообще говоря, каких-либо серьезных ограничений, связанных с внутренними явлениями в данной системе. Поэтому, во-первых, можно одновременно рассматривать не одну, а несколько конкурирующих функций $F_{M.1}, \dots, F_{M.i}, \dots$ — так называемый «веер моделей» [45, 46], во-вторых, в качестве одного из главных критериев выбора функции $F_{M.i}$ как модели объекта принять *точность описания* связи X_i с Y_j в изученном диапазоне Ω_x их изменения (см. гл. I), в-третьих, при прочих равных условиях ориентироваться на принцип «простоты» (например, простоты вида функции F_M — «прямая лучше кривой»; простоты построения — «меньше объем физического эксперимента и время расчетов на ЭВМ»; простоты интерпретации — «ясный инженерный смысл параметров Θ_i » и т. п.). Все это не отрицает необходимости тщательного анализа априорной информации для включения в веер моделей тех или иных концептуальных математических описаний (см. «Введение»), но подчеркивает допустимость их отсутствия в этом веере.

Для большинства реальных материаловедческих, рецептурно-технологических и технико-экономических задач целесообразно принять достаточно полно отвечающие вышеуказанным допущениям *полиномиальные модели* порядка m , которые для k факторов записываются так:

$$Y = A_0 + \sum_{i=1}^k A_i X_i + \sum_{i < j} A_{ij} X_i X_j + \sum_{i=1}^k A_{ii} X_i^2 + \dots + \sum_{i=1}^k A_{iii} X_i^3 + \dots \quad (\text{III.2})$$

Параметры модели Θ (коэффициенты полинома $A_0, A_i, A_{ij}, A_{ii}, \dots, A_{iii}, \dots$) эквивалентны частным производным кратного ряда Тейлора (см. прил. I.3) от явной относительно Y функции f_M (III.1) в окрестностях точки разложения с координатами g_1, \dots, g_k (время τ , как правило, включается в k факторов и принимает значение g_i):

$$\left. \begin{aligned} A_0 &= f_M(g_1, g_2, \dots, g_k) = \text{const}; & A_i &= \frac{\partial f_M}{\partial X_i}; \\ A_{ij} &= \frac{1}{2!} \frac{\partial^2 f_M}{\partial X_i \partial X_j}; & A_{ii} &= \frac{1}{2!} \frac{\partial^2 f_M}{\partial X_i^2}; \dots; \\ A_{iii} &= \frac{1}{3!} \frac{\partial^3 f_M}{\partial X_i^3}, \dots \end{aligned} \right\} \quad (\text{III.3})$$

Числовые оценки истинных коэффициентов (III.3) соответственно $a_0, a_j, a_{ij}, a_{ii}, \dots$ определяются как статистические величины по экспериментальным данным. В результате для расчетного значения выхода \hat{Y} будет получена *полиномиальная экспериментально-статистическая модель* (ЭС-модель)

$$\hat{Y} = a_0 + \sum_{i=1}^k a_i X_i + \sum_{i < j} a_{ij} X_i X_j + \sum_{i=1}^k a_{ii} X_i^2 + \dots + \\ + \sum_{i=1}^k a_{iii} X_i^3 + \dots + \varepsilon. \quad (\text{III.4})$$

Входящее в (III.4) слагаемое ε — случайная величина, которая отражает как влияние на поведение системы случайных факторов ξ , так и, в частности, последствия расчета оценок коэффициентов модели по экспериментальным данным (см. п. III, 7, III.8).

Полином (III.4) не является единственной или универсальной формой ЭС-модели; в то же время многие другие формы при использовании соответствующих преобразований (см. п. III.9) сводятся к нему. Например, многофакторная показательная функция (III.5), являющаяся базовой при моделировании методами теории подобия и анализа размерностей [52], после логарифмирования критериев подобия π_i и изменения обозначений ($Y = \ln \pi_0, a_0 = \ln K_0; X_i = \ln \pi_i$) оказывается эквивалентной линейной части полинома (III.4):

$$\pi_0 = K_0 \pi_1^{a_1} \pi_2^{a_2} \dots \pi_i^{a_i} \dots \pi_k^{a_k}; \quad (\text{III.5})$$

$$\left. \begin{aligned} \ln \pi_0 = \ln K_0 + a_1 \ln \pi_1 + \dots + a_k \ln \pi_k = \ln K_0 + \sum_{i=1}^k a_i \ln \pi_i; \\ \hat{Y} = a_0 + \sum_{i=1}^k a_i X_i. \end{aligned} \right\} (\text{III.6})$$

Параметры ЭС-моделей оцениваются по информации о *результатах экспериментов* с исследуемой системой (активное изменение значений X_i на входе и фиксация соответствующих значений Y_j) или о *результатах наблюдений* за ее поведением в режиме нормальной эксплуатации (пассивная фиксация одновременно значений X_i на входе и Y_j на выходе). Сбор информации при активном воздействии на объект (эксперимент), как правило, в задачах материаловедения и технологии предпочтительнее (по расходу ресурсов, по достоверности информации, по точности моделирования и т. п. ([4, 6, 52]), чем наблюдение, хотя не исключены ситуации, когда оно — единственный источник сведений о поведении объекта (в частности, при контроле качества сырья и продукции на производстве, эколого-социологических обследованиях и т. п.).

Вне зависимости от метода сбора информации о поведении системы результаты образуют *информационную таблицу* (табл. III.1), которая включает следующие элементы.

1. Сведения о количественных (или хотя бы качественных) уровнях всех исследуемых факторов $X_1, \dots, X_i, \dots, X_k$ в каждом u -м опыте для N разных опытов (*разными* являются те опыты, которые отличаются уровнями хотя бы одного из X_i факторов). Совокупность уровней таких опытов образует *спектр плана* (ГОСТ 24026—80). Если в нескольких опытах уровни всех факторов совпадают, то их следует считать *параллельными* (ГОСТ 24026—80). Если проводится не эксперимент, а наблюдение, то, как правило, фактор X_1 фиксирует время τ или его технологический эквивалент — номер цикла и т. п.

2. Сведения о числе *параллельных испытаний* $m_{j,u}$ каждого из показателей системы Y_j в каждом u -м опыте. В общем случае значение $m_{j,u}$ может быть переменным, но целесообразно с метрологической и вычислительных точек зрения стремиться к такой организации исследований, чтобы для данного j -го показателя сохранялось $m_u = \text{const}$ (например, $m_{1,1} = \dots = m_{1,u} = \dots = m_{1,N}$).

3. Совокупность элементов 1 и 2, определяющих число, условия и порядок реализации опытов, образует план эксперимента (ГОСТ 24026 — 80). Без потери общности это понятие можно отнести и к пассивным наблюдениям, поскольку значения X_i фиксируются в определенный момент времени τ , в определенных точках объекта и т. п.

4. Апостериорные сведения о количественных результатах эксперимента (наблюдения) $y_{j,uv}$ для каждого показателя функционирования системы в каждом u -м опыте и в каждом v -м параллельном испытании.

5. Информация, собранная в элементе 4 по каждому u -му опыту, «сворачивается» в числовые статистические характеристики результатов, из которых абсолютно необходимой является оценка среднего из $m_{j,u}$ определений $y_{j,uv}$, рассчитываемая как

$$\bar{y}_{j,u} = \left(\sum_{v=1}^{m_{j,u}} y_{j,uv} \right) / m_{j,u}. \quad (\text{III.7})$$

Кроме нее, в зависимости от целей исследования, объема информации $m_{j,u}$ и ее свойств, могут быть рассчитаны и другие статистические характеристики результатов (см. п. III.3 — III.5).

Уровни k факторов для всей совокупности опытов (часть из них может содержать повторяющиеся строки из спектра плана) образуют прямоугольную матрицу размером $(N_m \times k)$, называемую матрицей плана эксперимента \mathbf{X} . Если $m = \text{const}$, то она совпадает с повторенной m раз матрицей спектра плана, состоящей только из N разных строк табл. III.1 и имеющей размер $(N \times k)^*$.

Полученные в N опытах средние значения выходного j -го показателя системы образуют вектор \mathbf{y} размером $(N \times 1)$, называ-

* Далее везде, если нет специальных указаний, матрица плана эксперимента тождественна матрице спектра плана.

Таблица III.1. Информационная таблица о поведении системы

Характеристики системы						
входные (план эксперимента)			выходные (результат эксперимента)			
Номер опыта	Уровни факторов $X_1, \dots, X_i, \dots, X_k$	Число парал- лельных испытаний для оценки \bar{y}_1 и \bar{y}_2	Значение Y_1 в каждом ν -м параллельном испытании	Среднее значение \bar{y}_1	Значения Y_2 в каждом ν -м испытании	Среднее значение \bar{y}_2
1	$X_{11} \dots X_{i1} \dots X_{k1}$	m_{11}, m_{21}	$y_{1.11}, \dots, y_{1.1\nu}, \dots, y_{1.1m_{11}}$	$\bar{y}_{1.1}$	$y_{2.11}, \dots, y_{2.1m_{21}}$	$\bar{y}_{2.1}$

i	$X_{1i} \dots X_{ii} \dots X_{ki}$	m_{1i}, m_{2i}	$y_{1.i1}, \dots, y_{1.i\nu}, \dots, y_{1.im_{1i}}$	$\bar{y}_{1.i}$	$y_{2.i1}, \dots, y_{2.im_{2i}}$	$\bar{y}_{2.i}$

N	$X_{1N} \dots X_{iN} \dots X_{kN}$	m_{1N}, m_{2N}	$y_{1.N1}, \dots, y_{1.N\nu}, \dots, y_{1.Nm_{1N}}$	$\bar{y}_{1.N}$	$y_{2.N1}, \dots, y_{2.Nm_{2N}}$	$\bar{y}_{2.N}$

Таблица III.2. Фрагмент информационной таблицы об изменении пре под влиянием 9 факторов

Номер опыта	Номер точки плана	План эксперимента									Число образцов в опыте $m_1=m_2$
		Факторы									
		рецептурные					технологические				
		$X_1=$ =П/Ц	$X_2=$ =В/Ц	$X_3=$ =С _{сп} , %	$X_4=$ =С _э , %	$X_5=$ =С _{отх} , %	$X_6=$ =τ ₀ , ч	$X_7=$ =V _T , °C/ч	$X_8=$ =T, °C	$X_9=$ =τ _T , ч	
1	47	2	0,40	0,4	0,6	0,15	4	20	65	7	6
...
и	52	2,5	0,40	0,4	0,6	0,15	6	20	65	7	6
...
N=55	17	2	0,35	0	0	0	2	10	80	4	6

емый вектором результатов. Матрица X и вектор y являются информацией, необходимой и достаточной для оценки неизвестных параметров Θ ЭС-моделей (см. п. III.7 — III.9). Оценки случайной величины ϵ , входящей в такие модели, а также других статистических характеристик моделей определяются при регрессионном анализе (см. п. III.8), который основан на информации, содержащейся в матрице X и в совокупности результатов параллельных испытаний, как в каждом i -м опыте — векторы размером $(1 \times m_i)$, так и во всех N опытах.

Пример III.1. При экспериментальном изучении влияния на предел прочности и другие свойства мелкозернистого бетона рецептурно-технологических факторов получена [22] информационная таблица с 55 строками, фрагмент которой показан в табл. III.2. В исследование включено 5 рецептурных факторов из подсистемы X_p и 4 технологических фактора из подсистемы $X_{п.п}$ (см. «Введение»):

X_1 — песчано—цементное отношение $2 \leq \text{П/Ц} \leq 3$;
 X_2 — водоцементное отношение $0,35 \leq \text{В/Ц} \leq 0,45$;
 X_3 — концентрация, % сухого вещества от массы цемента, суперпластификатора (СП) $0 \leq C_{сп} \leq 0,8$;

X_4 — то же, электролита $0 \leq C_э \leq 1,2$, предназначенного для ускорения твердения бетона и пассивации коррозии арматуры;

X_5 — то же, промышленного отхода $0 \leq C_{отх} \leq 0,3$, содержащего поверхностно-активное вещество (ПАВ), которое вводится как добавка, снижающая расход СП;

X_6 — время выдержки $2 \leq \tau_0 \leq 6$ свежесформованного бетона до начала тепловлажностной обработки (ТВО) в камерах;

X_7 — скорость подъема температуры в камерах ТВО:

$$20 \leq V_T = \frac{\partial T}{\partial \tau} \leq 30 \text{ (К/ч} = \text{°C/ч)};$$

X_8 — температура изотермической выдержки в камере твердения $323 \leq T \leq 353 \text{ К}$ или $50 \leq T \leq 80 \text{ °C}$;

дела прочности мелкозернистого бетона на сжатие R , МПа, после ТВО

Результаты эксперимента							
Предел прочности бетона R , МПа							
через 1 сут				через 28 сут			
каждого v -ого образца			среднее \bar{R}_1	каждого v -ого образца			среднее \bar{R}_{28}
20,3	22,7	24,0	23,38	43,2	45,4	49,9	48,58
24,0	24,5	24,8		50,2	50,7	52,1	
17,9	19,3	20,8	20,88	42,7	44,1	45,7	46,55
21,9	21,9	23,5		47,5	49,4	49,4	
22,7	23,2	23,7	24,43	32,3	32,6	33,9	33,68
24,8	26,1	26,1		33,9	34,4	35,0	

X_9 — время $4 \leq \tau_T \leq 10$ ч изотермической выдержки бетона при температуре T .

Эти 9 факторов образовали матрицу X размером (55×9) , из которой 3 строки показаны в табл. III.2; идущим по порядку номерам опытов соответствуют выбранные из оптимального плана эксперимента случайным образом точки — см. п. III.10. Первая (точка № 47) и вторая (точка № 52) строки в этой таблице отличаются уровнями факторов X_1 и X_6 (во второй оба увеличены), третья строка (точка № 17) для всех факторов отлична от второй; по ней, в частности, требуется исследовать бездобавочный (эталонный) бетон. Для определения в разные сроки после ТВО предела прочности бетона на сжатие в каждом опыте испытано по 6 образцов ($m_1 = m_2 = 6$).

В правой части табл. III.2 приведены результаты эксперимента — шесть частных значений R_v для каждого из образцов. Они записаны упорядоченно от меньшего к большему, образуя *ранжированный* (от нем. *Rang* — разряд) ряд. Частные значения позволили по (III.7) найти средний результат, например в опыте № 1:

$$\bar{R}_{1.1} = (20,3 + 22,7 + \dots + 24,8) / 6 = 140,3/6 = 23,38 \text{ МПа.}$$

Частные результаты «рассеяны» по отношению к среднему (от меньшего результата к большему $R_{1.1v}$, например, увеличивается на 4,8 МПа). Различные оценки этого рассеяния рассчитаны в п. III.3.

Результаты испытаний образовали два вектора $y_1 = \bar{R}_1$ и $y_2 = \bar{R}_2$ размером 55×1 .

III.2. Экспериментально-статистические модели в нормализованном факторном пространстве

Введение при моделировании нормализованных по (В.2) — (В.4) переменных переводит k -мерное факторное пространство из прямоугольной гиперпризмы (см. рис. В. 2) в *гиперкуб*, причем через его центр $\{0, 0, \dots, 0\}$ проходят новые оси $x_i (i = \overline{1, k})$, вершины имеют координаты $\{\pm 1, \pm 1, \dots, \pm 1\}$, а середины граней — координаты, содержащие только комбинации ± 1 и 0, что можно наблюдать при $k = 2$ (рис. III.1, а).

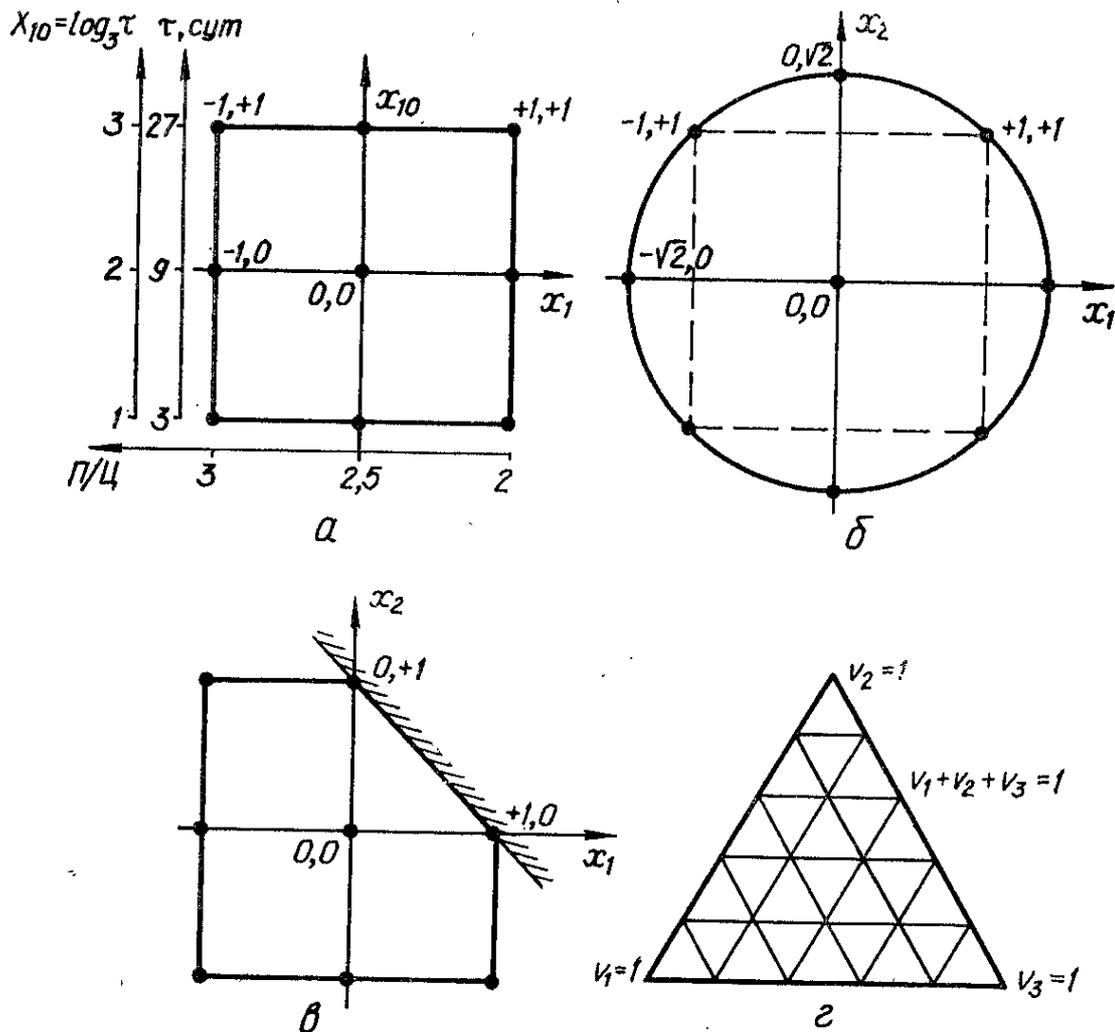


Рис. III.1. Факторные пространства:

a — квадрат после нормализации факторов; *b* — сфера, описанная вокруг квадрата; *c* — усеченный квадрат; *d* — симплекс

Пример III.2. Введение для рецептурно-технологических факторов, рассмотренных в пр. III.1, нормализованных переменных

$$\left. \begin{aligned}
 x_1 &= (2,5 - \text{П/Ц})/0,5; & x_2 &= (\text{В/Ц} - 0,4)/0,05; \\
 x_3 &= (C_{\text{сп}} - 0,4)/0,4; & x_4 &= (C_3 - 0,6)/0,6; \\
 x_5 &= (C_{\text{отх}} - 0,15)/0,15; & x_6 &= (\tau_0 - 4)/2; \\
 x_7 &= (V_T - 20)/10; & x_8 &= (T - 65)/15; \\
 x_9 &= (\tau_T - 7)/3
 \end{aligned} \right\} \quad (\text{III.8})$$

позволяет преобразовать левую часть информационной таблицы (табл. III.2) и записать ее в безразмерных переменных (табл. III.3). Следует обратить внимание на то, что для удобства интерпретации результатов моделирования значение $x_1 = +1$ совмещено не с верхним, а с нижним уровнем П/Ц (т.е. с верхним уровнем расхода цемента).

Не обязательно, чтобы нормализация производилась только при условии равенства интервалов варьирования $X_{i, \max} - X_{0, i} = X_{0, i} - X_{i, \min} = \Delta X$. Исходя из априорной информации, целесообразно нормализовать предварительно преобразованные факторы. Например, известно, что при нормальном твердении изменение структурно-механических свойств (в частности, предела прочности бетона [15]) пропорционально логарифму «возраста» τ ,

Таблица III.3. Фрагмент плана эксперимента в натуральных и нормализованных переменных

Номер опыта	Номер точки плана	Натуральные переменные									Нормализованные переменные								
		$X_1=П/Ц$	$X_2=В/Ц$	$X_3=C_{сп}, \%$	$X_4=C_{э}, \%$	$X_5=C_{отх}, \%$	$X_6=\tau_0, ч$	$X_7=VT, ^\circ C/ч$	$X_8=T, ^\circ C$	$X_9=\tau T, ч$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	47	2	0,40	0,4	0,6	0,15	4	20	65	7	+1	0	0	0	0	0	0	0	0
...
и	52	2,5	0,40	0,4	0,6	0,15	6	20	65	7	0	0	0	0	+1	0	0	0	
...
N=55	17	2	0,35	0	0	0	2	10	80	4	+1	-1	-1	-1	-1	-1	-1	+1	-1

сут, поэтому сначала проводится преобразование $\log_a \tau + X_\tau$, а затем по (В.2) — (В.4) выполняется нормализация

$$x_\tau = \frac{\log_a \tau - 0,5 (\log_a \tau_{\max} + \log_a \tau_{\min})}{0,5 (\log_a \tau_{\max} - \log_a \tau_{\min})} = \frac{X_\tau - X_{0,\tau}}{\Delta X_\tau}. \quad (III.9)$$

По (III.9) при $a = 3$ преобразована ось ординат на рис. III.1, а. Возврат к натуральным переменным τ , сут, в этом случае выполняется по вытекающему из (III.9) соотношению

$$\tau = a^{0,5[(1+x_\tau)\log_a \tau_{\max} + (1-x_\tau)\log_a \tau_{\min}]}. \quad (III.10)$$

Записанная в нормализованных переменных матрица плана x (табл. III.3) имеет *безотносительный к содержательной постановке задачи* характер. Именно такие матрицы планов эксперимента x размером $(N \times k)$ строятся и оптимизируются по некоторым критериям, важным с точки зрения и экспериментаторов, и пользователей моделей (см. п. III.10). Матрицы планов в безразмерных переменных x имеют универсальный характер, поэтому сводятся в каталоги [57] и хранятся в памяти ЭВМ.

Не следует представлять, что в инженерных задачах необходимо во всех ситуациях сводить факторное пространство к гиперкубу. В соответствии с целями исследования и возможностями их достижения для конкретной системы *факторное пространство* (даже в нормализованных переменных) может принимать самые *различные формы*, ограниченные «кусками» линейных и нелинейных (выпуклых или вогнутых — см. прил. I.1) поверхностей [6, 8]. Так, сферическое (для плоскости — круговое) ограничение, описанное радиусом \sqrt{k} вокруг k -мерного куба (рис. III.1, б), целесообразно в ряде задач, связанных с поиском области оптимума, расположенной в любом равновероятном направлении от центра эксперимента (см. п. III.10). На рис. III.1, в квадрат усечен добавочной функцией

$x_2 = f(x_1) = 1 - x_1$; такое ограничение наиболее часто наблюдается в системах с предельным физическим состоянием, возникающим при одновременном увеличении нескольких факторов (появляется взрывоопасность или расслоение смеси и т. п. — см. пр. III.44 и III.47). На рис. III.1, *г* показано ограниченное треугольником факторное пространство, характерное для специальных «смесевых» задач (см. п. III.13), когда сумма долей всех компонентов v_i смеси есть величина постоянная, например равная 1 (так называемый треугольник Гиббса — Розебома, благодаря работам школы акад. Н. С. Курнакова, нашел широкое применение в физической химии, в металлургии, в технологии цемента, стекла и т. п.).

При введении нормализованных факторов полиномиальная ЭС-модель (III.4) принимает форму

$$\hat{Y} = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i < j} b_{ij} x_i x_j + \sum_{i=1}^k b_{ii} x_i^2 + \dots + \varepsilon, \quad (\text{III.11})$$

в которой коэффициенты $b_0, b_i, b_{ij}, b_{ii} \dots$ через соотношение (B.7) функционально (детерминировано) связаны с коэффициентами a_0, a_i, a_{ij}, a_{ii} и установленными в данном исследовании ограничениями $X_{i, \min}$ и $X_{i, \max}$.

Все коэффициенты в (III.11), во-первых, являются случайными величинами — статистическими оценками истинных параметров полинома $\beta_0, \beta_i, \beta_{ij}, \beta_{ii}, \dots$; во-вторых, в силу безразмерности нормализованных факторов x_i (являющихся, по существу, симплексами подобия [52]) все коэффициенты имеют ту же размерность, что и исследуемый отклик системы Y ; в-третьих, *свободный член* b_0 тождественно равен расчетному значению отклика \hat{Y} в центре факторного пространства ($\hat{Y} = b_0$ при $x_i = 0, i = \overline{1, k}$).

Инженерный смысл оценок b_j и b_{ii} достаточно полно анализируется на однофакторной модели (парабола второго порядка, следующая из (III.11) в случае, когда все остальные $x_{i \geq 2}$ фиксированы):

$$\hat{Y} = b_0 + b_1 x_1 + b_{11} x_1^2. \quad (\text{III.12})$$

Каждый из коэффициентов в (III.12) имеет строгую геометрическую (рис. III.2) и физическую интерпретацию.

Коэффициент b_1 называется *линейным эффектом фактора* x_1 . Как ясно из (III.3), его следует интерпретировать как усредненную в исследуемом диапазоне $-1 \leq x_1 \leq +1$ *скорость изменения отклика* системы $\partial \hat{Y} / \partial x_1$ при управлении фактором x_1 . Если $b_{11} = 0$ и однофакторная модель линейна (рис. III.2, *а, б*), то эта скорость постоянная на всем диапазоне $|x_1| \leq 1$ и численно равна тангенсу угла φ наклона прямой к оси x_1 . Знак перед коэффициентом b_1 определяет, увеличивается (рис. III.2, *а*) или уменьшается (рис. III.2, *б*) выход системы \hat{Y} с ростом x_1 . Максимум линейной модели $\hat{Y}_{\max} = b_0 + |b_1|$ и ее минимум $\hat{Y}_{\min} = b_0 - |b_1|$ всегда ле-

$b_1 = \text{const}$ и определяется как линейная функция координат этой точки:

$$\frac{\partial Y}{\partial x_1} = \frac{\partial (b_0 + b_1 x_1 + b_{11} x_1^2)}{\partial x_1} = b_1 + 2b_{11} x_1. \quad (\text{III.13})$$

Однофакторная модель (III.12) — парабола, форма которой полностью определена квадратичным эффектом, а положение на плоскости (x_1, Y) — числовыми значениями b_0 и b_1 (рис. III.2, в — е). Для одной из комбинаций знаков при b_1 и b_{11} ($b_1 > 0, b_{11} < 0$) параболическая модель показана на рис. III.2, в. Экстремум (вершина) параболы находится в точке $x_{1, \text{ext}}$, которая определяется по выражению, полученному при приравнении частной производной (III.13) нулю (см. п. II.1):

$$x_{1, \text{ext}} = -0,5b_1/b_{11}. \quad (\text{III.14})$$

Численное значение \hat{Y}_{ext} можно найти прямой подстановкой (III.14) в (III.12), и выделить в результате важный с инженерной точки зрения элемент — прирост отклика Δb_0 в экстремальной точке по сравнению с центром эксперимента ($x_1 = 0$):

$$\begin{aligned} \hat{Y}_{\text{ext}} &= b_0 + b_1 (-0,5b_1)/b_{11} + b_{11} (-0,5b_1/b_{11})^2 = \\ &= b_0 - 0,25b_1^2/b_{11}; \end{aligned} \quad (\text{III.15})$$

$$\Delta b_0 = -0,25b_1^2/b_{11} - 0,5b_1 x_{1, \text{ext}}. \quad (\text{III.16})$$

Минимум модели на рис. III.2, в лежит на границе факторного пространства ($x_1 = -1$) и равен $\hat{Y}_{\text{min}} = b_0 - b_1 + b_{11}$; перепад отклика $\Delta \hat{Y} = Y_{\text{max}} - \hat{Y}_{\text{min}} = b_1 - b_{11} - 0,25b_1^2/b_{11}$. Ясно, что по модели (III.12) можно решить *оптимизационную задачу первого рода* — найти оптимальные условия функционирования системы при любом расходе ресурса в пределах $|x_1| \leq 1$ (если расчет по формуле III.14 даст значение $|x_{1, \text{ext}}| > 1$, то целесообразно принять $|x_1| = 1$ хотя бы до проведения контрольных экспериментов при $|x_1| > 1$).

График параболической модели строится по правилам, дополняющим построение ее линейной части: а) на обоих граничных вертикалях факторного пространства откладывается в одинаковом направлении (в соответствии со знаками при b_{11}) квадратичный эффект от точек, соответствующих суммам ($b_0 \pm b_1$); б) через две граничные точки, центральную точку $\{x_1 = 0, \hat{Y} = b_0\}$ и точку экстремума проводится парабола. При этом учитывается ее симметрия относительно вертикали, проходящей через точку $x_{1, \text{ext}}$ (если координаты $x_{1, \text{ext}}$ и \hat{Y}_{ext} не вычислялись, то необходимо сравнить знаки у b_1 и b_{11} : если они разные, то вершина параболы лежит справа от линии центра эксперимента).

С помощью однофакторной модели (III.12) можно решить и *оптимизационную задачу второго рода* — найти минимальный расход ресурса для обеспечения требуемого (нормативного) уровня $Y_{\text{тр}}$

показателя качества системы (рис. III.2, в). Он вычисляется как меньший (в пределах $|x_1| \leq 1$) корень квадратного уравнения, возникающего при подстановке $\hat{Y} = Y_{\text{тр}}$ в (III.12):

$$x_1 \{Y_{\text{тр}}\} = 0,5 [-b_1 \pm \sqrt{b_1^2 - 4(b_0 - Y_{\text{тр}})b_{11}}] / b_{11}. \quad (\text{III.17})$$

Слева от координаты $x_1 \{Y_{\text{тр}}\}$ лежит «запрещенная» область (рис. III.2, в) факторного пространства, справа — область допустимых решений, ограниченная сверху или уровнем $x_1 = +1$ или большим значением корня (III.17).

Пример III.3. На одном из этапов многофакторного исследования бетона (см. пр. III.1) была получена однофакторная модель влияния на предел прочности R_1 концентрации добавки $X_5 = C_{\text{отх}} = 0,15 \pm 0,15 \%$ (в пр. III.2 при условии $x_3 = -1$, $x_6 = +1$, а остальные $x_l = 0$):

$$\hat{R}_1 = 16,6 + 1,4 x_5 - 3,6 x_5^2. \quad (\text{III.18})$$

Решение первой оптимизационной задачи: а) по (III.14) координата $x_{5.\text{max}} = -0,5 \cdot 1,4 / (-3,6) = 0,194$; в натуральных переменных по (B.7) это соответствует расходу добавки $X_5 = C_{\text{отх}} = 0,194 \cdot 0,15 + 0,15 = 0,18 \%$ сухого ПАВ от массы цемента; б) по (III.16) в точке $x_{5.\text{max}}$ прирост отклика $\Delta b_0 = 0,5 \cdot 1,4 \cdot 0,194 = 0,14$ МПа, а значение $\hat{R}_{\text{max}} = 16,6 + 0,14 = 16,74$ МПа; в) по рис. III.2, в $\hat{R}_{\text{min}} = 16,6 - 1,4 - 3,6 = 11,6$ МПа в точке $x_1 = -1$ (нет добавки $C_{\text{отх}}$); г) перепад прочности $\Delta R = 16,74 - 11,6 = 5,14$ МПа во всем диапазоне изменения $C_{\text{отх}}$.

Решение второй оптимизационной задачи для $R_{1.\text{тр}} = 14$ МПа (что соответствует 70 % «марочной» прочности бетона М200) находится по (III.17):

$$x_5 \{R_{1.\text{тр}}\} = 0,5 (-1,4 \pm \sqrt{1,96 - 4(16,6 - 14)(-3,6)}) / (-3,6) = -0,586,$$

что соответствует по (B.7) расходу добавки $X_5 = C_{\text{отх}} = 0,062 \%$ сухого вещества ПАВ от массы цемента. Так как больший корень $x_5 \{R_{1.\text{тр}}\} = 1,02$ выходит за границу факторного пространства, то область допустимых решений ограничена сверху $x_5 = +1$ ($C_{\text{отх}} = 3 \%$).

В многофакторной ($k \geq 2$) полиномиальной модели (III.11) есть группа элементов $b_{ij}x_ix_j$, представляющих попарные произведения факторов x_i и x_j . Коэффициент b_{ij} у такого произведения называется *эффектом взаимодействия*. Возможное число $L\{b_{ij}\}$ эффектов взаимодействия равно числу сочетаний $C_k^2 = k! / [2!(k-2)!] = 0,5k(k-1)$: при $k=2$ число $L\{b_{ij}\} = 1$, при $k=3$ число $L\{b_{ij}\} = 3$, ..., при $k=9$ число $L\{b_{ij}\} = 36$ и т. д. Смысл эффекта взаимодействия выявляется при сравнительном анализе двух простых двухфакторных моделей (без квадратичных эффектов):

$$\hat{Y} = b_0 + b_1x_1 + b_2x_2; \quad (\text{III.19})$$

$$\hat{Y} = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2. \quad (\text{III.20})$$

Каждую из этих моделей можно отразить на обычном «плоскостном» графике только после того, как значение одного из факторов будет стабилизировано на некотором уровне (в частности, на гра-

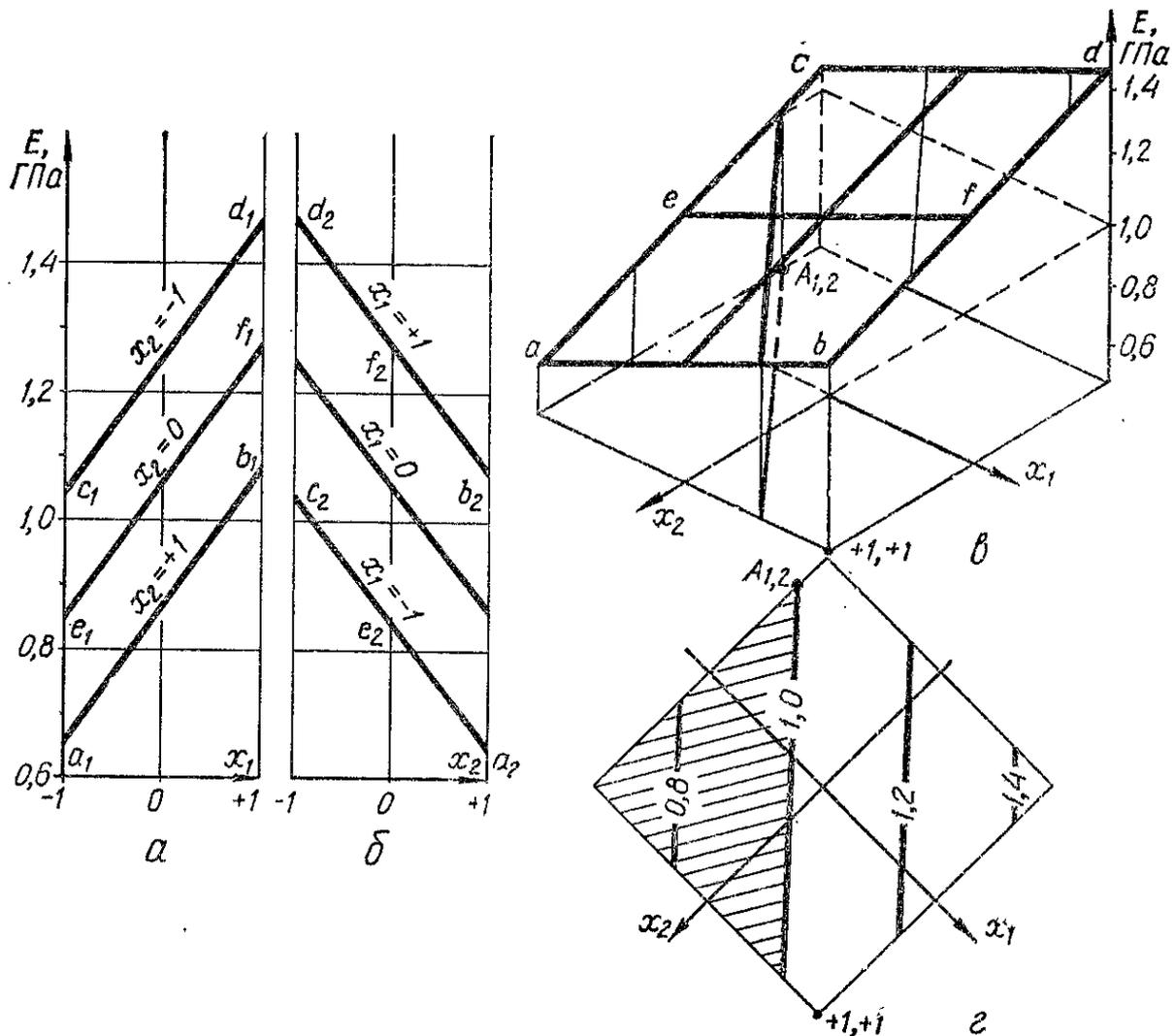


Рис. III.3. Геометрические образы линейной модели:

a — проекции прямых $E = f(x_1, x_2)$ при $x_2 = \text{const}$ на плоскость $\{x_1, E\}$; *б* — то же, при $x_1 = \text{const}$ на плоскость $\{x_2, E\}$; *в* — аксонометрия; *г* — прямые изолинии $E = f(x_1, x_2) = \text{const}$ на плоскость $\{x_1, x_2\}$

ницах и в центре факторного пространства). Для линейной модели (III.19), описывающей плоскость (рис. III.3), графики — семейство параллельных прямых, так как стабилизация $x_i = C_i = \text{const}$ вызывает только изменение свободного члена при постоянном угле наклона прямых $\varphi_i = \text{arctg } b_i$, например $\hat{Y} = (b_0 + b_2 C_2) + b_1 x_1$.

Пример III.4. На базе экспериментальных данных о влиянии пяти рецептурных (X_e, X_p) и четырех технологических ($X_{см}, X_{п.п}$) факторов [42] на показатели качества ячеистого бетона (структурные Y_s , механические и теплофизические Y_o) был получен комплекс многофакторных моделей. Из него, в частности, следует, что модуль упругости $E, \text{ГПа}$, может быть описан (при прочих фиксированных условиях) как линейная функция

$$\hat{E}_{12} = 1,06 + 0,22x_1 - 0,19x_2, \quad (\text{III.21})$$

где x_1 — нормализованное содержание оксида кальция в смеси извести и песка (так называемая «активность смеси» $a_c, \%$), изменяющееся в диапазоне $14 \leq a_c \leq 20 \%$; x_2 — нормализованный показатель вязкости вспучиваемой технологической смеси D_s , регулируемый водотвердым отношением и оцениваемый по ее расплыву (прибор Суттарда): $23 \leq D_s \leq 31 \text{ см}$.

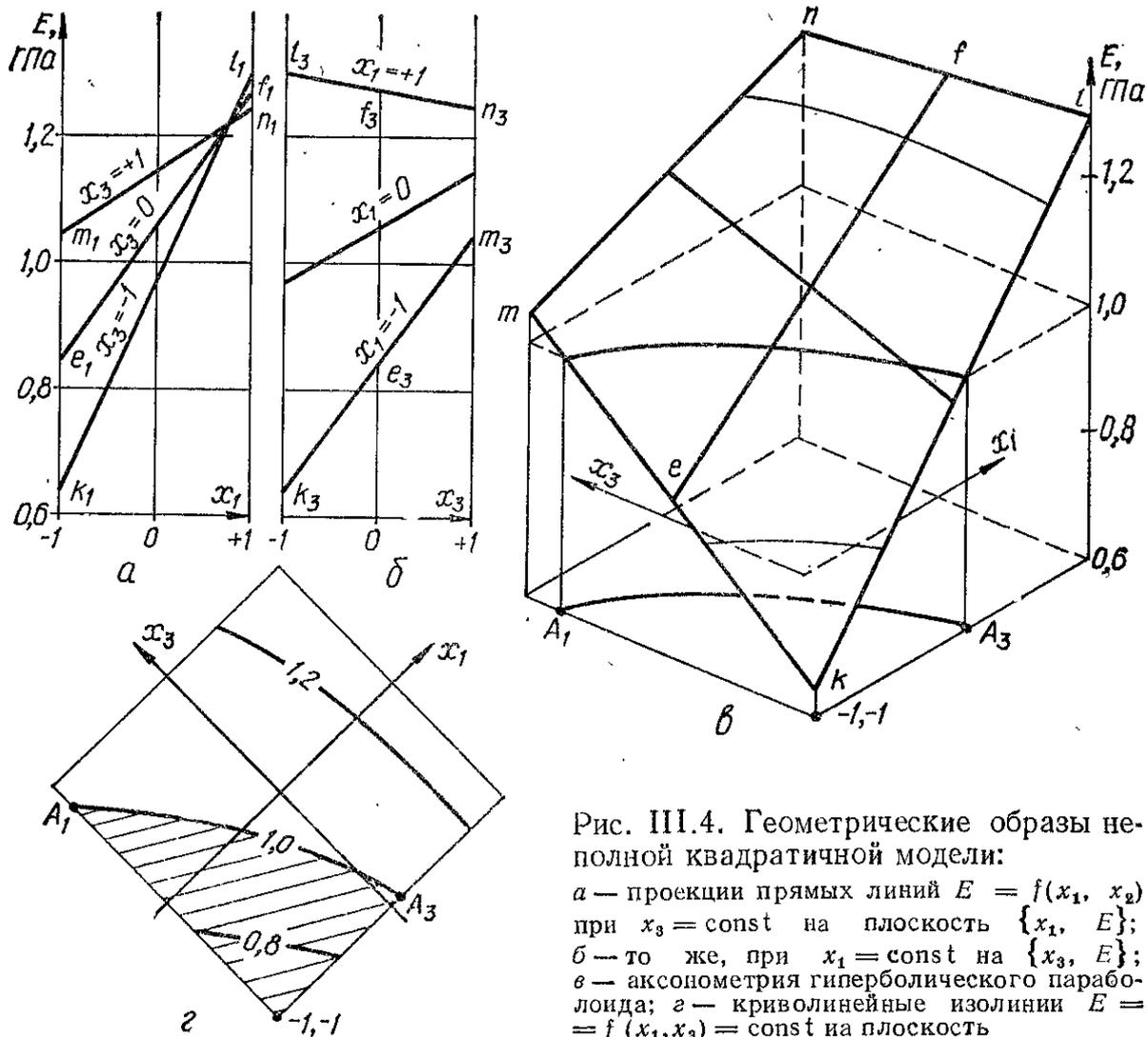


Рис. III.4. Геометрические образы неполной квадратичной модели:

a — проекции прямых $E = f(x_1, x_2)$ при $x_3 = \text{const}$ на плоскость $\{x_1, E\}$; *б* — то же, при $x_1 = \text{const}$ на $\{x_3, E\}$; *в* — аксонометрия гиперболического параболоида; *г* — криволинейные изолинии $E = f(x_1, x_3) = \text{const}$ на плоскость

Результаты построения однофакторных линейных графиков, полученных для граничных точек и центра факторного пространства, а также соответствующей плоскости, показаны на рис. III.3, *a*, *б*, *в*.

Минимум функции $E_{12, \text{min}} = 0,65$ ГПа находится в точке *a* с координатами $x_1 = -1$ (малая активность смеси) и $x_2 = +1$ (большой расплав смеси, связанный с повышением V/T), а максимум $E_{12, \text{max}} = 1,47$ МПа — в противоположном углу факторного пространства *d* при $x_1 = +1$ и $x_2 = -1$. Перепад модуля упругости $\Delta E_{12} = 0,82$ ГПа достигается за счет управления обоими факторами.

Поскольку $|b_2| > |b_1|$, то можно считать, что в исследованном диапазоне влияние показателя вязкости технологической смеси сильнее, чем влияние активности a_c (примерно в полтора раза).

Для модели с эффектом взаимодействия (III.20) однофакторные графики (рис. III.4) — семейство непараллельных прямых, так как стабилизация $x_j = C_j = \text{const}$ вызывает не только изменение свободного члена, но и изменение угла наклона прямых $\varphi_i = \arctg(b_i + b_{ij}C_j)$, например $\hat{Y} = (b_0 + b_2C_2) + (b_1 + b_{12}C_2)x_1$.

Таким образом, эффект взаимодействия меняет усредненную скорость изменения отклика системы \hat{Y} под влиянием управления фактором x_1 в зависимости от того, на каком уровне находится другой фактор x_2 . Возвращаясь к анализу скорости изменения выхода Y в точке x_i для квадратичной модели по (III.13) и положе-

ния ее экстремума по (III.14), нетрудно сделать вывод, что наличие в модели эффекта взаимодействия приведет к перемещению в факторном пространстве и линейной функции для скорости $\partial Y / \partial x_i$ и координат вершины параболы.

Содержащая эффект взаимодействия b_{12} модель (III.20) нелинейна, так как описывает не плоскость (поверхность первого порядка), а *криволинейную поверхность второго порядка* (см. прил. I.2), сечения которой могут проектироваться на плоскости $\{x_1, \hat{Y}\}$ и $\{x_2, \hat{Y}\}$ в виде прямых. Такими свойствами обладает определенным образом ориентированный *гиперболоид* — центральная поверхность второго порядка.

Максимум Y_{\max} и минимум Y_{\min} для модели (III.20) находятся на границах факторного пространства и определяются соотношением абсолютных значений коэффициентов $|b_1|$, $|b_2|$ и $|b_{12}|$ и знаками при них (см. п. III.15).

Пример III.5. По экспериментальным данным пр. III.4 построена и другая модель, описывающая влияние на E одновременного изменения x_1 и x_3 — нормализованной влажности песка $2 \leq \omega_{\text{п}} \leq 8\%$ (при его совместном помоле с известью для улучшения реакционной способности и однородности технологической смеси):

$$E_{13} = 1,06 + 0,22x_1 + 0,09x_3 - 0,12x_1x_3. \quad (\text{III.22})$$

Геометрическим образом модели (III.22) является гиперболоид (см. прил. I.2) с координатами центра S , находящегося вне области эксперимента $|x_i| \leq 1$:

$$\left. \begin{aligned} x_{1S} &= -b_3/b_{13} = -0,09/-0,12 = 0,75; \\ x_{2S} &= -b_1/b_{13} = -0,22/-0,12 = 1,83; \\ \hat{Y}_S &= b_0 - b_1b_3/b_{13} = 1,06 - 0,22 \cdot 0,9/-0,12 = 1,225 \text{ ГПа.} \end{aligned} \right\} \quad (\text{III.23})$$

Стабилизация x_i на трех уровнях ($-1, 0, +1$) приводит к прямым (рис. III.4, а, б с соотношением масштабных единиц по осям E и x_i , равным $m_E / m_x = 6,67$), имеющим разные углы наклона $\varphi_i = \arctg [(b_i + b_{12}C_j) \times m_E / m_x]$ и, следовательно, пересекающимся в точке S (III.23). Так,

если $x_3 = -1$, то $E_{13} = 0,97 + 0,34x_1$ и $\varphi_1 = \arctg (0,34 \cdot 6,67) = 66,2^\circ$,

если $x_3 = 0$, то $E_{13} = 1,06 + 0,22x_1$ и $\varphi_1 = 55,72^\circ$,

если $x_3 = +1$, то $E_{13} = 1,15 + 0,10x_1$ и $\varphi_1 = 33,7^\circ$.

Из рис. III.4, а следует важный инженерный вывод: если при малой активности смеси ($x_1 \rightarrow -1$) лучшие результаты дает помол извести с песком высокой влажности (точка m_1), то при повышенной активности ($x_1 \rightarrow +1$) это технологическое решение становится неэффективным (даже худшим из возможных — точка n_1 самая низкая). Необходимо подчеркнуть, что такая неоднозначность выводов, отражающая *диалектичность реальных систем*, возможна именно благодаря включению в ЭС-модели эффектов взаимодействия факторов.

Прямые на рис. III.4, а, б — образующие поверхности второго порядка (гиперболоида) — видны и на рис. III.4, в. Минимум поверхности отклика $E_{13.\min} = 0,66$ ГПа находится в точке k с координатами $x_1 = -1$ (малая активность смеси) и $x_3 = -1$ (низкая влажность песка при помоле). При том же значении $x_3 = -1$ находится и $E_{13.\max} = 1,3$ ГПа, хотя активность смеси в этом случае максимальна ($x_1 = +1$). Перепад модуля упругости $\Delta E_{13} = 0,64$ ГПа достигается за счет управления только x_1 при условии стабилизации $x_3 = -1$.

Из-за наличия эффекта взаимодействия однозначного решения о приоритете факторов x_1 и x_3 в изменении E_{13} получить нельзя до тех пор, пока не будет фиксирован уровень взаимодействующего фактора. Например, на рис. III.4, а влияние фактора x_1 на E при $x_3 = -1$ (прямая $k_1 l_1$) больше, чем влияние фактора x_3 (рис. III.4, б) для любого уровня фиксации $x_1 = C_1 \ll \leq |1|$, но если $x_3 = +1$ (прямая $m_1 n_1$), то влияние фактора x_1 слабее, чем влияние x_3 при $x_1 = -1$ (прямая $k_3 m_3$).

Двухфакторные модели удобнее всего представлять в виде плоских диаграмм, образуемых семейством изолиний $\hat{Y} = C_Y = \text{const}$ (см. «Введение»), как правило, с равным шагом ΔY . Показатель качества системы Y во всех точках изолинии имеет постоянное значение C_Y , независимо от координат x_1 и x_2 . При этом если заданному набору координат x_1 и x_2 соответствует единственное значение Y , то одному заданному значению $Y_{\text{тр}}$ соответствует множество точек $\{x_1, x_2\}$, лежащих на изолинии. Для того чтобы при изменении одного фактора (например, x_1) уровень свойства $Y_{\text{тр}}$ сохранялся неизменным, необходимо регулировать уровень другого фактора (в частности, x_2), поскольку они связаны жесткой функциональной связью $Y_{\text{тр}} = f_{12}(x_1, x_2)$, или в явном виде $x_2 = f_{1Y}(x_1, Y_{\text{тр}})$. В связи с этим двухфакторные диаграммы в рецептурно-технологических задачах называются регулировочными [4, 6].

Регулировочные диаграммы позволяют расширить рассмотренный выше круг инженерных задач, решаемых на основе ЭС-моделей: можно не только найти значения $Y_{\text{max (min)}}$ и соответствующие координаты, но и выделить «запрещенную» область факторного пространства (см. гл. II), определить условия функционирования системы при наименьшем расходе каждого из видов ресурсов и т. п.

Пример III.6. Если задан уровень модуля упругости $Y_{\text{тр}} = E_{\text{тр}} = 1$ ГПа, то после построения диаграмм (рис. III.3,2 и III.4,2) по моделям E_{12} и E_{13} (пр. III.4 и III.5 соответственно) можно найти области «запрещенных» решений (заштрихованы); их границей являются изолинии $E = 1$ ГПа.

Минимальный расход ресурсов для обеспечения $E = 1$ ГПа как по x_1 , так и по x_2 (рис. III.3, 2) соответствует координатам одной точки — $A_{1,2}$. Их легко определить аналитически: а) по диаграмме $x_1 = -1$; б) после подстановки этого значения в (III.21) получено $E_{12} = 0,84 - 0,19x_2$; в) поскольку $E_{\text{тр}} = 1 = 0,84 - 0,19x_2$, то вторая координата определяется как $x_2 = 0,16 / (-0,19) = 0,84$ (или $D_s = 30,4$ см).

Минимальный расход ресурсов x_1 и x_3 (рис. III.4, 2) одновременно обеспечить нельзя; или минимизируется активность смеси в точке $A_1(x_1 = -1)$, тогда x_3 должно быть не менее 0,77 ($\omega_{\text{п}} = 7,3\%$), или минимизируется влажность песка в точке $A_2(x_3 = -1)$, тогда x_1 должно быть не менее 0,09 ($a_c = 17,3\%$). Для принятия окончательного решения нужно сформулировать некоторый дополнительный критерий, например экономический. Тогда, вообще говоря, точка A_1 будет предпочтительнее, так как затраты на регулировку влажности должны быть относительно меньше.

Двухфакторная модель с квадратичными эффектами b_{11} и (или) b_{22} описывает поверхности второго порядка (см. прил. I.2). Трактовка эффектов b_0, b_i, b_{12} и b_{ii} при этом отличается от вышеизложенного незначительно. То же можно констатировать и в отноше-

нии существа инженерных задач, решаемых по каждой из таких моделей. Модели такого вида будут проанализированы в дальнейшем как при планировании многофакторных экспериментов, так и при решении комплексных рецептурно-технологических задач (см. п. III.15 — III.17).

III.3. Эксперимент в одной точке факторного пространства и статистические характеристики его результатов (малая выборка, одно свойство)

Каждая u -я строка информационной таблицы (см. табл. III.1) в своей левой части или строка плана эксперимента x детерминированно задает координаты одной u -й точки в k -факторном пространстве. В этой точке, т. е. в *неизменных* рецептурных, технологических, эксплуатационных и других фиксируемых *условиях*, экспериментатор реализует m_u испытаний. Однако, несмотря на такую «неизменность» условий, результаты эксперимента $y_{u1}, \dots, y_{uw}, \dots, y_{um_u}$ не будут, как правило, тождественны — они оказываются рассеянными в некотором диапазоне $y_{u, \min} \leq y_{uw} \leq y_{u, \max}$. Результаты отражают диалектическое единство необходимости и случайности, поскольку фиксированные уровни факторов $x_{iu} = \text{const}$ в u -й точке определяют относительное положение этого диапазона на шкале Y , а случайные факторы ξ (III.1) определяют как случайное значение y_{uw} внутри диапазона $y_{u, \min}, \dots, y_{u, \max}$, так и возможную его ширину $\Delta y_u = y_{u, \max} - y_{u, \min}$.

Результаты эксперимента в u -й точке являются, с одной стороны, совершенно *определенной* (детерминированной) *совокупностью* из m_u чисел, поэтому любые алгебраические действия с ними, повторяемые по неизменному алгоритму, приведут к однозначному числовому результату вычислений Θ_u^* . Но, с другой стороны, они являются случайно образованной конечной выборкой объемом m_u единиц из генеральной совокупности объемом N_r единиц ($\infty \geq N_r > m$), структура и свойства которой обусловлены объективными закономерностями, присущими исследуемым системам. При повторном эксперименте в u -й точке эта случайность отразится в том, что выборка объемом m_u будет содержать иной набор экспериментальных результатов y_{uw} , причем выполненные с ними по неизменному алгоритму действия приведут к однозначному числовому результату Θ_{u2}^* , но отличному от предыдущего Θ_{u1}^* .

Таким образом, результаты одинаковых вычислений $\Theta_{u1}^*, \Theta_{u2}^*, \dots, \Theta_{uN}^*$, выполненных $N\{m_u\}$ раз над выборками объемом m_u , образуют *новую случайную выборку* из генеральной совокупности величины $\Theta = f\{Y\}$.

Одна из конечных целей технолога — по информации, ограниченной выборкой, сделать правдоподобные заключения о свойствах исследуемой генеральной совокупности. Отраслевые стандарты, нормирующие свойства строительных материалов и конструкций, регламентируют правила формирования выборок при контроле

качества (отбор и изготовление образцов, их количество, периодичность и т. п.— см., например, для бетонов ГОСТ 10180—78, 27006—86). Специальные стандарты регламентируют правила построения технологических выводов о качестве продукции по результатам анализа выборок (см., например, для бетонов ГОСТ 18105—86, 27005—86).

Свойства генеральной совокупности для случайной величины Y полностью описываются (см. прил. 1.4) функцией распределения вероятностей $P\{Y < y_{тр}\} = F\{Y\}$, для непрерывных величин — функцией плотности вероятности $f(Y) = \partial F(Y)/\partial Y$, называемыми соответственно *интегральным* и *дифференциальным законами распределения* Y .

Во многих инженерных, научно-технических и социально-экономических задачах достаточно знания не самого закона распределения, а лишь некоторых его *числовых характеристик* $\Theta_i = \varphi_i(Y)$, среди которых важнейшими являются два *параметра распределения*:

характеристика *центра группирования* случайной величины, называемая *математическим ожиданием* или *генеральным средним* и обозначаемая $E\{Y\}$ или η :

$$E\{Y\} = \eta = \int_{-\infty}^{\infty} Y f(Y) dY; \quad (\text{III.24})$$

характеристика *рассеяния* случайной величины Y , называемая *дисперсией* и обозначаемая $D\{Y\}$ или σ^2 :

$$D\{Y\} = \sigma^2 = \int_{-\infty}^{\infty} (Y - \eta)^2 f(Y) dY. \quad (\text{III.25})$$

К сожалению, в задачах технологии и материаловедения законы распределения случайных величин Y , как правило, неизвестны, поэтому нельзя воспользоваться формулами (III.24) и (III.25) и приходится ограничиваться *оценками* по выборочным (экспериментальным) данным как генерального среднего η и дисперсии σ^2 , так и любых других истинных параметров и других характеристик генеральной совокупности Θ . Если эти оценки (обозначаются в общем случае Θ^* — «со звездочкой», а нередко латинскими буквами: среднее — \bar{Y} , дисперсия — s^2 и т. д.) выражаются одним числом, то они называются *точечными*, а если множеством чисел, то — *интервальными*. Взаимосвязь между этими двумя видами оценок учитывает, с одной стороны, законы распределения случайных величин Θ^* , изучаемые теорией вероятностей (см. прил. 1.4), с другой — точность и объем экспериментальной информации. Поэтому ряд вычислительных процедур статистических выводов трансформируются в зависимости от того, относятся они к малым или большим выборкам. В строительно-технологических задачах (за исключением некоторых задач нормирования и контроля качества) используются *малые выборки* с числом параллельных испытаний

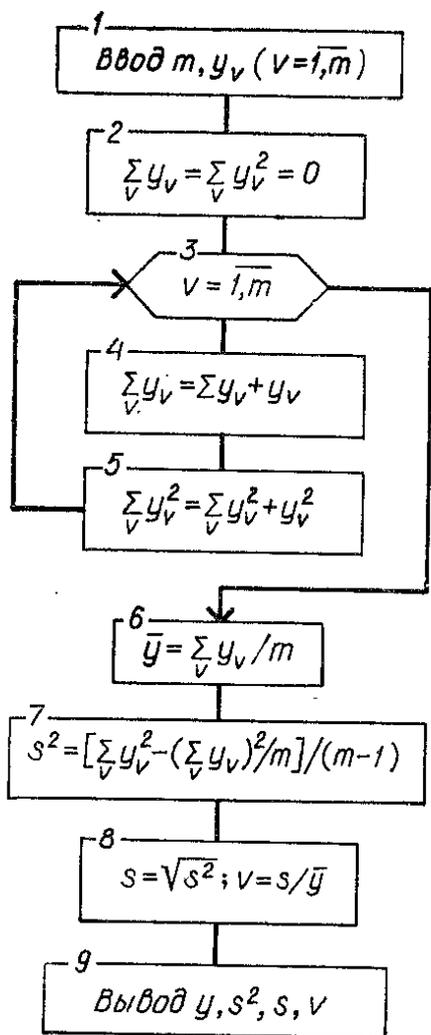


Рис. III.5. Блок-схема алгоритма расчета оценок \bar{y} , s^2 , s , v

$m_u < 50$, причем в большинстве экспериментальных работ (как стандартных, так и исследовательских) диапазон ограничен $2 \leq m_u \leq 6$.

В u -й точке (далее индекс « u » в этом разделе опущен) на основании $m \leq 50$ экспериментальных данных y_v определяются оценки среднего \bar{y} (III.7) и дисперсии s^2 :

$$s^2 = \frac{1}{m-1} \sum_{v=1}^m (y_v - \bar{y})^2 = \frac{1}{m-1} \left[\sum_{v=1}^m y_v^2 - \frac{1}{m} \left(\sum_{v=1}^m y_v \right)^2 \right]. \quad (\text{III.26})$$

При этом используются следующие свойства числовых характеристик в операциях с константами C_c и C_n :

$$\frac{1}{m} \sum_{v=1}^m (y_v C_n + C_c) = \bar{y} C_n + C_c; \quad (\text{III.27})$$

$$\frac{1}{m-1} \sum_{v=1}^m [(y_v C_n + C_c) - (\bar{y} C_n + C_c)]^2 = s^2 C_n^2. \quad (\text{III.28})$$

Оценки \bar{y} (III.7) и s^2 (III.26) позволяют рассчитывать две другие важные для инженерной практики оценки производных числовых характеристик:

среднеквадратического (аналогичный термин — среднеквадратичное), или стандартного, отклонения σ с размерностью, тождественной Y :

$$s = + \sqrt{s^2}; \quad (\text{III.29})$$

безразмерного коэффициента вариации $v = \sigma/\eta$ (относительную величину допускается выражать в процентах)

$$v = s/\bar{y}. \quad (\text{III.30})$$

Необходимо отметить, что оценка коэффициента вариации v положена в основу контроля качества бетона (ГОСТ 18105—86, 27005—86) и изменения по его результатам технологических решений.

Блок-схема алгоритма расчета оценок \bar{y} , s^2 , s и v приведена на рис. III.5.

Пример III.7. Расчет вспомогательных величин для определения оценок \bar{R}_1 , s^2 , s и v прочности бетона в опыте №1 табл. III.2 представлен в табл. III.4. В ее правой части расчет выполняется с использованием $Z = (R_1 - 24) \cdot 10$.

Таблица III.4. К расчету оценок числовых характеристик по малой выборке

Номер образ-ца	Прямой расчет			Расчет через суммы преобразованных результатов		
	Результат $y_v = R_{1v}$, МПа	$y_v - \bar{y}$	$(y_v - \bar{y})^2$	$C_c = -24$	$C_{\Pi} = 10$	Z_v^2
				$Z'_v = R_{1v} + C_c$	$Z_v = Z'_v C_{\Pi}$	
1	20,3	-3,08	9,4864	-3,7	-37	1369
2	22,7	-0,68	0,4624	-1,3	-13	169
3	24,0	+0,62	0,3844	0	0	0
4	24,0	+0,62	0,3844	0	0	0
5	24,5	+1,12	1,2544	+0,5	+5	25
6	24,8	+1,42	2,0164	+0,8	+8	64
Сумма	$\sum_{v=1}^6 y_v = 140,3$	контроль: $\sum_{v=1}^6 (y_v - \bar{y}) = 0,02 \approx 0$	$\sum_{v=1}^6 (y_v - \bar{y})^2 = 13,9884$	$\sum_{v=1}^6 Z'_v = -3,7$	$\sum_{v=1}^6 Z_v = -37$	$\sum_{v=1}^6 Z_v^2 = 1627$

Расчет оценки среднего по (III.7) и (III.27):

$$\bar{R}_1 = \frac{1}{6} \sum_{v=1}^6 y_v = 140,3/6 = 23,38;$$

$$\bar{Z} = \frac{1}{6} \sum_{v=1}^6 Z_v = -37/6 = -6,17;$$

$$\bar{R}_1 = \bar{Z}/C_{\Pi} - C_c = -6,17/10 + 24 = 23,38.$$

Оценка дисперсии по (III.26) и (III.28):

$$s^2 \{R_1\} = \frac{1}{6-1} \sum_{v=1}^6 (y_v - \bar{y})^2 = 13,9884/5 = 2,7977;$$

$$s^2 \{Z\} = \frac{1}{6-1} \left[\sum_{v=1}^6 Z_v^2 - \frac{1}{6} \left(\sum_{v=1}^6 Z_v \right)^2 \right] = \left[1627 - \frac{(-37)^2}{6} \right] / 5 = 279,77;$$

$$s^2 \{R\} = s^2 \{Z\} / C_{\Pi}^2 = 279,77/10^2 = 2,7977.$$

Оценки среднеквадратического отклонения по (III.29) и коэффициента вариации по (III.30):

$$s \{R_1\} = +\sqrt{s^2 \{R_1\}} = +\sqrt{2,7977} = 1,67 \text{ МПа};$$

$$v = s \{R_1\} / \bar{R}_1 = 1,67/23,38 = 0,070 \text{ или } 7 \%.$$

Кроме вышеуказанных оценок числовых характеристик, называемых *параметрическими*, поскольку они тесно связаны с параметрами закона распределения случайной величины Y (см. прил. 1.4), вид которого предполагается известным исследователю, можно в ряде

случаев ограничиваться *непараметрическими* оценками. Они статистически менее эффективны (см. далее этот раздел о свойствах оценок Θ^*), однако не требуют никаких предпосылок о знании закона распределения Y и более просты для вычислений. Среди непараметрических оценок наиболее широко применяются порядковые, или ранговые, получаемые при анализе ранжированного ряда из m результатов:

$$y_{(1)\min} \leq y_{(2)} \leq \dots \leq y_{(v)} \leq \dots \leq y_{(m)\max}. \quad (\text{III.31})$$

Частота $p_l = m_l\{y\}/m$ — относительное (ко всей выборке объемом m) число случаев $m_l\{y\}$ появления случайной величины Y в некотором интервале от $y = l - \Delta l$ до $y = l + \Delta l$.

Размах — широта рассеяния в ряду (III.31)

$$W = y_{(m)\max} - y_{(1)\min}. \quad (\text{III.32})$$

При малых выборках ($m \leq 10$) из нормально распределенных генеральных совокупностей величиной размаха допустимо пользоваться (см., в частности, ГОСТ 18 106—86, 27 005—86) для быстрого определения несмещенной оценки $s = K_W W$ с помощью табл. III.5 [2].

Таблица III.5. K_W для оценки s по размаху W

m	K_W	m	K_W	m	K_W
2	0,886	5	0,430	8	0,351
3	0,591	6	0,395	9	0,337
4	0,486	7	0,370	10	0,325

Медиана Me — значение случайной величины Y , приходящееся на середину ранжированного ряда (частоты слева и справа равны 0,5); для четного числа членов определяется как среднее из двух срединных элементов ряда.

Квантиль Q_α — значение случайной величины Y , занимающее в ряду такое положение, что ниже его находится часть совокупности, равная α (следовательно, часть ряда выше Q_α равна $1 - \alpha$); медиана — частный случай квантиля $Q_{0,5} = Me$.

Пример III.8. Ранжированный ряд

$$(y_{(1)} = 20,8) < (y_{(2)} = 22,7) < (y_{(3)} = 24) = (y_{(4)} = 24) < \\ < (y_{(5)} = 24,5) \leq (y_{(6)} = 24,8)$$

прочности мелкозернистого бетона R_1 , МПа (см. пр. III.7), характеризуется следующими непараметрическими оценками:

размах $W = y_{(6)} - y_{(1)} = 24,8 - 20,8 = 4$ МПа;

частота появления случайной величины R_1 в диапазоне, например, $(24 \pm \pm 0,5)$ МПа составляет $p_{24} = 3/6 = 0,5$ (учтены $y_{(3,4,5)}$);

медиана $Me = (24 + 24) / 2 = 24$ МПа (между $y_{(3)}$ и $y_{(4)}$);

квантиль для части ряда, например $\alpha = 0,2$, определяется по алгоритму:

а) порядковый номер m_α^0 члена ряда, соответствующего Q_α ,

$$m_\alpha^0 = \alpha (m + 1) = 0,2 (6 + 1) = 0,2 (6 + 1) = 1,4; \quad (\text{III.33})$$

- б) предшествующий m_α^0 член ранжированного ряда $y'_\alpha = y_{(1)} = 20,8$ МПа;
 в) последующий за m_α^0 член ряда $y'' = y_{(2)} = 22,7$ МПа;
 г) интерполяция между ними дает (с учетом доли 0,4 — см. (III.33))

$$Q_\alpha = 20,8 + 0,4(22,7 - 20,8) = 21,56 \text{ МПа.}$$

Экспериментальные данные всегда представляют собой статистическую совокупность ограниченного объема m . По мере того как для этой совокупности определяются различные точечные оценки $\Theta_i^* = \varphi_i(y_1, \dots, y_m)$, возможности для последующих вычислений, приводящих к независимым результатам, постепенно исчерпываются (вплоть до полного их запрещения). Простейшим пояснением этого является результат вычисления среднего, например, из двух величин $(y_a + y_b)/2 = \bar{y}$. После фиксации \bar{y} лишь одна из них (или y_a или y_b) может изменяться случайно и независимо, поскольку другая следует из нее детерминированно ($y_b = 2\bar{y} - y_a$). Число независимых элементов статистической совокупности, оставшихся после наложения на нее тех или иных связей (вычисления λ различных оценок Θ^*), называется *числом степеней свободы*:

$$f = m - \lambda. \quad (\text{III.34})$$

Если при вычислении среднего число степеней свободы равно числу элементов ряда m , то при вычислении дисперсии $f = m - 1$, что и учтено в формуле (III.26). По мере использования других оценок будет оговариваться и для них число степеней свободы.

Для того чтобы инженерные и другие логико-профессиональные выводы, сделанные по статистической совокупности, были «доброкачественны», к точечным оценкам Θ^* предъявляется ряд требований. Оценка должна быть *состоятельной*, то есть по мере увеличения объема m данной выборки она должна сходиться по вероятности [10, 32, 53] к истинному значению параметра Θ . Целесообразно, чтобы оценка Θ удовлетворяла требованию *несмещенности*, заключающемуся в том, что результат усреднения значений Θ_u^* , полученных по разным выборкам заданного объема, дает в точности истинное значение оцениваемого параметра ($E\{\Theta^*\} = \Theta$) [10]. Так, замена в формуле (III.26) в знаменателе числа степеней свободы $f = m - 1$ на число параллельных измерений m приводит к оценке, смещенной в меньшую сторону на σ^2/m (что будет значительно искажать результаты анализа в малых выборках). Желательно, чтобы используемая оценка Θ^* была *эффективной*, т. е. среди себе подобных по смыслу оценок она обладала бы наименьшей относительно истинного значения Θ дисперсией

$$D\{\Theta^*\} = E\{(\Theta^* - \Theta)^2\}. \quad (\text{III.35})$$

Так, дисперсия медианы $D\{Me\} = 1,576 s^2/m$ [4, 10], а дисперсия вычисленного по (III.7) среднего $D\{\bar{Y}\} = s^2/m$, следовательно, вторая оценка эффективней. Понятие об эффективности оценок является одним из базовых в математической теории эксперимента (см. п. III.10 — III.14), хотя иногда эффективностью (так же,

как и несмещенностью) приходится на практике жертвовать, в частности в связи со сложностью расчета оценок.

Точечная оценка $Q^* = \varphi(y_1, \dots, y_m)$ числовой характеристики распределения случайной величины Y , рассчитанная по результатам ограниченного числа экспериментов, является, как уже подчеркивалось выше, *случайной величиной* и поэтому может не совпадать с истинным (неслучайным!) значением оцениваемого параметра Θ . Однако знание закона распределения случайной величины Θ^* позволяет построить вокруг точечной оценки некоторую ограниченную область, которая должна с заданной вероятностью q накрыть истинное значение параметра Θ . Такая область (III.36) называется *доверительной*, а образуемая ею оценка — *интервальной* с нижней Θ_n и верхней Θ_b границей:

$$P(\Theta_n \leq \Theta \leq \Theta_b) = q = 1 - \alpha. \quad (\text{III.36})$$

Вероятность α противоположного события — интервальная оценка (III.36) не накрыла детерминированный параметр (точку Θ) — называется *уровнем значимости статистической гипотезы* (гр. *hypothesis* — предположение). Действительно, при построении доверительного интервала (III.36) выдвигается следующее предположение о свойствах генеральной совокупности: математическое ожидание оценки $E\{\Theta^*\}$ равно истинному значению параметра Θ , т. е. разность $\Theta - E\{\Theta^*\} = 0$. Такие статистические гипотезы, в которых предполагается *отсутствие разницы* между двумя истинными параметрами (отклонения возникают не по систематическим причинам, а как случайные колебания в выборках Θ^*), называются *нуль-гипотезами* и обозначаются $H_0: \Theta = E\{\Theta^*\}$.

Вообще любые суждения (гипотезы) о свойствах генеральной совокупности на основании экспериментальных (выборочных) данных представляют собой перенос знаний *от части к целому*. Этот перенос, обусловленный диалектическим единством необходимости и случайности, может быть осуществлен лишь как *вероятностное суждение*, а значит, объективно содержит некоторый *риск* (вероятность) совершить *ошибку в выводах*.

Статистические гипотезы проверяются с помощью специальных правил — *статистических критериев*. Количественно они оцениваются с помощью некоторой специально подобранной функции от результатов эксперимента $\gamma = \psi(y_1, \dots, y_m)$, для которой в теории вероятности хорошо изучена (и протабулирована) функция распределения вероятностей $f(\gamma)$, показанная в общем виде на рис. III.6. Если рассчитанный по данным i -й выборки критерий γ_i попадает в область *правдоподобных* значений, то гипотеза *допускается*; если в область неправдоподобно малых (ограничена сверху значением γ_n) или неправдоподобно больших (ограничена снизу γ_b) — *отклоняется*.

Все критерии по своей природе *отрицательны* — попадание γ в область принятия гипотезы означает лишь то, что она *не противоречит* экспериментальным данным. Эту гипотезу с риском $\alpha =$

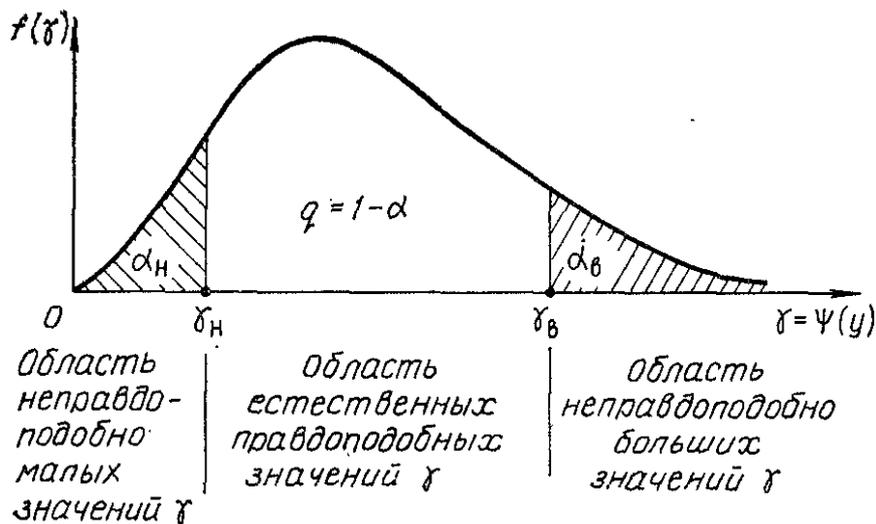


Рис. III.6. Функция распределения вероятностей статистического критерия γ и области ошибок первого рода α

$= \alpha_H + \alpha_B$ можно признать правомерной, по крайней мере до тех пор, пока исследования по расширенной информации или с помощью более совершенных (статистически более «мощных») критериев не приведут к противоположному результату. Следует особо подчеркнуть, что неотрицательный результат проверки H не означает, что эта гипотеза лучшая и единственная (свойством непротиворечивости могут обладать и другие гипотезы). Ее следует рассматривать лишь как одно из правдоподобных (а не абсолютно достоверных) утверждений.

Ошибка в выводах может быть двух родов. Если отвергнута правильная гипотеза, то совершается *ошибка первого рода*. Вероятность α допустить ее и названа выше уровнем значимости, а соответствующие ей области (рис. III.6) — критическими. Если будет допущена неправильная гипотеза, то совершается *ошибка второго рода*. Ее вероятность β дополняет значение мощности критерия $B = 1 - \beta$. В технико-экономических задачах, в частности при контроле качества продукции, α и β называются соответственно *рисками производителя* и *потребителя*. Все четыре возможные при проверке гипотез ситуации приведены в табл. III.6.

Одной из основных задач теории статистической проверки гипотез является построение математиками таких критериев γ , которые бы при заданном диапазоне риска α минимизировали риск β . Как правило, назначается риск первого рода α , численное значение которого зависит от технико-экономической *тяжести негативных последствий*, возникающих от допущения неправильных гипотез, а потом выбирают среди критериев такой, который бы обладал наибольшей мощностью B .

Назначение во всех задачах моделирования «стандартного» 5 %-ного уровня риска $\alpha = 0,05$ следует рассматривать не более, чем некоторую исторически сложившуюся традицию. Уровень риска α должен снижаться с ростом цены потерь от неправильных выводов, т. е. зависеть как от объекта анализа и оптимизации, так

Таблица III.6. Возможности ошибок при проверке статистических гипотез (на примере проверки равенства генерального среднего $\eta\{Y\}$ заданному значению $\eta_{тр}$)

Какая из двух гипотез истинна?	Что рекомендует критерий γ ?	
	Допустить нуль-гипотезу	Отклонить нуль-гипотезу, а допустить альтернативную гипотезу
Фактически истинна нуль-гипотеза $H_0: \eta\{Y\} = \eta_{тр}$	Решение <i>правдоподобно</i> : допущена гипотеза $H_0: \eta\{Y\} = \eta_{тр}$	Решение <i>ложно</i> , так как совершена ошибка <i>первого рода</i> , поскольку отклонена верная гипотеза H_0 ; <i>риск поставщика α (producer's risk)</i> — вероятность ошибки, при которой <i>годную</i> партию продукции могут в результате колебаний выборочной оценки признать не соответствующей техническим требованиям
Фактически истинна альтернативная гипотеза $H_1: \eta\{Y\} < \eta_{тр}$	Решение <i>ложно</i> , так как совершена ошибка <i>второго рода</i> , поскольку допущена ложная гипотеза H_0 вместо истинной H_1 ; <i>риск β потребителя (consumer's risk)</i> — вероятность ошибки, при которой негодную партию продукции могут в результате колебания выборочной оценки ошибочно признать годной	Решение <i>правдоподобно</i> , так как допущена гипотеза $H_1: \eta\{Y\} < \eta_{тр}$

и от стадии его исследования. При назначении риска можно руководствоваться следующим [4, 6, 23]:

- а) предварительные поисковые работы — $0,1 \leq \alpha \leq 0,2$,
- б) окончательные рецептурно-технологические решения в лабораторных и производственных условиях (кроме п. «в, г») — $0,05 \leq \alpha \leq 0,1$;
- в) анализ и контроль качества материалов в производственных условиях (кроме п. «г») — $0,01 \leq \alpha \leq 0,05$;
- г) анализ и контроль качества конструкционных и защитных материалов, а также токсичных, взрывоопасных и других веществ, применение которых связано с опасностью для жизни и здоровья людей — $\alpha \leq 0,01$.

Пример III.9. Проверить гипотезу о равенстве истинной прочности $\eta\{R_1\}$ мелкозернистого бетона (в условиях пр. III.1 и III.7) требуемому уровню $R_{1,тр} = 25$ МПа.

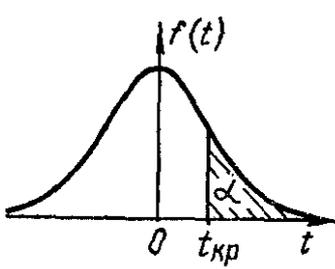
Основные этапы проверки гипотезы с помощью критерия

$$t = |\bar{Y} - \eta_{тр}| \frac{\sqrt{m-1}}{s\{Y\}}, \quad (III.37)$$

допускающего отсутствие информации об истинной дисперсии $\sigma^2\{Y\}$, показаны в табл. III.7.

Результаты диалектически связаны со стадией исследования, на которой получены экспериментальные данные:

Таблица III.7. Основные этапы проверки статистических гипотез (на примере проверки равенства генерального среднего заданному значению $\eta_{тр}$)

Номер этапа	Цель	Решение	Численный пример
1	Определить нуль-гипотезу и альтернативу	Нуль-гипотеза: «истинное среднее $\eta\{Y\}$ равно требуемому значению $\eta_{тр}$ »	$H_0: \eta\{Y\} = 25$ МПа $H_1: \eta\{Y\} < 25$ МПа
2	Установить уровень значимости	а) 0,1 — поисковые работы б) 0,05 — контроль качества	
3	Выбрать критерий γ и определить распределение $f(\gamma)$	Случайная величина $t = [(\bar{Y} - \eta_{тр}) \sqrt{m-1}] / s\{Y\}$ подчиняется t -распределению с числом степеней свободы $f = m - 1$	
4	Установить критическую область для проверки нуль-гипотезы	 <p>График функции плотности вероятности $f(t)$ для t-распределения. Ось абсцисс — t, ось ординат — $f(t)$. Вертикальная линия от 0 до пика кривой. Другая вертикальная линия от $t_{кр}$ до оси абсцисс. Область под кривой к правую от $t_{кр}$ заштрихована и обозначена α.</p> <p>Табулировано в прил. II.5.</p>	<p>При $f = 6 - 1 = 5$ и одностороннем риске α граничное значение критерия равно:</p> <p>а) для $\alpha = 0,1$ $t_{кр}^a = 1,476$ б) для $\alpha = 0,05$ $t_{кр}^b = 2,571$</p>
5	Вычислить значение критерия по данным выборки	$t_{факт} = \frac{ \bar{Y} - \eta_{тр} }{s\{Y\}} \sqrt{m-1}$	$t_{факт} = \frac{ 23,38 - 25 }{1,67} \times \sqrt{5} = 2,169$
6	Сравнить $t_{факт}$ с $t_{кр}$ и решить, не отвергается ли гипотеза H_0	Если $t_{факт} \leq t_{кр}$, то H_0 не противоречит наблюдениям и не отвергается; если $t_{факт} > t_{кр}$, то нуль-гипотеза отклоняется как неправдоподобная и $\eta\{Y\} < \eta_{тр}$	<p>а) поскольку $t_{факт} > t_{кр}^a$, H_0 отклоняется и на поисковом этапе признается, что R_1 не достигает требуемого значения;</p> <p>б) поскольку $t_{факт} < t_{кр}^b$, то H_0 принимается и при контроле качества можно считать, что $\eta\{Y\} = 25$ МПа</p>

а) если это поисковый этап работ и можно назначить повышенный риск отклонить правильную гипотезу, то критический уровень критерия снижается и нуль-гипотеза отклоняется (с риском 10 %);

б) если те же экспериментальные данные соответствуют контролю качества производителем продукции, то риск ошибки первого рода назначается сниженным и нуль-гипотезу $\eta\{R_1\} = 25$ МПа рекомендуется допустить как правдоподобную.

Построение доверительных интервалов Θ_H и Θ_B (III.36), по сути, один из видов проверки статистических гипотез. Так, в частности, при построении доверительных интервалов для генерального среднего $\eta\{Y\}$ можно исходить из факта распределения величины t , использованной в пр. III.9, по закону Стьюдента (см. прил. I.4 и п. II.5). Тогда вероятность попадания ее в интервал от t_H до t_B определяется соотношением

$$P\left\{t_H \leq \frac{|\bar{Y} - \eta|}{s\{Y\}} \sqrt{m-1} < t_B\right\} = 1 - \alpha.$$

Поскольку распределение Стьюдента симметрично, то односторонний риск снизу α_H целесообразно принять равным одностороннему риску сверху α_B , после чего справедливо

$$P\left\{-t\{\alpha/2\} \leq \frac{|\bar{Y} - \eta|}{s\{Y\}} \sqrt{m-1} < +t\{\alpha/2\}\right\} = 1 - \alpha,$$

откуда следуют:

$$P\left\{\bar{Y} - t\{\alpha/2\} s\{Y\} (m-1)^{-1/2} \leq \eta < \bar{Y} + t\{\alpha/2\} s\{Y\} (m-1)^{-1/2}\right\} = 1 - \alpha; \quad (\text{III.38})$$

$$P\left\{\bar{Y} - \Delta_\alpha\{\bar{Y}\} \leq \eta < \bar{Y} + \Delta_\alpha\{\bar{Y}\}\right\} = 1 - \alpha; \quad (\text{III.39})$$

$$\Delta_\alpha\{\bar{Y}\} = t\{\alpha/2\} s\{Y\} \frac{1}{\sqrt{m-1}}. \quad (\text{III.40})$$

Соотношения (III.38) — (III.40) позволяют сделать важные для материаловеда и технолога метрологические выводы. Во-первых, абсолютная ошибка в определении среднего $\Delta_\alpha\{\bar{Y}\}$ при постоянном риске α будет уменьшаться как по мере увеличения количества образцов m в данной выборке, так и по мере увеличения точности эксперимента $s\{Y\} \rightarrow \min$. Во-вторых, при заданных уровнях риска $\alpha_{\text{тр}}$ и ошибки среднего $(\Delta_\alpha\{\bar{Y}\})_{\text{тр}}$ можно найти обеспечивающее их число образцов:

$$m_{\min} = 1 + t^2\{\alpha/2\} s_{\text{тр}}^2\{Y\} / (\Delta_\alpha\{\bar{Y}\})_{\text{тр}}^2. \quad (\text{III.41})$$

Если же заданы ошибка эксперимента $s_{\text{тр}}\{Y\}$ или минимальное (по технико-экономическим условиям) число образцов $m_{\text{тр}}$, то можно оценить предельную ошибку эксперимента

$$s_{\max}\{Y\} = (\Delta_\alpha\{\bar{Y}\})_{\text{тр}} (\sqrt{m_{\text{тр}} - 1}) / t\{\alpha/2\} \quad (\text{III.42})$$

и выбрать соответствующее испытательное оборудование и измерительные приборы.

Следует обратить внимание на то, что расчет m по (III.41) итерационный — последовательно задаются m' при нахождении $t\{\alpha/2\}$

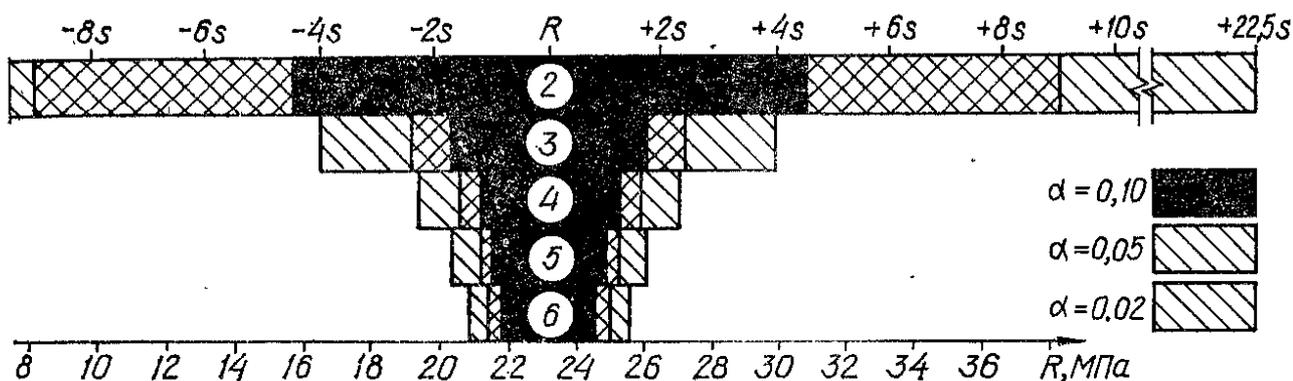


Рис. III.7. Изменение доверительного интервала в зависимости от риска α и объема информации m

по таблицам и находятся m'' , потом задаются m''' и т. д., пока процесс не сойдется.

Пример III.10. Исходя из условий пр. III.7 ($\bar{R}_1 = 23,38$ МПа; $s\{R_1\} = 1,67$; $m = 6$), определить доверительный интервал истинного среднего $\eta\{R\}$ при двустороннем риске $\alpha = 0,1$.

Полудиализон (III.40) рассчитывается при $t = 2,015$ ($f = m - 1 = 5$; $\alpha = 0,1$ — см. прил. II.5):

$$\Delta_{\alpha}\{\bar{R}_1\} = 2,015 \cdot 1,67 / \sqrt{5} = 1,5 \text{ МПа,}$$

следовательно, интервальная оценка $\eta\{R_1\}$

$$\left. \begin{aligned} P\{23,38 - 1,5 \leq \eta\{R_1\} \leq 23,38 + 1,5\} &= 1 - 0,1; \\ P\{21,88 \leq \eta\{R_1\} < 24,88\} &= 0,9. \end{aligned} \right\} \text{ (III.43)}$$

Изменение интервала для $\eta\{R_1\}$ в зависимости от риска α показано на рис. III.7; там же для сравнения дано изменение величины $2\Delta_{\alpha}\{R_1\}$ при уменьшении числа образцов в выборке.

Можно определить метрологические показатели эксперимента, если заданы уровень риска, например $\alpha = 0,05$, и абсолютная ошибка, например $(\Delta_{\alpha}\{R_1\})_{\text{тр}} = 1,25$ МПа (взята как 5 %-ная доля от $R_{1,\text{тр}} = 25$ МПа в пр. III.9).

Если задана ошибка эксперимента $s_{\text{тр}}\{R_1\}$ (пусть она, например, равна 1,2 МПа), то определяется число образцов:

пусть $m' = 6$, тогда при $f = 5$ и $t\{0,05\} = 2,571$

$$m_{\text{min}} = 2,571^2 \cdot 1,2^2 / 1,25^2 + 1 = 7,1 \approx 8 \text{ шт.};$$

пусть $m'' = 8$, тогда при $f = 7$ и $t\{0,05\} = 2,365$

$$m_{\text{min}} = 2,365^2 \cdot 1,2^2 / 1,25^2 + 1 = 6,2 \approx 7 \text{ шт.};$$

пусть $m''' = 7$, тогда при $f = 6$ и $t\{0,05\} = 2,447$

$$m_{\text{min}} = 2,447^2 \cdot 1,2^2 / 1,25^2 + 1 = 6,5 \approx 7 \text{ шт.};$$

процесс сошелся при $m = 7$ образцов.

Если задано предельное число образцов, например $m_{\text{т}} = 3$ (стандартное число гнезд в формах), то предельная ошибка эксперимента (при $f = 2$, $t\{0,05\} = 4,303$)

$$s_{\text{max}}\{R_1\} = 1,25 \sqrt{3 - 1} / 4,303 = 0,41 \text{ МПа.}$$

Такое решение потребовало перехода к испытательным машинам более высокого класса, так как относительная ошибка снизилась примерно до 2% (0,41 МПа/25 МПа).

Доверительные интервалы числовых характеристик Θ могут быть как симметричными (III.40), так и несимметричными, если случайные величины Θ^* подчиняются асимметричным распределениям (χ^2 , F и др. — прил. I.4). Так, при малых выборках доверительные интервалы для истинного среднеквадратического отклонения σ (III.29) вычисляются по точной формуле

$$P \left\{ s \sqrt{\frac{m-1}{\chi^2\{\alpha/2\}}} \leq \sigma < s \sqrt{\frac{m-1}{\chi^2\{1-\alpha/2\}}} \right\} = 1 - \alpha, \quad (\text{III.44})$$

а для истинного коэффициента корреляции γ (III.30) — по приближенной

$$P \left\{ \sqrt{\frac{m}{\chi^2\{\alpha/2\}(v^2-1)-1}} \leq \gamma < \sqrt{\frac{m}{\chi^2\{1-\alpha/2\}(v^2-1)-1}} \right\} = 1 - \alpha. \quad (\text{III.45})$$

В них входит критерий χ^2 (с числом степеней свободы $f = m - 1$, см. прил. I.4 и п. II.4), распределенный с правой асимметрией, особенно сильной при малых m .

Для непараметрических оценок (III.31) — (III.33) интервалы, в силу сложности функций, определяются по специальным таблицам. Некоторые доверительные интервалы проанализированы для больших выборок (см. п. III.5).

Пример III.11. Исходя из условий пр. III.7 ($\bar{R}_1 = 23,38$ МПа, $s\{R_1\} = 1,67$; $v = 0,07$; $m = 6$), определить доверительный интервал для дисперсии σ^2 , среднеквадратического отклонения σ и коэффициента вариации γ при доверительном двустороннем риске $\alpha = 0,1$.

При $f = 6 - 1 = 5$ критерий $\chi^2\{0,05\} = 11,07$, а $\chi^2\{0,95\} = 1,145$ (см. прил. II.4), поэтому

$$P \left\{ 1,67 \sqrt{\frac{5}{11,07}} \leq \sigma < 1,67 \sqrt{\frac{5}{1,145}} \right\} = 1 - 0,1;$$

$$P\{1,12 \leq \sigma < 3,49\} = 0,9.$$

От экспериментального $s\{R_1\} = 1,67$ диапазон влево $-0,55$, а вправо $+1,82$; для σ^2 асимметрия еще больше:

$$P\{1,25 \leq \sigma^2 < 12,18\} = 0,9.$$

Сильная асимметрия и у величины γ :

$$P \left\{ \sqrt{\frac{6}{11,070(1/0,07^2-1)-1}} \leq \gamma < \sqrt{\frac{6}{1,145(1/0,07-1)-1}} \right\} = 1 - 0,1;$$

$$P\{0,052 \leq \gamma \leq 0,161\} = 0,9.$$

Определение статистических характеристик (точечных и интервальных оценок) результатов эксперимента в одной точке факторного пространства относится к этапу первичной обработки эмпирической информации, который обязателен в исследованиях любого вида. Эти задачи имеют хорошее типовое программное обеспечение для ЭВМ всех типов, а также для программируемых микрокалькуляторов [16, 49 и др.].

III.4. Эксперимент в одной точке факторного пространства и его статистические характеристики [малая выборка, несколько свойств]

В u -й точке факторного пространства, вообще говоря, может измеряться не одно, а M различных по физической природе свойств объекта. Это нашло отражение в структуре информационной табл. III.1, где фиксируются как значения $y_{j,uv}$ в каждом v -м измерении, так и средние величины $\bar{y}_{j,u}$. Ясно, что в общем случае каждой средней величине $\bar{y}_{1,u}$ (обозначена индексом «1» без потери общности) соответствует одно среднее значение других свойств $\bar{y}_{j,u}$ ($j = \overline{2, M}$), поэтому для установления степени взаимосвязи между случайными величинами $\bar{y}_{1,u}$ и $\bar{y}_{j,u}$ данных, содержащихся в одной u -й строке информационной таблицы, недостаточно — нужно рассматривать таблицу в целом.

Но может быть частный случай, когда все измерения свойства $1, \dots, j, \dots, M$ последовательно проводятся на каждом uv -м образце с фиксированным номером, например: балочка взвешивается (G_{uv}^c), определяется ее объем (v_{uv}) и вычисляется плотность ($\rho_{uv} = G^c/v$); измеряется скорость прохождения в ней ультразвука (c_{uv}^c); образец помещается в воду, вновь взвешивается (G_{uv}^b) и вычисляется его водопоглощение по массе ($\omega_{uv} = [G^b - G^c]/G^c$), повторно измеряется скорость прохождения ультразвука (c_{uv}^b), разрушением этого образца определяется предел прочности на изгиб (R_{uv}^b), по шлифу из этой балочки оценивается количество новообразований, пор и дефектов в структуре ... Такая схема организации эксперимента позволяет оценить в условиях постоянства уровней основных факторов $X_i = \text{const}$ степень взаимосвязи между случайными величинами $Y_1, \dots, Y_i, \dots, Y_M$ — показателями свойств исследуемого объекта. При этом корректные заключения будут получены, если правдоподобны следующие *предположения* [11]. Во-первых, результаты v -х ($v = \overline{1, m_u}$) измерений свойств взаимонезависимы, т. е. образуют случайную выборку их генеральной совокупности. Во-вторых, координаты точек случайного множества, образованного результатами определения M свойств, распределены по M -мерному нормальному закону, следовательно, для каждой пары (без потери общности Y_1 и Y_2) свойств множество $\{Y_1, Y_2\}$ имеет колоколообразное (рис. III.8, а) двумерное нормальное распределение $f_N(Y_1, Y_2)$. В любом сечении этой поверхности одномерный (так называемый условный) закон распределения нормален, в связи с чем функция изменения дисперсии при изменении первой случайной величины $D\{Y_2\} = \sigma_2^2 h^2(Y_1)$ (или наоборот) известна. В-третьих, гипотетическая зависимость между Y_1 и Y_2 — *линейная*.

Закон распределения $f_N(Y_1, Y_2)$ (рис. III.8, а) кроме средних η_1 и η_2 , а также частных дисперсий σ_1 и σ_2 имеет пятый параметр — $|\rho\{Y_1, Y_2\}| \leq 1$, который характеризует взаимосвязь Y_1 и Y_2 и называется *коэффициентом парной корреляции*. Оценка $r\{Y_1, Y_2\}$

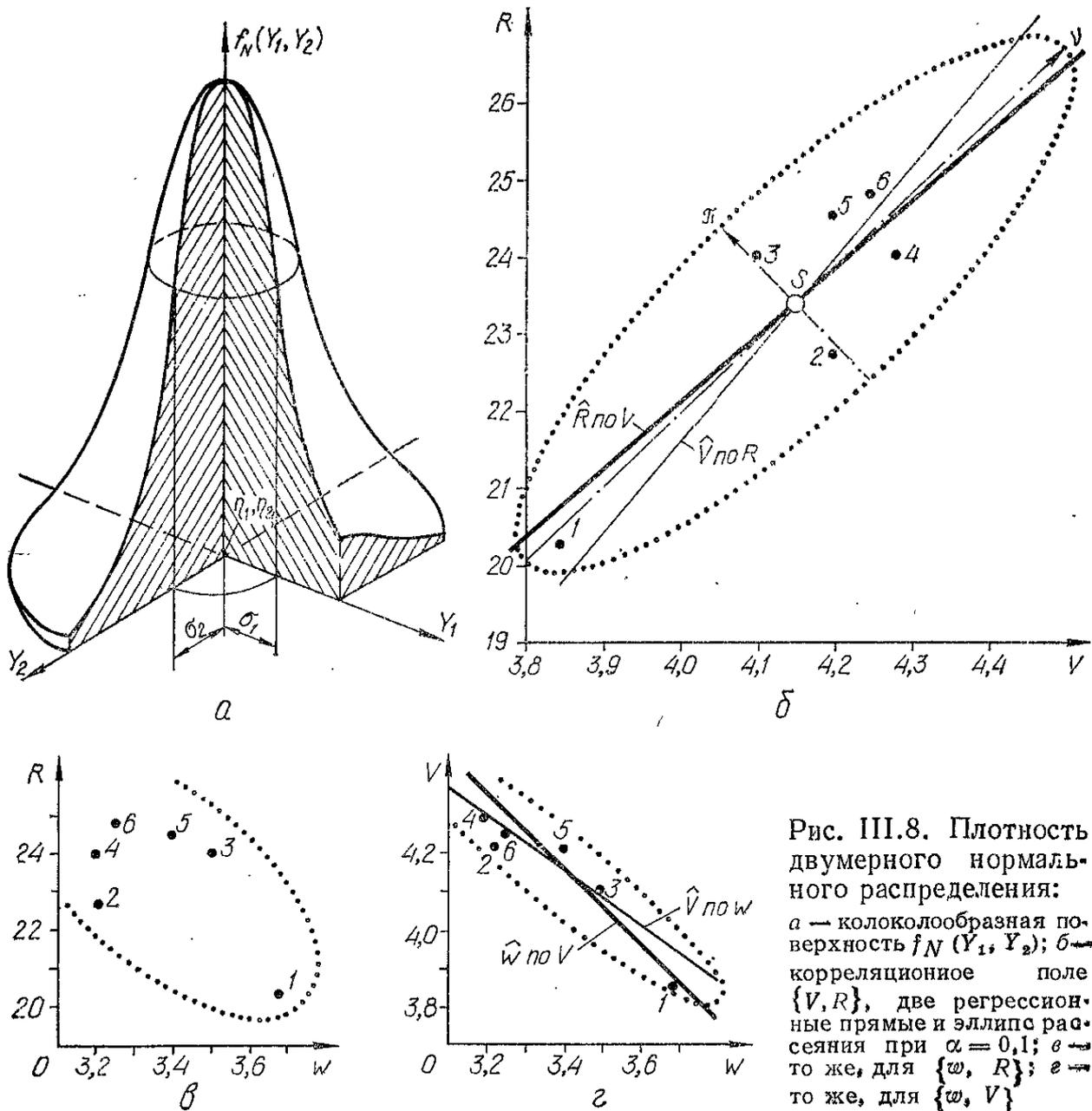


Рис. III.8. Плотность двумерного нормального распределения: а — колоколообразная поверхность $f_N(Y_1, Y_2)$; б — корреляционное поле $\{V, R\}$, две регрессионные прямые и эллипс рассеяния при $\alpha = 0,1$; в — то же, для $\{w, R\}$; г — то же, для $\{w, V\}$

вычисляется через оценку ковариации $\text{cov}^* \{Y_1, Y_2\}$ двумерного распределения (в отраслевых стандартах, например ГОСТ 24544 — 81 и ГОСТ 24545 — 81, по методам испытания бетона ковариация названа корреляционным моментом) — математического ожидания (ср. (III.24)) произведения отклонения Y_1 и Y_2 от своих средних:

$$r \{Y_1, Y_2\} = \text{cov}^* \{Y_1, Y_2\} / s_1, s_2; \quad (\text{III.46})$$

$$\begin{aligned} \text{cov}^* \{Y_1, Y_2\} &= [\sum_{v=1}^{m_u} (y_{1,uv} - \bar{y}_{1,u})(y_{2,uv} - \bar{y}_{2,u})] / (m_u - 1) = \\ &= \frac{1}{m_u - 1} \left[\sum_{v=1}^{m_u} (y_{1,uv} y_{2,uv}) - \frac{1}{m_u} \left(\sum_{v=1}^{m_u} y_{1,uv} \right) \left(\sum_{v=1}^{m_u} y_{2,uv} \right) \right]. \quad (\text{III.47}) \end{aligned}$$

При равенстве $\rho = 0$ взаимосвязь между случайными величинами Y_1 и Y_2 отсутствует, а при $\rho \rightarrow 1$ она приближается к детерминированной функциональной. При этом необязательно, чтобы связь

между Y_1 и Y_2 была причинно-следственной, поскольку оба свойства могут независимо (но параллельно) изменяться под действием общей причины, недоступной для наблюдения технологу в данном исследовании. Знак при $r\{Y_1, Y_2\}$ указывает направление взаимосвязи: при $r > 0$ рост первого показателя должен сопровождаться линейным увеличением второго (и наоборот!), а при $r < 0$ за ростом 1-го показателя должно следовать линейное снижение второго.

Для множества M случайных величин вычисляется оценка $R_{1...M}$ коэффициента множественной корреляции $0 \leq R \leq 1$ всех свойств в i -й точке:

$$1 - R_{1...M}^2 = \left| \begin{array}{cccc} 1 & r_{12} & \dots & r_{1M} \\ r_{21} & 1 & \dots & r_{2M} \\ \dots & \dots & \dots & \dots \\ r_{M1} & r_{M2} & \dots & 1 \end{array} \right| : \left| \begin{array}{cccc} 1 & r_{23} & \dots & r_{2M} \\ r_{32} & 1 & \dots & r_{3M} \\ \dots & \dots & \dots & \dots \\ r_{M2} & r_{M3} & \dots & 1 \end{array} \right|. \quad (\text{III.48})$$

Пример III.12. Результаты испытаний (в условиях пр. III.7) трех свойств бетона представлены в табл. III.8. Кроме предела прочности R_1 , МПа, для каждого из шести образцов измерены скорость ультразвука V , км/с, и водопоглощение ω , % по массе.

Таблица III.8. Результаты эксперимента и расчета сумм преобразованных величин $Z_R = 10(R_1 - 24)$, $Z_V = 100(V - 4,1)$ и $Z_\omega = 100(\omega - 3,4)$

Номер измерения	Эксперимент			Расчет через преобразованные величины								
	R_1	V	ω	Z_R	Z_V	Z_ω	Z_R^2	Z_V^2	Z_ω^2	$Z_R Z_V$	$Z_R Z_\omega$	$Z_V Z_\omega$
1	20,3	3,85	3,68	-37	-25	+28	1369	625	784	+925	-1036	-700
2	22,7	4,21	3,22	-1	+11	-18	169	121	324	-143	+234	-198
3	24,0	4,10	3,50	0	0	+10	0	0	100	0	0	0
4	24,0	4,29	3,20	0	+19	-20	0	361	400	0	0	-380
5	24,5	4,21	3,40	+5	+11	0	25	121	0	+55	0	0
6	24,8	4,25	3,25	+8	+15	-15	64	225	225	+120	-120	-225
Суммы по столбцам				-37	+31	-15	1627	1453	1833	+957	-922	-1503

Оценки средних:

$$\bar{Z}_R = -6,17; \quad \bar{Z}_V = +5,17; \quad \bar{Z}_\omega = -2,5;$$

$$\bar{R}_1 = 23,38 \text{ МПа}; \quad \bar{V} = 4,15 \text{ км/с}; \quad \bar{\omega} = 3,375 \text{ \%}.$$

Оценки среднеквадратичных отклонений:

$$s\{Z_R\} = 16,73; \quad s\{Z_V\} = 16,08; \quad s\{Z_\omega\} = 18,95;$$

$$s\{R_1\} = 1,67 \text{ МПа}; \quad s\{V\} = 0,16 \text{ км/с}; \quad s\{\omega\} = 0,19 \text{ \%}.$$

Оценки ковариаций (достаточен расчет по (III.47) для преобразованных величин)

$$\text{cov}^*\{Z_R, Z_V\} = [957 - (-37) \cdot 31/6]/5 = 229,6;$$

$$\text{cov}^*\{Z_R, Z_\omega\} = -202,9; \quad \text{cov}^*\{Z_V, Z_\omega\} = -285,1.$$

Оценки коэффициентов корреляции (расчет по (III.46) для преобразованных величин дает тождественный результат, так как к \bar{Y}_1 и \bar{Y}_2 централизованы и числитель, и знаменатель)

$$r(R, V) = r\{Z_R, Z_V\} = 229,6 / (16,73 \cdot 16,08) = 0,854;$$

$$r\{R, \omega\} = r\{Z_R, Z_\omega\} = -0,639; \quad r\{V, \omega\} = r\{Z_V, Z_\omega\} = -0,936.$$

Судя по абсолютному значению оценок коэффициента корреляции сила взаимосвязи может быть ранжирована как

$$(V \rightleftharpoons \omega) > (R \rightleftharpoons V) > (R \rightleftharpoons \omega).$$

Оценка коэффициента множественной корреляции проводится для трех случайных величин по формуле, которая следует из соотношения (III.48) при $M = 3$:

$$R = \sqrt{(r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}) / (1 - r_{23}^2)} =$$

$$= \sqrt{(0,854^2 + 0,639^2 - 2 \cdot 0,854 \cdot 0,639 \cdot 0,936) / (1 - 0,936^2)} = 0,972. \quad (\text{III.49})$$

При проектировании на координатную плоскость изолинии $f_N\{Y_1, Y_2\} = p$ колоколообразной поверхности двумерного нормального распределения (рис. III.8, а) будет получен эллипс рассеяния. Риск появления случайной величины (результата эксперимента) вне ограниченного поля составляет $\alpha = 1 - p$. Эллипс рассеяния — важная информационная характеристика степени взаимосвязи Y_1 и Y_2 . Его площадь, диаметры и угол наклона служат (см. п. III.10) численными показателями как поведения объекта, так и организации эксперимента.

Центр эллипса с каноническим уравнением

$$(\nu/\sigma_\nu \sqrt{\chi_\alpha^2})^2 + (\pi/\sigma_\pi \sqrt{\chi_\alpha^2})^2 = 1 \quad (\text{III.50})$$

расположен в точке с координатами \bar{Y}_1 и \bar{Y}_2 . Главные диаметры эллипса $\sigma_\nu \sqrt{\chi_\alpha^2}$ и $\sigma_\pi \sqrt{\chi_\alpha^2}$ лежат на осях ν и π , повернутых к оси Y_1 на угол ψ :

$$\text{tg } 2\psi = 2r_{12} s_1 s_2 / (s_1^2 - s_2^2). \quad (\text{III.51})$$

Критерий χ^2 выбирается по допустимому риску α при $f = 2$: по прил. II.4 $\chi_{0,05}^2 = 5,991$, $\chi_{0,1}^2 = 4,605$ и $\chi_{0,2}^2 = 3,219$; с ростом риска эллипс уменьшается. Стандартные отклонения σ_ν и σ_π вычисляются из системы

$$\left. \begin{aligned} (\sigma_\nu + \sigma_\pi)^2 &= (s_1^2 + s_2^2) + 2s_1 s_2 \sqrt{1 - r_{12}^2}; \\ (\sigma_\nu - \sigma_\pi)^2 &= (s_1^2 + s_2^2) - 2s_1 s_2 \sqrt{1 - r_{12}^2}. \end{aligned} \right\} \quad (\text{III.52})$$

Нормированная площадь эллипса Q_{12} (именованные величины неудобны при сравнительном анализе) определяется как

$$Q_{12} = (s_1/\bar{y}_1)(s_2/\bar{y}_2) \sqrt{1 - r_{12}^2} = v_1 v_2 \sqrt{1 - r_{12}^2}. \quad (\text{III.53})$$

Чем ближе к нулю оценка коэффициента корреляции r , тем угол ψ наклона эллипса меньше, а отношение главных диаметров σ_ν/σ_π ближе к 1.

Блок-схема алгоритма корреляционного анализа (включая описанную ниже оценку коэффициентов линейных моделей) приведена на рис. III.9.

Пример III.13. По данным пр. III.12 построены на рис. III.8, б, в, г эллипсы рассеяния при $\alpha = 0,1$. Для пары $\{R, V\}$ расчет выполнен в следующем порядке:

а) $\bar{R} = 23,38$ МПа и $\bar{V} = 4,15$ км/с — координаты центра;

б) при расчете угла наклона ψ следует учесть соотношение масштабов (на рис. III.8, б — $m_V/m_R = 10$, что дает

$$\begin{aligned} \operatorname{tg} 2\psi &= 2 \cdot 0,854 \cdot 1,67 (0,16 \cdot 10) / [1,67^2 + \\ &+ (0,161 \cdot 10)^2] = 21,67; \\ \operatorname{arctg} (21,67) &= 87,36, \quad \psi = 43^\circ 41'; \end{aligned}$$

в) соотношение $m_V/m_R = 10$ учтено также при расчете σ_v и σ_π :

$$\begin{aligned} (\sigma_v + \sigma_\pi)^2 &= 1,61^2 + 1,67^2 + \\ &+ 2 \cdot 1,61 \cdot 1,67 \sqrt{1 - 0,857^2} = 8,183; \\ (\sigma_v - \sigma_\pi)^2 &= 1,61^2 + 1,67^2 - \\ &- 2 \cdot 1,61 \cdot 1,67 \sqrt{1 - 0,857^2} = 2,584; \end{aligned}$$

$$\left. \begin{aligned} \sigma_v + \sigma_\pi &= 2,86, \\ \sigma_v - \sigma_\pi &= 1,608, \end{aligned} \right\}; \quad \left. \begin{aligned} \sigma_v &= 2,27, \\ \sigma_\pi &= 0,65, \end{aligned} \right\};$$

$$(\sigma_v/\sigma_\pi)_{12} = 3,49;$$

г) главные диаметры эллипса

$$\begin{aligned} \sigma_v \sqrt{\chi_\alpha^2} &= 2,27 \sqrt{4,605} = 4,86; \\ \sigma_\pi \sqrt{\chi_\alpha^2} &= 0,65 \sqrt{4,605} = 1,4; \end{aligned}$$

д) площадь эллипса рассеяния для R и V (при v_i в %)

$$Q_{12} = (1,67/23,38)(0,161 \cdot 4,15) \cdot 10^4 \sqrt{1 - 0,865^2} = 14,24.$$

Аналогичные расчеты для пары свойств $\{R, \omega\}$ и $\{V, \omega\}$ дают соотношения главных диаметров и площади эллипсов:

для $\{R, \omega\}$

$$(\sigma_v/\sigma_\pi)_{13} = 2,29/1,06 = 2,16, \quad Q_{13} = 30,91;$$

для $\{V, \omega\}$

$$(\sigma_v/\sigma_\pi)_{23} = 2,47/0,44 = 5,61, \quad Q_{23} = 7,67.$$

Полученный в пр. III.12 вывод о силе связи внутри пар свойств подтверждается и ранжированием величин (σ_v/σ_π) и Q_{ij} :

$$\begin{aligned} (\sigma_v/\sigma_\pi)_{23} &> (\sigma_v/\sigma_\pi)_{12} > (\sigma_v/\sigma_\pi)_{13}, \\ Q_{23} &< Q_{12} < Q_{13}. \end{aligned}$$

При анализе корреляционной связи величин, распределенных по двумерному нормальному закону, необходима проверка стати-

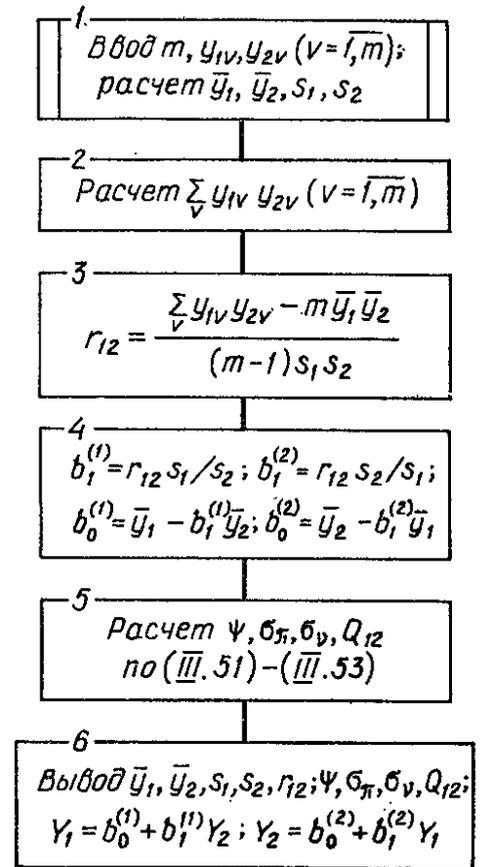


Рис. III. 9. Блок-схема алгоритма корреляционного анализа

стической гипотезы о равенстве истинного коэффициента корреляции ρ нулю, т. е. о взаимонезависимости $y_{1.u}$ и $y_{2.u}$. Нуль-гипотеза $H_0: \rho = 0$ принимается с риском α , если полученная по экспериментальным значениям оценка r_{12} по абсолютному значению меньше критического уровня $r_{кр}$ случайной величины r , распределение которой для числа степеней свободы $f = m - 2$ приведено в прил. II.8. Принятие нуль-гипотезы означает, что линейная связь между показателями $y_{1.u}$ и $y_{2.u}$ не обнаружена, а отклонение H_0 говорит о том, что между $y_{1.u}$ и $y_{2.u}$ линейная связь существует с достоверностью $(1 - \alpha)$.

Если нуль-гипотеза отклонена, то можно построить две линейные однофакторные корреляционные модели: одну для оценки второго показателя по данным о первом, другую — для оценки первого по информации о втором:

$$\hat{Y}_2 - \bar{y}_2 = r_{12}(s_2/s_1)(Y_1 - \bar{y}_1); \quad (\text{III.54})$$

$$\hat{Y}_1 - \bar{y}_1 = r_{12}(s_1/s_2)(Y_2 - \bar{y}_2). \quad (\text{III.55})$$

При этом второе уравнение не является алгебраическим преобразованием первого, за исключением случая, когда $r_{12} = r_{12}^{-1} = 1$ (функциональная связь). Прямые (III.54) и (III.55) называются уравнениями регрессии Y_2 по Y_1 и регрессии Y_1 по Y_2 соответственно (более подробно см. п. III.7).

Пример III.14. По данным пр. III.12 проверка нуль-гипотезы $H_0: \rho = 0$ при риске $\alpha = 0,1$ ($m = 6$, $f = 6 - 2 = 4$, $r_{кр}\{0,1\} = 0,729$) приводит к выводам:

а) линейная связь между R и V правдоподобна ($r_{12} = 0,857 > r_{кр}$ и нуль-гипотеза отклонена); риск отклонить правильную гипотезу (при дополнительном анализе по прил. II.8) $0,02 < \alpha < 0,05$;

б) линейная связь между R и ω не обнаружена ($|r_{13}| = |-0,639| < r_{кр}$ и нуль-гипотеза принимается);

в) линейная связь между V и ω правдоподобна ($|r_{23}| = |-0,936| > r_{кр}$) с риском $0,001 < \alpha < 0,01$.

Можно построить пары корреляционных моделей:
для варианта «а»

$$\left. \begin{aligned} \hat{R} - 23,38 &= 0,854(1,67/0,16)(V - 4,15), \\ \hat{R} &= -13,61 + 8,21V; \end{aligned} \right\} \quad (\text{III.56})$$

$$\hat{V} = 2,24 + 0,082 R; \quad (\text{III.57})$$

для варианта «в»

$$\left. \begin{aligned} \hat{\omega} &= 7,96 - 1,10V, \\ \hat{V} &= 6,83 - 0,79\omega. \end{aligned} \right\} \quad (\text{III.58})$$

Для пары $\{R, \omega\}$ модель смысла не имеет.

Прямые (III.56) — (III.58) нанесены на эллипсах рассеяния на рис. III,8, б и г. Если в нарушение смысла корреляционного анализа рассчитать $V = \varphi(R)$ алгебраически из (III.56), т. е. получить $\hat{V} = 1,53 + 0,112 R$, то будет получено некорректное решение.

Вся вышеизложенная процедура корреляционного анализа применима и в том случае, когда в одной точке факторного пространства опыты воспроизводятся целиком (полностью повторяются стро-

ки спектра плана в информационной табл. III.1). При этом в расчет можно вводить средние оценки $\bar{Y}_j (j = \overline{1, M})$ по каждому воспроизведенному опыту.

III.5. Эксперимент в одной точке факторного пространства (большая выборка)

При числе параллельных испытаний в u -й точке факторного пространства $m_u \geq 50$ наблюдается ряд методических особенностей по сравнению с малой выборкой (п. III.3, III.4). Во-первых, информация, как правило, представляется не в виде ряда из m индивидуальных значений y_{uv} (в частности, ранжированных рядов (III.31)), а в виде группировок, для каждой из которых известны границы каждого j -го интервала и количество попавших в него индивидуумов m_j . Такая форма представления информации связана не только с вычислительными преимуществами, но и нередко с самой сутью работы измерительного оборудования, использующего принципы дискретизации, или квантования (лат. *quantum* — сколько), шкал — приборы акустической эмиссии, лазерной порометрии и др. Во-вторых, по мере увеличения объема выборки m_u (в частности, из нормально распределенной генеральной совокупности) распределение состоятельных и несмещенных оценок Θ^* приближается к нормальному, поэтому жесткость требований к использованию ряда специальных распределений (t , F , χ^2 и др.) практически снимается. В-третьих, анализ больших выборок позволяет гипотезу о распределении генеральной совокупности по закону, близкому к нормальному (которую приходится принимать в малых выборках без специальной проверки) рассматривать не как единственную, а выдвинуть веер альтернативных гипотез и выбрать по экспериментальным данным наиболее точную по отношению к данному объекту функцию распределения его состояний, что значительно уменьшает риск от принятия неверных инженерных решений.

При группировке результатов m параллельных испытаний, изменяющихся с размахом $y_{m(\max)} - y_{1(\min)}$ (III.31), общее число K интервалов выбирается в пределах 8 — 40 (ГОСТ 8735 — 75, 8269 — 76 по методам испытания заполнителей рекомендуют $K_{\min} = 8 \dots 10$). Целесообразно, чтобы длина интервалов l_y была постоянной.

Окончательное решение о l_y и K , а также о положении на шкале Y середины интервалов y_i выбирается, исходя из удобства работы. При анализе распределений показателей качества продукции и технико-экономических показателей, включаемых в нормативно-техническую документацию, число интервалов K выбирается в зависимости от объема выборки: $K = 18 \dots 20$ при $m = 200$, $K = 25 \dots 30$ при $m = 400$, $K = 35 \dots 40$ при $m = 1000$.

При подсчете частот m_j попадания случайных величин в j -ю группировку учитывается, что величина, попавшая на границу

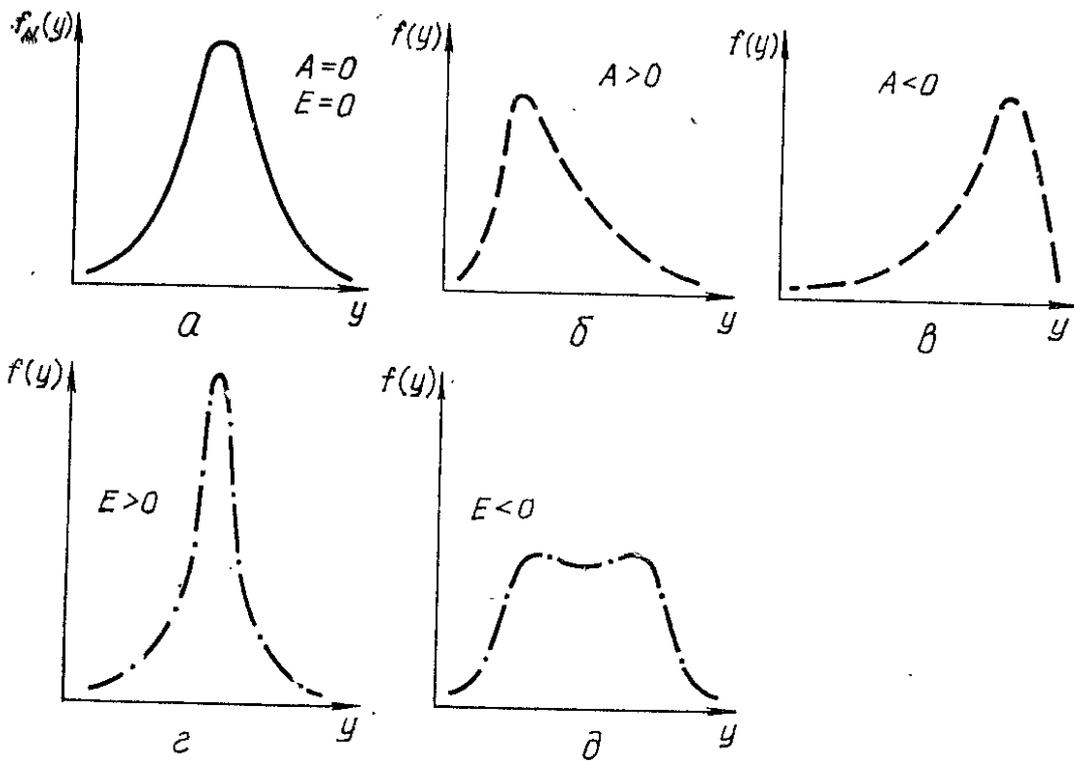


Рис. III.10. Формы кривых плотности вероятности:

a — нормальная $A = 0, E = 0$; *б* — положительная асимметрия; *в* — отрицательная асимметрия; *г* — вытянутая при положительном эксцессе; *д* — упрощенная при отрицательном эксцессе

интервала $\overset{\circ}{q}_j \pm l_y/2$, рассматривается как $1/2$ точки и входит в оба смежных интервала.

Оценки параметров распределения при $m \geq 50$ предпочтительно определять по методу моментов. Рассчитываются оценки четырех первых начальных $m_s^* \{y\}$ и центральных $\mu_s^* \{y\}$ моментов порядка $s (s = \overline{1,4})$ с учетом частот $\overset{\circ}{m}_j$ (иногда их называют весами, а расчет с их учетом — взвешиванием оценок):

$$m_s^* \{y\} = \frac{1}{m} \sum_{j=1}^K \overset{\circ}{m}_j \overset{\circ}{y}_j^s; \quad (\text{III.59})$$

$$\mu_s^* \{y\} = \frac{1}{m} \sum_{j=1}^K \overset{\circ}{m}_j (\overset{\circ}{y}_j - \bar{y})^s; \quad (\text{III.60})$$

$$\mu_2^* = m_2^* - (m_1^*)^2; \quad (\text{III.61})$$

$$\mu_3^* = m_3^* - 3m_2^* m_1^* + 2(m_1^*)^3; \quad (\text{III.62})$$

$$\mu_4^* = m_4^* - 4m_3^* m_1^* + 6m_2^* (m_1^*)^2 - 3(m_1^*)^4. \quad (\text{III.63})$$

Далее рассчитываются оценки среднего \bar{Y} и дисперсии s^2 , которые являются соответственно первым начальным и вторым центральным моментами распределения

$$\bar{Y} = m_1^*; \quad s^2 = \mu_2^*, \quad (\text{III.64})$$

а также оценки параметров формы кривой (рис. III.10): коэффициента асимметрии распределения A^* и эксцесса E^* , характеризу-

ющего степень островершинности распределения по сравнению с нормальным (рис. III.10), где $\mu_4^*/(\mu_2^*)^2 = 3$;

$$A^* = \mu_3^*/(\sqrt{\mu_2^*})^3; E^* = \mu_4^*/(\mu_2^*)^2 - 3. \quad (\text{III.65})$$

В особо точных расчетах в дополнение к (III.64), (III.65) могут вводиться поправки на интервал группирования [10, 20 и др.].

Пример III.15. Для анализа качества продукции взята большая выборка $m = 537$ паспортной активности цемента [2]: от $R_{\text{мин}} = 29,5$ МПа до $R_{537\text{мак}} = 52,4$ МПа. Согласно нормативам, для такого m рекомендуется 20 — 25 интервалов; целесообразно принять $K = 24$, $l_y = 1$ МПа, а середину интервалов совместить с целыми числами.

Таблица III.9. К расчету оценок числовых характеристик распределения при большой выборке

$\overset{\circ}{R}_j$, МПа	$\overset{\circ}{m}_j$	Z_j	Z_j^2	Z_j^3	Z_j^4	$\overset{\circ}{m}_j Z_j$	$\overset{\circ}{m}_j Z_j^2$	$\overset{\circ}{m}_j Z_j^3$	$\overset{\circ}{m}_j Z_j^4$
29	2	-11	121	-1331	14641	-22	242	-2662	29282
30	2	-10	100	-1000	10000	-20	200	-2000	20000
31	2	-9	81	-729	6561	-18	162	-1458	13122
32	8	-8	64	-512	4096	-64	512	-4096	32768
33	3	-7	49	-343	2401	-21	147	-1029	7203
34	19	-6	36	-216	1296	-114	684	-4104	24624
35	16	-5	25	-125	625	-80	400	-2000	10000
36	22	-4	16	-64	256	-88	352	-1408	5632
37	37	-3	9	-27	81	-111	333	-999	2997
38	57	-2	4	-8	16	-114	228	-456	912
39	63	-1	1	-1	1	-63	63	-63	63
40	64	0	0	0	0	0	0	0	0
41	51	+1	1	1	1	51	51	51	51
42	51	+2	4	8	16	102	204	408	816
43	39	+3	9	27	81	117	351	1053	3159
44	25	+4	16	64	256	100	400	1600	6400
45	21	+5	25	125	625	105	525	2625	13125
46	22	+6	36	216	1296	132	792	4752	28512
47	17	+7	49	343	2401	119	833	5831	40817
48	8	+8	64	512	4096	64	512	4096	32768
49	2	+9	81	729	6561	18	162	1458	13122
50	2	+10	100	1000	10000	20	200	2000	20000
51	2	+11	121	1331	14641	22	242	2662	29281
52	2	+12	144	1728	20736	24	288	3456	41472
Суммы						159	7883	9717	376127
m_s^*						0,30	14,68	18,09	700,42

Результаты группирования (середины интервалов $\overset{\circ}{R}_j$, частоты $\overset{\circ}{m}_j$, преобразованные величины $Z_j = R - 40$ и т. д.) показаны в табл. III.9; там же приведены оценки четырех начальных моментов, с помощью которых рассчитываются оценки μ_s^* , \bar{R} , s^2 , s , ν , A^* , E^* ;

$$\begin{aligned}\bar{R} &= m_1^* \{Z\} - 40 = 0,3 + 40 = 40,3 \text{ МПа}; \\ s^2 \{Z\} &= s^2 \{R\} = \mu_2^* = 14,68 - (0,3)^2 = 14,59; \\ s &= \sqrt{s^2} = \sqrt{14,59} = 3,82 \text{ МПа}; \quad v = s/R = 3,82/40,3 = 0,0948; \\ \mu_3^* &= 18,09 - 3 \cdot 14,68 \cdot 0,3 + 2 \cdot 0,3^3 = 508; \\ \mu_4^* &= 700,42 - 4 \cdot 18,1 \cdot 0,3 + 6 \cdot 14,68 \cdot 0,3^2 - 3 \cdot 0,3^4 = 686,7; \\ A^* \{Z\} &= A^* \{R\} = 5,08/3,82^3 = 0,09; \\ E^* \{Z\} &= E^* \{R\} = 686,7/14,59^2 - 3 = 0,22.\end{aligned}$$

По данным табл. III.9 определяются и непараметрические числовые характеристики по аналогии с пр. III.8, но с учетом интервалов группирования.

Размах $W = 52,4 - 29,3 = 23,1$ МПа. Частота появления случайной величины в заданном диапазоне, например $p\{R = 50 \pm l_y/2\} = 2/537 = 0,0037$, а $p\{R \leq 40\} = 263/537 = 0,4898$ (в последнем случае учтена половина частоты $m_{40} = 64$ в силу гипотезы о равномерном распределении случайных величин внутри интервала [2]). Эмпирический квантиль Q_p для заданной частоты p вычисляется на основе информации, связанной с интервалом $\overset{\circ}{I}\{Q_p\}$, содержащим $Q_p =$ квантиль:

$$Q_p = (\overset{\circ}{y}_Q - l_y/2) + (pm - m_{\Sigma Q}^{\circ}) l_y / m_Q^{\circ}, \quad (\text{III.66})$$

где $\overset{\circ}{y}_Q$ — середина интервала $\overset{\circ}{I}\{Q_p\}$; m_Q° — частота в интервале $\overset{\circ}{I}\{Q_p\}$; $m_{\Sigma Q}^{\circ}$ — накопленная частота до интервала $\overset{\circ}{I}\{Q_p\}$.

Например, заданной частоты $p = 0,1$ соответствует член ранжированного ряда $pm = 0,1 \cdot 537 = 53,7$. При суммировании от первой группы $2 + 2_0 + \dots + 19 + 16 = 52$ определяется накопленная частота до интервала $\overset{\circ}{I}\{Q_p\}$, а следовательно, и сам интервал $\overset{\circ}{I}\{Q_p\}$, где $\overset{\circ}{y}_Q = 36$ МПа и $m_Q^{\circ} = 22$. С учетом $l_y = 1$ МПа

$$Q_p = (36 - 0,5) + (53,7 - 52) \cdot 1/22 = 35,5 + 0,08 = 35,58 \text{ МПа}.$$

Медиана, т. е. $Q_{0,5}$, в данном примере

$$Me = (40 - 0,5) + (268,5 - 231) \cdot 1/64 = 40,09 \text{ МПа}$$

и практически совпадает с \bar{Y} , что говорит о симметрии распределения.

Доверительные интервалы для истинных значений числовых характеристик Θ при больших выборках можно (за исключением специально оговариваемых ситуаций) приближенно рассчитывать на основе нормального распределения, квантиль которого ϵ_N зависит только от допустимого риска α (при $f \rightarrow \infty$ распределения t , χ^2 и другие совпадают с нормальным):

$$P\{\Theta^* - \epsilon_N s\{\Theta^*\} \leq \Theta < \Theta^* + \epsilon_N s\{\Theta^*\}\} = 1 - \alpha, \quad (\text{III.67})$$

где для двухстороннего риска $\epsilon_N\{\alpha = 0,2\} = 1,282$; $\epsilon_N\{\alpha = 0,1\} = 1,645$ и $\epsilon_N\{\alpha = 0,05\} = 1,96$ (см. прил. II.1).

Формулы для расчета оценок среднеквадратических отклонений $s\{\Theta^*\}$, а также оценок основных параметрических и непараметрических числовых характеристик (см. п. III.3) при больших m приведены в табл. III.10. Некоторые из оценок $s\{\Theta^*\}$ оказываются доста-

Таблица III.10. Формулы для расчета оценок среднеквадратических отклонений $s\{\theta^*\}$ оценок основных числовых характеристик при больших информационных массивах и пример определения для них доверительного интервала

Числовая характеристика θ	Среднеквадратическое отклонение $s\{\theta^*\}$	Расчет $s\{\theta^*\}$	θ^*	Доверительный интервал при $\alpha=0,1$ и $\varepsilon_N=1,645$
Среднее η	s/\sqrt{m}	$3,82/\sqrt{537} = 0,168$	40,3	$40,03 \leq \eta < 40,57$
Среднеквадратическое отклонение σ	$s/\sqrt{2m}$	$3,82/\sqrt{2 \cdot 537} = 0,117$	3,82	$3,63 \leq \sigma < 4,01$
Коэффициент вариации γ	$v \sqrt{\frac{1+2v^2}{2m}} \approx v/\sqrt{2m}$	$0,095/\sqrt{2 \cdot 537} = 0,003$	0,095	$0,09 \leq \gamma < 0,1$
Коэффициент асимметрии A	$[6(m-2)(m+1)^{-1}(m+3)]^{0,5} \approx \sqrt{6/m}$	$\sqrt{6/537} = 0,106$	0,09	$-0,08 \leq A < +0,26$
Экцесс E	$[24m(m-2)(m-3)(m+1)^{-2}(m+3)(m+5)]^{0,5} \approx \sqrt{24/m}$	$\sqrt{24/537} = 0,211$	0,22	$-0,13 \leq E < +0,57$
Частость p	$\sqrt{\frac{p(1-p)}{m}}$, если $p(1-p) \geq 0,05$	$\sqrt{\frac{0,49 \cdot 0,51}{537}} = 0,022$	0,49	$0,454 \leq p < 0,525$
Квантиль $\eta\{Q_p\}$	$\frac{\sqrt{p(1-p)}}{\Phi_N\{\varepsilon_p\}} \frac{s}{\sqrt{m}}$	$\frac{\sqrt{0,1 \cdot 0,9}}{0,103 \sqrt{537}} 3,82 = 0,48$	35,58	$34,79 \leq \eta\{Q_p\} < 36,37$
Медиана $\eta\{Me\}$	$1,2533 \frac{s}{\sqrt{m}}$	$1,2533 \frac{3,82}{\sqrt{537}} = 0,21$	40,09	$39,75 \leq \eta\{Me\} > 40,43$

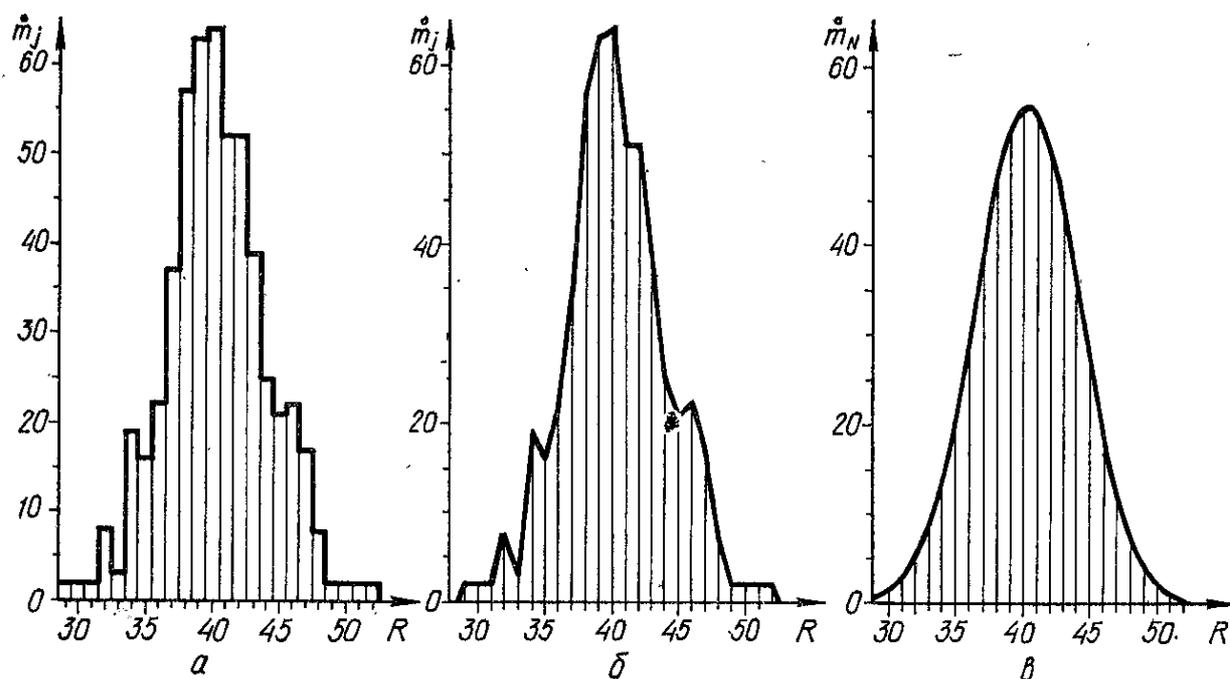


Рис. III.11. Гистограмма (а), полигон (б) и кривая нормального распределения (в) частот активности цемента

точно эффективными только при условии распределения генеральной совокупности по нормальному закону. При расчете $s\{Q_p\}$ ордината плотности нормального распределения $\varphi_N\{\epsilon_p\}$ (прил. II.1) относится к середине соответствующего интервала группирования; коэффициент в формуле для медианы, как частного случая квантиля при $p = 1/2$, получен при $\varphi_N\{\epsilon_{0,5} = 0\} = 0,3989$; для малых частот p^* (редкие события) нужно использовать специальные рекомендации, основанные на более сложных предпосылках [17].

Пример III.16. Для числовых характеристик Θ^* , определенных в пр. III.15, в табл. III.10 даны расчеты оценок среднеквадратических отклонений $s\{\Theta^*\}$ и показаны доверительные интервалы (III.67) при двухстороннем риске $\alpha = 0,1$.

Следует отметить, что на основе анализа совокупности из 537 результатов испытания цемента оценки получены с различной относительной точностью (для среднего эта точность $s\{\Theta^*\}/\Theta^* = 0,004$ или 0,4 %, для среднеквадратического отклонения и коэффициента вариации — 3,1 %, для частот — 4,5 % и т. п.).

Доверительные интервалы для коэффициента асимметрии и эксцесса накрывают ноль, следовательно, нужно допустить нуль-гипотезы $H_0 : A = 0$ и $H_0 : E = 0$ и признать, что распределение активности цемента R не отличается (на данном заводе) от нормального закона.

Графически (рис. III.11) эмпирические распределения частот отражаются гистограммами (в пределах каждого диапазона проводится прямая на соответствующем уровне шкалы m_j) или полигонами (ломаная с углами в середине каждого диапазона в точке на уровне m_j). Следует отметить, что при $n < 100$ форма гистограммы существенно изменяется в зависимости от длины интервалов группирования (или от числа интервалов). Однако анализ общего вида графиков совместно с информацией о параметрах A и E

позволяет выбирать [2, 10] по каталогу близкий *аппроксимирующий* закон *распределения*, для которого с учетом эмпирических данных строится непрерывная кривая по табулированным нормированным значениям случайной величины.

Пример III.17. Для частот в табл. III.9 построены на рис. III.11 гистограмма и тождественный ей полигон. В пр. III.16 показано, что нуль-гипотезы $H_0: A = 0$ и $H_0: E = 0$, соответствующие нормальному распределению, не отклоняются. Алгоритм построения ординат аппроксимирующей кривой $\varphi_N\{R\}$ реализован в табл. III.11: в столбце 4 учтено $s\{R\} = 3,82$ МПа; соответствующие нормированным величинам (столбец 5) значения функции $\varphi_N\{\varepsilon_R\}$ взяты по прил. II.1; переход к теоретическим частотам (столбец 6) осуществлен при умножении $\varphi_N\{\varepsilon_R\}$ на величину $(ml_y)/s\{R\} = 531 \cdot 1/3,82 = 140,6$. Кривая распределения $\varphi_N\{R\}$ показана на рис. III.11, в.

Окончательное решение о выборе того или иного теоретического аппроксимирующего закона распределения $\varphi_T\{Y\}$ можно сделать после проверки нуль-гипотезы $H_0: \varphi_T\{Y\} = \varphi_{\varepsilon}\{Y\}$ — об отсутствии значимых расхождений между этим законом (в частности, нормальным $\varphi_N\{Y\}$) и законом распределения случайных величин $\varphi_{\varepsilon}\{Y\}$ той генеральной совокупности, о которой собрана экспериментальная информация. Проверка гипотезы, называемой гипотезой согласия опытного и теоретического распределений случайной величины, ведется по стандартной процедуре с помощью критериев Колмогорова (λ), а также χ^2 , ω^2 и ряда других нестандартных критериев, в том числе непараметрических [10].

При $m > 100$ рекомендуются критерии Колмогорова и χ^2 , которые для выборки из нормально распределенной совокупности обладают примерно равной эффективностью.

Критерий χ^2 вычисляется на основе информации об эмпирических m_j и теоретических $m_{T,j}^*$ частотах, полученных для середин соответствующих интервалов \hat{Y}_j для аппроксимирующего закона

$$\chi^2 = \sum_{j=1}^K (m_j - m_{T,j}^*)^2 / m_{T,j}^*. \quad (\text{III.68})$$

Критерий (III.68) применим к группировкам с частотой более 10, поэтому в расчет вводятся значения частот $m_{T,j}^*$ в объединенных интервалах, где было $m_{T,j} < 10$. Если при числе степеней свободы $f = K - N\{\Theta\} - 1$ критерий χ^2 для закона распределения, имеющего $N\{\Theta\}$ параметров (у нормального закона $N\{\Theta\} = 2$ — среднее η и дисперсия σ^2), соответствует риску $\alpha > 0,05$, то гипотеза $H_0: \varphi_N\{Y\} = \varphi_T\{Y\}$ допускается как правдоподобная.

Критерий Колмогорова вычисляется по максимальной абсолютной разности D_{\max} между эмпирическими накопленными частотами $p_{\Sigma,j}$ в каждом j -м интервале и теоретической функцией распределения $F_T\{\varepsilon_B\}$, рассчитанной для верхней границы $y_{j,B} = y_j + l/2$ этого интервала:

$$\lambda = D_{\max} \sqrt{m} = \max |p_{\Sigma,j} - F_T\{\varepsilon_B\}| \sqrt{m}. \quad (\text{III.69})$$

Таблица III.11. К расчету кривой нормального распределения и критериев согласия

Опытные данные		Расчет ординат кривой нормального распределения				Расчет критерия χ^2		Расчет критерия Колмогорова λ					
$\overset{\circ}{R}_j$	$\overset{\circ}{m}_j$	$\Delta_j = \overset{\circ}{R}_j - \bar{R}$	$\varepsilon_R = \overset{\circ}{R}_j / s$	$\Phi_N \{ \varepsilon_R \}$	$\overset{\circ}{m}_N^*$	$ \overset{\circ}{m}_j - \overset{\circ}{m}_N^* $	$\frac{(\overset{\circ}{m}_j - \overset{\circ}{m}_N^*)^2}{\overset{\circ}{m}_N^*}$	$\varepsilon_{R.B} = \frac{\Delta_j + l_y/2}{s}$	$F_N \{ \varepsilon_{R.B} \}, \%$	$\overset{\circ}{m}_{\Sigma,j}$	$\frac{p_{\Sigma,j} = \overset{\circ}{m}_{\Sigma,j}}{\overset{\circ}{m}} \cdot 100, \%$	$D, \%$	
1	2	3	4	5	6	7	8	9	10	11	12	13	
29	2	-11,3	-2,96	0,005	0,7	} 2,4	0,29	-2,83	0,2	2	0,4	0,2	
30	2	-10,3	-2,70	0,010	1,4			-2,57	0,5	4	0,8	0,3	
31	2	-9,3	-2,43	0,021	3,0			-2,30	1,0	6	1,1	0,1	
32	8	-8,3	-2,17	0,038	5,3			-2,04	2,1	14	2,6	0,5	
33	3	-7,3	-1,91	0,064	9,0			-1,78	3,8	17	3,2	0,6	
34	19	-6,3	-1,65	0,102	14,3			4,7	1,54	6,4	36	6,7	0,3
35	16	-5,3	-1,39	0,152	21,4			5,4	1,36	10,4	52	9,7	0,7
36	22	-4,3	-1,13	0,211	29,7			7,7	2,00	15,9	74	13,8	2,1
37	37	-3,3	-0,86	0,276	38,8			1,8	0,08	23,3	111	20,7	2,6
38	57	-2,3	-0,60	0,333	46,8			10,2	2,22	31,9	168	31,3	0,6
39	63	-1,3	-0,34	0,377	53,0	10,0	1,89	41,7	231	43,0	1,3		
40	64	-0,3	-0,08	0,398	55,9	8,1	1,17	52,0	295	54,9	2,9		
41	51	+0,7	+0,18	0,393	55,2	4,2	0,32	62,2	346	64,4	2,2		
42	51	+1,7	+0,45	0,360	50,6	0,4	0,00	71,9	397	73,9	2,0		
43	39	+2,7	+0,71	0,310	43,6	4,6	0,49	80,0	436	81,2	1,2		
44	25	+3,7	+0,97	0,249	35,0	10,0	2,86	86,4	461	85,8	0,6		
45	21	+4,7	+1,23	0,187	26,3	5,3	1,07	91,3	482	89,8	1,5		
46	22	+5,7	+1,49	0,131	18,4	3,6	0,70	94,7	504	93,8	0,9		
47	17	+6,7	+1,75	0,086	12,1	4,9	0,40	97,0	521	97,0	0,0		
48	8	+7,7	+2,02	0,052	7,3	} 0,6	0,02	+1,88	97,0	521	97,0	0,0	
49	2	+8,7	+2,28	0,030	4,2			+2,15	98,4	529	98,5	0,1	
50	2	+9,7	+2,54	0,016	2,2			+2,41	99,2	531	98,9	0,3	
51	2	+10,7	+2,80	0,008	1,1			+2,67	99,6	533	99,3	0,3	
52	2	+11,7	+3,06	0,004	0,6			+2,93	99,8	535	99,6	0,2	
								+3,19	99,9	537	100,0	0,1	
								$\chi^2 = 16,41$					$D_{\max} = 2,9\%$

Для значения λ по прил. II.9 определяется риск α ; если $1 - \alpha \geq 0,8$, то нуль-гипотеза о согласии опытного распределения с теоретическим не отклоняется.

Пример III.18. Для данных пр. III.17 определить критерий χ^2 весьма просто, используя столбец 6 в табл. III.11. Крайние частоты объединены в столбце 7 так, чтобы выполнялось условие $\overset{\circ}{m}^* \geq 10$. Сумма по столбцу 8 — $\chi^2 = 16,41$, которое при $K = 16$ (учтены объединенные интервалы!) дает $0,3 > \alpha > 0,2$ при $f = 16 - 2 - 1$ (прил. II.9), что позволяет принять гипотезу о нормальном распределении величины R .

При использовании критерия λ сначала для верхней границы интервала находится нормированная случайная величина $\varepsilon_{R_{в}}$ (столбец 9), которая используется для определения значения $F_N\{\varepsilon_{R_{в}}\}$ (столбец 10) по прил. II.2, причем для отрицательных значений $\varepsilon_{R_{в}}$ учитывается, что $F_N\{\varepsilon_{R_{в}} < 0\} = 1 - F_N\{|\varepsilon_{R_{в}}|\}$. Накопленные частоты $\overset{\circ}{m}_{\Sigma, j}$ (столбец 11) вычисляются по данным $\overset{\circ}{m}_j$ (столбец 2) и нормируются к m (столбец 12). По максимальной разности $D_{\max} = 2,9\% = 0,029$ (в интервале $\overset{\circ}{R}_j = 40$ МПа) определяется критерий

$$\lambda = 0,029 \sqrt{537} = 0,672,$$

который соответствует $0,3 > \alpha > 0,2$ (прил. II.9), следовательно гипотеза о нормальности распределения активности цемента на данном заводе принимается как правдоподобная.

Свертка эмпирической информации о проведении системы в u -й точке факторного пространства в виде определенной (по виду и по численным значениям параметров) *плотности распределения* $\varphi_{\varepsilon}(Y)$ является наиболее *высокой ступенью* анализа опытных данных в одной точке факторного пространства. На основе $\varphi_{\varepsilon}(Y)$ можно по аналогии с методикой оценки доверительных интервалов (III.36) решить большинство задач, связанных с оценкой вероятности появления случайной величины показателя системы внутри (вне) заданных границ (задачи управления объектом, контроля его качества и надежности).

III.6. Сравнение результатов эксперимента в двух и нескольких точках факторного пространства

В информационной таблице о поведении объекта, как уже указывалось в п. III.1, любая пара опытов — u -й и $(u + 1)$ -й — раздвинута в k -факторном пространстве на некоторое расстояние в результате изменения уровней одного или нескольких факторов X_i . Эти изменения в уровнях могут быть измерены [4, 9, 10 и др.] как на *количественных*, или *метрических*, шкалах (имеют единицу измерения МПа, кг/м³, руб. и т. п.), так и на шкалах более низкого уровня упорядочения. Для *ранговых*, или *порядковых*, шкал известно, что « A больше B », а « B больше C » и т. п., но не определено, «насколько A больше B » и «насколько отличие между A и B существеннее, чем отличие между B и C » и т. п. Несмотря на ограниченность

ранговые шкалы приносят огромную пользу в науке и технике (баллы за эстетические и органолептические качества продукции, оценки педагогические и спортивные, категории специалистов и т. п.). Простейшие шкалы — *номинальные*, или *классификационные*, позволяют лишь разделить объекты на классы «*A* не есть *B*», «*B* не есть *C*» и т. п. без указания их порядка (можно присвоить кодовые номера объектам *A*, *B*, *C*, ..., например студенческим группам на потоке, но это не содержит никакой количественной информации).

Если состояние системы изменяется вдоль метрических шкал *X*, то, изучив в ряде точек отклики Y_j , можно, как правило, их изменение ΔY_j в зависимости от ΔX описать непрерывными функциями $Y_j = \varphi_j(X)$, построив комплекс моделей (см. п. III. 7 и далее). В некоторых ситуациях такие модели можно получать и для порядковых шкал (хотя и с заведомо невысокой точностью конечных инженерных выводов).

Если же изменение состояния системы (по входам) удастся отметить лишь на классификационной шкале, то непрерывные функции как модели непригодны. При этом задачи оценки изменения откликов $Y_{j,u}$, $Y_{j,(u+1)}$, ..., $Y_{j,(u+w)}$, ... сводятся к задаче сравнения; естественно, что ее постановка правомерна и для шкал более высокого уровня.

Сравнение результатов в двух точках факторного пространства *A* (оценки обозначены Θ_1^* без потери общности) и *B* (оценки Θ_2^*) проводится с помощью нуль-гипотез $H_0: \Theta_1 = \Theta_2$ при альтернативе $H_0: \Theta_1 \neq \Theta_2$ (или $H_1: \Theta_1 > \Theta_2$, если договориться далее считать Θ_1 всегда больше чем Θ_2).

Анализ начинается со *сравнения дисперсий* при гипотезе $H_0: \sigma_1^2 = \sigma_2^2$, которая проверяется по *F-критерию Фишера*:

$$\hat{F} = s_1^2/s_2^2 \text{ (при } s_1^2 \geq s_2^2 \text{ критерий } F \geq 1). \quad (\text{III.70})$$

Если \hat{F} меньше чем F_τ при риске α (обычно 0,05) и числе степеней свободы в числителе $f_1 = m_1 - 1$ и в знаменателе $f_2 = m_2 - 1$ (см. прил. II.6), то гипотеза $H_0: \sigma_1^2 = \sigma_2^2$ допускается и можно считать, что изменения рассеяния результатов *Y* при переходе от точки *A* к точке *B* не происходит — *результаты однородны*. Это может быть очень хорошим инженерным результатом, если технолог заменяет основное сырье на вторичное, например вводит в вяжущее золу-унос. Однако это может быть весьма неприятным выводом, если за счет смены смесителя *A* на смеситель *B* технолог хотел повысить однородность смеси!

Если \hat{F} больше чем F_τ , то при $q = 1 - \alpha$ допускается гипотеза $H_0: \sigma_1^2 > \sigma_2^2$ и, следовательно, рассеяние результатов в первой точке считается больше, чем во второй.

По *F-критерию* можно проверить и гипотезу о равенстве двух безразмерных показателей рассеяния — *коэффициентов вариации* $\gamma_1 = \sigma_1/\eta_1$ и $\gamma_2 = \sigma_2/\eta_2$. Для этого при $m_u \geq 6$ используются [2]

оценки $v_1 = s_1/\bar{Y}_1$ и $v_2 = s_2/\bar{Y}_2$ с числом степеней свободы $f_i = m_i - 1$, тогда

$$\hat{F} = \left[\frac{v_2^2}{1 + v_1^2} \left(\frac{m_1}{m_1 + 1} \right) \right] / \left[\frac{v_2^2}{1 + v_2^2} \left(\frac{m_2}{m_2 + 1} \right) \right]. \quad (\text{III.71})$$

Сравнение двух средних — проверка гипотезы $H_0: \eta_1 = \eta_2$ — осуществляется по t -критерию

$$\hat{t} = |Y_1 - Y_2| / (\tilde{s}\{\bar{Y}\} \sqrt{2}). \quad (\text{III.72})$$

При этом вид обобщенной (для двух точек A и B) дисперсии средних $\tilde{s}^2\{\bar{Y}\}$ и число степеней свободы критерия зависят от того, принята ли первая гипотеза $H_0: \sigma_1^2 = \sigma_2^2$ (III.73) — (III.74) или отклонена (III.75) — (III.76), а также от соотношения числа испытаний в этих точках; все четыре возникающие при этом ситуации показаны в табл. III.12. При получении дробных значений f они

Таблица III.12. Формулы для расчета обобщенной дисперсии $s\{Y\}$ и числа степеней свободы при сравнении двух средних

Число испытаний	Гипотеза однородности дисперсий	
	принята и $\sigma_1^2 = \sigma_2^2$	отклонена и $\sigma_1^2 \neq \sigma_2^2$
$m_1 \neq m_2$	$\tilde{s}^2\{\bar{Y}\} = \tilde{s}^2(m_1^{-1} + m_2^{-1})/2;$ $\tilde{s}^2 = \frac{(m_1 - 1)s_1^2 + (m_2 - 1)s_2^2}{m_1 + m_2 - 2};$ $f = m_1 + m_2 - 2;$ <p style="text-align: right;">(III.73)</p>	$\tilde{s}^2\{\bar{Y}\} = \left(\frac{s_1^2}{m_1} + \frac{s_2^2}{m_2} \right) / 2;$ $f = \tilde{s}^2\{\bar{Y}\} / \left[\frac{s_1^2}{m_1 + 1} + \frac{s_2^2}{m_2 + 1} \right] - 2;$ <p style="text-align: right;">(III.75)</p>
$m_1 = m_2 = m$	$\tilde{s}^2\{\bar{Y}\} = (s_1^2 + s_2^2)/(2m);$ $f = 2(m - 1);$ <p style="text-align: right;">(III.74)</p>	$\tilde{s}^2\{\bar{Y}\} = (s_1^2 + s_2^2)/(2m);$ $f = (m - 1)(s_1^2 + s_2^2)^2 / [(s_1^2)^2 + (s_2^2)^2];$ <p style="text-align: right;">(III.76)</p>

округляются до целого. Из (III.71) следует простое правило: если доверительные интервалы для средних не перекрываются, то разница между η_1 и η_2 существует (однако пересечение интервалов не доказывает, что $\eta_1 = \eta_2$). Критерии (III.72) — (III.76) наиболее распространены в инженерной практике, но не единственны [2 и др.].

Если \bar{t} меньше чем t_T при риске α (обычно двусторонний $\alpha = 0,1$) и f (табл. III.12 и прил. II.5), то гипотеза $H_0: \eta_1 = \eta_2$ допускаются; можно считать, что изменений в среднем при переходе от точки A к точке B не происходит. Это может быть положительным результатом, если анализируется стабильность технологии в условиях «старения» сырья; однако этот вывод будет негативным, если рассматривается попытка улучшения свойств материала за счет изменения рецептуры;

Если \bar{t} больше чем t_T , то гипотеза $H_0: \eta_1 = \eta_2$ отклоняется.

Пример III.19. По данным информационной табл. III.2 в опыте в точке № 47 получено $\bar{R}_{1.47} = \bar{Y}_1 = 23,38$ МПа и $s\{R_{1.47}\} = s_2 = 1,67$ МПа (пр. III.7), а в точке № 52 (отличается уровнем П/Ц и τ_0) $\bar{R}_{1.52} = \bar{Y}_2 = 20,88$ МПа и $s\{R_{1.52}\} = s_1 = 2,02$ МПа.

Проверка гипотезы об однородности дисперсий по (III.70) при $s_2 > s_1$ показывает, что при $f_2 = f_1 = 6 - 1 = 5$

$$\hat{F} = 2,02^2/1,67^2 = 1,46 < F_{0,05} = 5,05,$$

поэтому ее можно допустить и использовать для проверки гипотезы о равенстве средних (III.73). Поскольку

$$f = 2(6 - 1) = 10;$$

$$\tilde{s}^2\{\bar{Y}\} = (2,02^2 + 1,67^2)/(2 \cdot 6) = 6,8693/12 = 0,5724;$$

$$\hat{t} = |23,38 - 20,88|/\sqrt{2 \cdot 0,5724} = 2,34 < t_{0,1} = 1,81,$$

то разница между средними признается значимой, и переход от технологии бетона в точке № 52 к технологии в точке № 47 повышает среднюю прочность \bar{R}_1 .

В заключение следует отметить, что если бы риск пришлось уменьшить в 5 раз, то при $t_{0,02}$ (см. прил. II.5) такой вывод был бы ошибочным.

Сравнение двух истинных частот v_1 и v_2 появления некоторого события Q при гипотезе $H_0: v_1 = v_2$ (против альтернативы $H_1 = v_1 > v_2$) может [20] производиться на основе информации, собранной в классификационную таблицу типа 2×2 (табл. III.13),

Таблица III.13. Информация о частоте Q (дефектность образцов) в классификационной таблице 2×2

Точка факторного пространства	Событие Q произошло (образец дефектен)	Событие Q не произошло; произошло противоположное событие \bar{Q} (образец остался без изменений)	Сумма частот
Среда A (точка 2)	$m_{Q2} = 12$ $p_{Q2} = 0,250$	$m_{\bar{Q}2} = 36$	$m_2 = 48$
Среда B (точка 1)	$m_{Q1} = 13$ $p_{Q1} = 0,361$	$m_{\bar{Q}1} = 25$	$m_1 = 36$
Сумма частот	$m_{Q1} + m_{Q2} = 25$ $p = 0,298$	$m_{\bar{Q}1} + m_{\bar{Q}2} = 61$	$m_1 + m_2 = 84$

с помощью эмпирических частостей

$$p_{Q1} = m_{Q1}/m_1; p_{Q2} = m_{Q2}/m_2; \bar{p} = (m_{Q1} + m_{Q2})/(m_1 + m_2); \quad (\text{III.77})$$

по t -критерию (III.78) при $f = m_1 + m_2 - 2$

$$\hat{t} = p_{Q1} - p_{Q2} - 0,5 (m_1^{-1} + m_2^{-1}) / \sqrt{\bar{p}(1-\bar{p})(m_1^{-1} + m_2^{-1})}. \quad (\text{III.78})$$

Критерий (III.78) применим, если $m_i m_j / (m_1 + m_2) > 5$.

Пример III.20. Две серии образцов находились в агрессивных средах A и B , причем после равного количества циклов набухания и высыхания на некоторых из них появились признаки разрушения (событие Q). Информация представлена в табл. III.13.

Проверка применимости t -критерия по самому малому значению $(m_{Q1} + m_{Q2}) m_1 / (m_1 + m_2) = 25 \cdot 36/84 = 11 > 5$ дает положительный результат, поэтому рассчитывается

$$\hat{t} = \frac{0,301 - 0,205 - 0,5 (48^{-1} + 36^{-1})}{\sqrt{0,298(1-0,298)(48^{-1} + 36^{-1})}} = 1,306.$$

Гипотеза о равенстве долей дефектных изделий при риске $\alpha = 0,1$ принимается, поскольку \hat{t} меньше $t_{\alpha} = 1,664$ при $f = 84 - 2 = 82$, и допустим вывод о равной опасности сред A и B для сохранности качества материала (если же риск уменьшить, то гипотеза $H_0: \nu_1 = \nu_2$ может быть отклонена).

Сравнение результатов в *нескольких* (N) *точках* факторного пространства является логическим расширением вышеизложенных задач сравнения для точек A и B . В такой постановке задачи не только потребовали специальных (нередко весьма сложных) критериев и методов для проверки статистических гипотез, но и в силу многообразия возможных вариантов выделились в самостоятельную область анализа многомерных статистических комплексов. Ниже будет проанализировано несколько (хотя и простейших, но весьма важных для инженерных выводов в области материаловедения и технологии) постановок таких задач.

В практике анализа, моделирования и оптимизации систем весьма часто необходима проверка гипотезы об *однородности* (равенстве) *дисперсий* в некоторой области факторного пространства, т. е. нуль-гипотезы $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_u^2 = \dots = \sigma_N^2 = \bar{\sigma}^2$. Если она будет признана правдоподобной, то можно рассчитать *обобщенную* на все исследованное пространство оценку дисперсии \bar{s}^2 , которая будет стремиться к $\bar{\sigma}^2$. Такое допущение, в частности, существенно облегчает построение последующих методик моделирования.

Наиболее прост критерий проверки гипотезы однородности дисперсии для *равного числа испытаний* во всех точках $m_u = \text{const}$ (к этому, как указывалось в п. III.1, целесообразно стремиться при организации экспериментальных работ). Рассчитывается *G-критерий* Кокрена, который представляет собой отношение максимальной оценки дисперсии $s_{u,\text{max}}^2$ в сравниваемом ряду к сумме всех N дисперсий (включая $s_{u,\text{max}}^2$):

$$\hat{G} = s_{u,\text{max}}^2 / \sum_{u=1}^N s_u^2. \quad (\text{III.79})$$

Критические значения G -критерия зависят не только от допустимого риска α (они табулированы для $\alpha = 0,05$ в прил. II.7) и числа степеней свободы $f = m_u - 1$, но и от количества N проверяемых на однородность групп дисперсий. Если $\hat{G} < G_T$, то гипотеза однородности принимается как правдоподобная. Если $\hat{G} > G_T$, то для обеспечения однородности дисперсий целесообразно преобразование Y , а в некоторых инженерных задачах допустимо исключение из анализа той точки, к которой относится $s_{u,\max}^2$, с последующей проверкой ряда из $K - 1$ элементов на однородность дисперсий.

Пример III.21. В информационной табл. III.2 в пр. III.1 показаны результаты определения прочности трех разных бетонов после ТВО через 1 и 28 сут на 6 образцах в каждом опыте. Рассчитанные по ним оценки \bar{R} и $s^2\{Y\}$ приведены в табл. III.14. Необходимо проверить однородность этих данных с целью дальнейшего описания единой многофакторной моделью, включающей время твердения.

Таблица III. 14. Характеристики опытов (точки № 47, 52 и 17 в табл. III. 2) и проверка однородности дисперсий

Статистические оценки	Оценки при определении прочности бетона						Σs^2	\hat{G}	
	через 1 сут для точек плана			через 28 сут для точек плана					
	47	52	17	47	52	17			
$s^2\{R\}$	2,78	4,07	2,15	12,06	7,30	1,09	29,45	0,410	
Расчет по коэффициентам вариаций	$s\{R\}$, МПа	1,67	2,02	1,47	3,47	2,70	1,04	—	—
	\bar{R} , МПа	23,38	20,88	24,43	48,58	46,55	33,68	—	—
	$v\{R\}$, %	7,1	9,7	6,0	7,1	5,8	3,1	—	—
	$v^2\{R\}$	50,41	94,09	36,00	50,41	33,64	9,61	274,16	0,343

При сравнении суммы шести дисперсий $s^2\{R\}$ и $s_{\max}^2 = 12,06$ выборку можно признать однородной, так как $\hat{G} = 0,41$ меньше $G_T = 0,44$ (см. прил. II.7) при $f = 5$ и $K = 6$; дисперсии нужно объединить и использовать $s^2\{R\} = 29,45/6 = 4,9$.

Однородность данной совокупности можно несколько повысить, если вместо дисперсий рассмотреть квадраты коэффициентов вариации $v^2\{R\}$, однако тогда нужно будет моделировать (см. п. III.7) не прочность бетона, а $\ln R$.

Для проверки однородности ряда из K дисперсий при неравном числе испытаний нужно при условии $f_u = m_u - 1 > 2$ [10] воспользоваться критерием Бартлетта B , в котором кроме частных оценок s_u^2 используется «взвешенная» по f_u дисперсия \bar{s}^2 с $f = K - 1$:

$$B = \frac{1}{C} \left[\ln \bar{s}^2 \cdot f_{\Sigma} - \sum_{u=1}^K (f_u \ln s_u^2) \right], \quad (\text{III.80})$$

$$\left. \begin{aligned} \text{где } C &= 1 + \frac{1}{3(K-1)} \left(\sum_{u=1}^K f_u^{-1} - f_{\Sigma}^{-1} \right), \quad f_{\Sigma} = \sum_{u=1}^K f_u, \\ \bar{s}^2 &= \left(\sum_{u=1}^K f_u s_u^2 \right) / f_{\Sigma}. \end{aligned} \right\} \quad (\text{III.81})$$

Случайная величина B распределена как χ^2 при f_K , поэтому, если $B < \chi_{\alpha}^2 \{K-1\}$, то гипотеза однородности дисперсий с риском α (обычно 0,05) принимается.

Пример III.22. Если предположить, что в пр. III.21 при определении R_{28} число образцов $m_u = 4$ (а не 6, как для R_1), то для проверки нуль-гипотезы нужно рассчитать B -критерий по процедуре, представленной в табл. III.15.

Таблица III.15. Расчет критерия Бартлетта

Расчет вспомогательных величин для (III.80)	Значения величин при определении прочности бетона						Сумма
	через 1 сут для точек плана			через 28 сут для точек плана			
	47	52	17	47	52	17	
s_u^2	2,78	4,07	2,15	12,06	7,30	1,09	—
f_u	5	5	5	3	3	3	24
$f_u s_u^2$	13,90	20,35	10,75	36,18	21,90	3,09	106,17
f_u^{-1}	0,2	0,2	0,2	0,33	0,33	0,33	1,59
$\ln s_u^2$	1,022	1,404	0,765	2,490	1,988	0,086	—
$(f \ln s^2)_u$	5,110	7,020	3,825	7,470	5,964	0,258	29,647

По результатам расчетов вспомогательных величин определяются:

$$\bar{s}^2 = 106,17/24 = 4,424; \quad \ln \bar{s}^2 = 1,487;$$

$$C = 1 + \frac{1}{3(6-1)} (1,59 - 1/24) = 1,103;$$

$$B = \frac{1}{1,103} [1,487 \cdot 24 - 29,647] = 5,477 < \chi_{0,05}^2 = 11,07.$$

Ряд дисперсий при $f_K = 5$ можно признать однородным, поскольку B меньше критического значения (см. прил. II.4).

Нуль-гипотеза о равенстве средних в K точках факторного пространства ($H_0: \eta_1 = \dots = \eta_u = \dots = \eta_k$) проверяется в ходе так называемого дисперсионного анализа, который составляет особую, весьма развитую ветвь в прикладном статистическом анализе [20 и др.]. Ниже он проанализирован для простейшего случая: все испытания классифицированы по одному признаку A , т. е.

образуют ряд $A_1, \dots, A_u, \dots, A_k$ при равном числе испытаний в группе $m_u = \text{const}$.

Дисперсионный анализ удобно вести в табл. III.16, где выделяются два источника вариации Y : *внутренний* (за счет рассеяния

Таблица III.16. Дисперсионный анализ при классификации по одному признаку

Источник рассеяния	Сумма квадратов	Число степеней свободы	Оценка дисперсии
Различие между уровнями A_1, \dots, A_k групп	$SS_{\text{МГ}} = SS_2 - C_0$	$m - 1$	$S_{\text{МГ}}^2 = \frac{SS_{\text{МГ}}}{m - 1}$
Различие внутри групп при фиксированном уровне A_u	$SS_{\text{ВГ}} = SS_1 - SS_2$	$K(m - 1)$	$S_{\text{ВГ}}^2 = \frac{SS_{\text{ВГ}}}{K(m - 1)}$
Общее рассеяние	$SS_0 = SS_1 - C_0$	$Km - 1$	

в каждой u -й точке) и *внешний* (за счет рассеяния в результате перехода системы с уровня A_u на уровень A_{u-1}). Первый источник дает, как было показано выше, *внутригрупповую дисперсию* $s_{\text{ВГ}}^2 = s_u^2$, а второй — *межгрупповую* $s_{\text{МГ}}^2$. Если при проверке по F -критерию отношения $s_{\text{МГ}}^2/s_{\text{ВГ}}^2$ (межгрупповая всегда в числителе!) будет показано, что $\hat{F} < F_T$, то следует признать отсутствие значимого изменения Y при переходе между уровнями $A_1 \rightarrow \dots \rightarrow A_u \rightarrow \dots \rightarrow A_K$ по сравнению с «собственными» случайными колебаниями при фиксации системы на u -м уровне, а следовательно, справедлива нуль-гипотеза $H_0: \eta_1 = \dots = \eta_u = \dots = \eta_K$.

В табл. III.16 выделены, как это общепринято в дисперсионном анализе, суммы квадратов SS упорядоченных определенным образом наблюдений:

SS_1 — сумма квадратов всех наблюдений Y или преобразованных величин $Z = (Y - C_c)C_{\text{п}}$, причем при $C_{\text{п}} = 1$ значения $s_{\text{МГ}}^2$ и $s_{\text{ВГ}}^2$ не меняются (см. III.28):

$$SS_1 = \sum_{u=1}^K \sum_{v=1}^{m_u} y_{uv}^2; \quad (\text{III.82})$$

SS_2 — отнесенная к m_u сумма квадратов итогов по группам

$$SS_2 = \frac{1}{m_u} \sum_{u=1}^K \left(\sum_{v=1}^{m_u} y_{uv} \right)^2; \quad (\text{III.83})$$

C_0 — квадрат общего итога, отнесенный к общему числу наблюдений:

$$C_0 = \frac{1}{Km_u} \left(\sum_{u=1}^K \sum_{v=1}^{m_u} y_{uv} \right)^2. \quad (\text{III.84})$$

Пример III.23. Опыты по определению прочности бетона в условиях точки № 17 в табл. III.2 (нормальное твердение; факторов $X_6 - X_9$ нет) были полностью повторены (от приготовления компонентов до испытания кубов) четыре раза и образовали группы *A, B, C* и *D*. Расчет выполняется по преобразованным данным $Z = 10(R - 20)$ в табл. III.17; там же определены вспомогательные суммы для использования в (III.82) — (III.84).

Таблица III.17. Исходные и расчетные данные для дисперсионного анализа прочности бетона

Порядковый номер <i>и</i> -го испытания	Исходные данные				Квадраты			
	Z_A	Z_B	Z_C	Z_D	Z_A^2	Z_B^2	Z_C^2	Z_D^2
1	-36	-39	-20	-25	1296	1521	400	625
2	-33	-39	-6	4	1089	1521	36	16
3	-17	-20	-2	10	289	400	4	100
4	8	10	4	18	64	100	16	324
5	12	10	12	26	144	100	144	676
6	12	12	18	47	144	144	324	2209
Суммы по столбцам	-54	-66	6	80	3026	3786	924	3950
Итоговая сумма	$-34 = \sqrt{Km_u C_0}$				11 686 = SS_1			
Квадраты сумм	2916	4356	36	6400				
Итог	13 708 = mSS_2							
Числовые характеристики	$\bar{Z}_A =$ = -9	$\bar{Z}_B =$ = -11	$\bar{Z}_C =$ = -1	$\bar{Z}_D =$ = 13,3	$s_A^2 =$ = 508	$s_B^2 =$ = 612	$s_C^2 =$ = 183	$s_D^2 =$ = 577
	$\bar{Z} = 1,9; \bar{s}^2 \{Z\} = 124,3$				$\bar{s}^2 = 470$			

Сумма квадратов SS_1 определена в табл. III.17;

$$SS_2 = 13\,708/6 = 2285; C_0 = (-34)^2/(4 \cdot 6) = 48.$$

Дисперсионный анализ выполнен в табл. III.18. Фактическое дисперсионное отношение

$$\hat{F} = \frac{s_{MG}^2 \{Z\}}{s_{BG}^2 \{Z\}} = \frac{s_{MG}^2 \{R\}}{s_{BG}^2 \{R\}} = \frac{7,46}{4,7} = 1,58 < F_T = 3,1, \quad (III.85)$$

т.е. меньше критического ($\alpha = 0,05, f_1 = f_{MG} = 3, f_2 = f_{BG} = 20$), что позволяет принять гипотезу о равенстве дисперсий $\sigma_{MG}^2 = \sigma_{BG}^2$ и, следовательно, рассчитать объединенную дисперсию

$$\begin{aligned} \bar{s}^2 \{Z\} &= [(K-1)s_{MG}^2 + K(m_u-1)s_{BG}^2]/[(K-1) + K(m_u-1)] = \\ &= [(4-1) \cdot 746 + 4(6-1) \cdot 470]/[(4-1) + 4(6-1)] = 506 \text{ МПа}. \end{aligned} \quad (III.86)$$

Очевидно, что результат (III.85) может быть получен и по строке «Общее рассеяние» в табл. III.18. Оценка $\sqrt{\bar{s}^2 \{Y\}} = s\{Z\}/10 = \sqrt{5,06} = 2,25 \text{ МПа}$

Таблица III.18. Дисперсионный анализ прочности мелкозернистого бетона при повторении одинаковых опытов

Источник рассеяния	Сумма квадратов	f_j	Оценка дисперсий
Различие между группами A, B, C, D	$SS_{\text{МГ}} = 2285 - 48 = 2237$	$4 - 1 = 3$	$s_{\text{МГ}}^2 \{Z\} = 746$ $s_{\text{МГ}}^2 \{Y\} = 7,46$
Различие внутри групп (опытов)	$SS_{\text{ВГ}} = 11\,686 - 2285 = 9401$	$4 \times 5 = 20$	$s_{\text{ВГ}}^2 \{Z\} = 470$ $s_{\text{ВГ}}^2 \{Y\} = 4,70$
Общее расстояние	$SS_0 = 11\,686 - 48 = 11\,638$	23	—

является оценкой абсолютной среднеквадратической ошибки эксперимента $s_{\text{э}}\{Y\}$. В данном случае она одинакова и для характеристик параллельных испытаний (повторных измерений) внутри u -го опыта (внутри группы), и для характеристики полного воспроизведения u -го опыта K раз в серии (между опытами-группами).

В таких простых ситуациях анализ дисперсий можно, вообще говоря, провести и минуя дисперсионный анализ — через прямой расчет числовых характеристик K выборок. В табл. III.17 (в нижней части) приведены числовые характеристики обсужденных групп A, B, C и D . Видно, что

$$\bar{s}^2 = s_{\text{ВГ}}^2 = s_{\text{восп}}^2 = 470, \quad \text{а} \quad s_{\text{МГ}}^2 = m_u s^2 \{Z\} = 6 \cdot 124,3 = 746.$$

Методы, формулы и результаты сравнения нескольких оценок числовых характеристик параметров генеральной совокупности систематически используются как на этапах подготовки, проведения и первичной обработки эксперимента (особенно при принятии метрологических решений [2, 4]), так и на этапах анализа самих многофакторных ЭС-моделей (последовательный регрессионный анализ и др. — см. п. III.7 — III.4) и принятия по ним инженерных решений (см. п. III.15 — III.17).

III.7. Построение и статистический анализ линейных моделей

Одними из частных видов полиномиальных ЭС-моделей с натуральными X_i (II.4) или с нормализованными x_i (II.11) факторами являются *линейные модели*

$$\hat{Y} = b_0 x_0 + b_1 x_1 + \dots + b_i x_i + \dots + b_k x_k = \sum_{i=0}^k b_i x_i = \mathbf{b} \cdot \mathbf{x}, \quad (\text{III.87})$$

описывающие в k -факторном пространстве поверхности, которые в любом однофакторном сечении дают *прямую*

$$\hat{Y} = b_0 + b_1 x_1, \quad (\text{III.88})$$

а в любом двухфакторном сечении (без потери общности оставлены факторы x_1 и x_2) — *плоскость*

$$\hat{Y} = b_0 + b_1 x_1 + b_2 x_2 \quad (\text{при } |x_i| \leq 1). \quad (\text{III.89})$$

В модель (III.87), называемую также *моделью первого порядка* ГОСТ 24026 — 80), иногда вводится *фиктивный фактор* $x_0 \equiv +1$ для удобства ее записи в матричной форме; при этом матрица плана x (см. табл. III.2) приобретает размер $k + 1$; модель содержит $L\{b_i\} = k + 1$ линейных эффектов (см. п. III.2) и описывает k -мерную гиперплоскость. Изучение методики построения и статистического анализа таких моделей целесообразно начать с простейших моделей (III.88), (III.89), обобщая выводы, как правило, в матричной форме на общий случай (III.87).

При выдвижении гипотезы об описании на основе экспериментов поведения системы линейной моделью целесообразно различать две ситуации. В *первой* ситуации эта гипотеза выдвигается на основе глубокого анализа с применением фундаментальных законов существа структуры и функций системы. Поскольку *гипотеза единственна* и не имеет альтернатив (простейший пример — линейная модель себестоимости продукции как сумма произведений цены за единицу ресурса на число таких единиц! — см. п. II.5), то, строго говоря, для построения прямой (III.88) необходимы только две экспериментальные точки с x_1 на разных уровнях, для плоскости (III.89) — только три разные точки, а для гиперплоскости (III.87) — только $k + 1$ точка. Все остальные, так сказать, «промежуточные» точки не нужны. Во *второй* ситуации кроме гипотезы линейности выдвигаются *конкурирующие* альтернативные гипотезы о *нелинейных* (параболических, экспоненциальных, тригонометрических и т. п.) моделях, и для их проверки число экспериментальных точек должно быть, как правило, больше, чем в первой ситуации. Однако на первых поисковых этапах исследования допустимо вторую ситуацию временно сводить к первой и исследовать только модели (III.87).

При построении, в частности, однофакторной линейной модели необходимо подобрать такие значения оценок коэффициентов b_0 и b_1 , чтобы они были как можно ближе к истинным параметрам β_0 и β_1 , т. е. наиболее точно описывали данные опытов в информационной табл. III.1 (которая в этом случае содержит два столбца — x_1 и y_u — с N строками). Геометрически это означает, что нужно прямую в координатах $\{x_1, Y\}$ (рис. III.12) провести с учетом N экспериментальных точек «наилучшим образом». Если $N > 2$, то, как правило, провести прямую через все точки невозможно — всегда останется некоторое расстояние («невязка») Δ_u между экспериментальным значением y_u и расчетным \hat{y}_u (по выбранной линейной модели)

$$\Delta_u = y_u - \hat{y}_u. \quad (\text{III.90})$$

Пожелание провести прямую «наилучшим образом» конкретизируется как *минимизация нормы* $\|\Delta_u\|$ (см. п. I.2). Отличные от нуля значения может дать минимизация октаэдрической нормы — суммы абсолютных значений

$$\sum_{u=1}^N |\Delta_u| = \sum_{u=1}^N |y_u - \hat{y}_u| \rightarrow \min \quad (\text{III.91})$$

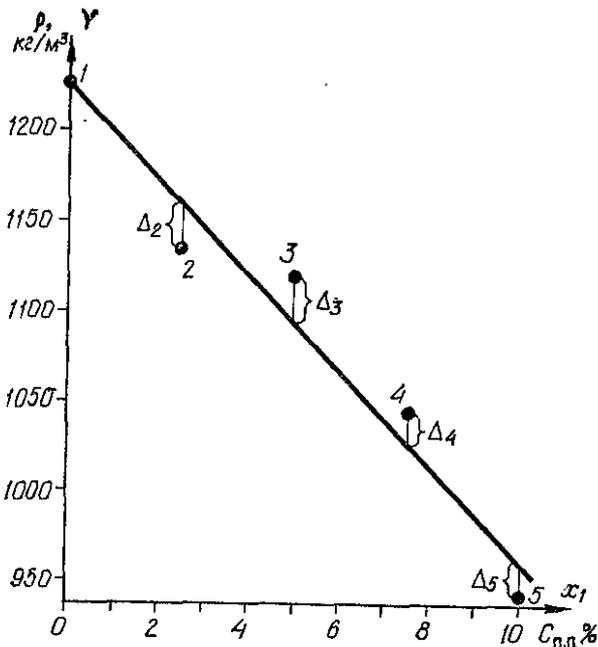


Рис. III.12. Однофакторная линейная ЭМ $\hat{\rho} = f(x)$ и ее невязки с эмпирическими точками

эффективной характеристике рассеяния, как дисперсия (ср. III.25 и III.26):

$$s_{\text{ост}}^2 = \sum_{u=1}^N \Delta_u^2 / f_{\text{ост}}. \quad (\text{III.93})$$

Дисперсия (III.93) называется *остаточной*; она имеет число степеней свободы $f_{\text{ост}}$, равное разности между числом неодинаковых (по x_i) опытов N и числом $L\{b_i\}$ оцененных коэффициентов модели:

$$f_{\text{ост}} = N - L\{b_i\}. \quad (\text{III.94})$$

Для (III.88) $L\{b_i\} = 2$, а для (III.87) $L\{b_i\} = k + 1$.

Метод построения моделей, основанный на соотношениях (III.92) — (III.94), т. е. на *методе наименьших квадратов* — МНК, хорошо теоретически обоснован и является *базовым* при построении ЭС-моделей в задачах материаловедения и технологии. Другой эффективный способ оценки неизвестных параметров модели — метод максимального правдоподобия [10, 11, 40] — в этой области науки и техники имеет несравненно меньшее приложение (используются результаты его применения к решению задач математической теории эксперимента).

Для однофакторной модели выражение (III.92) с учетом $x_0 \equiv \equiv + 1$ конкретизируется как

$$\sum_u \Delta_u^2 = SS_{\text{ост}} = \sum_u (y_u - \hat{y}_u)^2 = \sum_u [y_u - (b_0 x_0 + b_1 x_{1u})]^2 \rightarrow \min, \quad (\text{III.95})$$

а для k -факторной как

$$SS_{\text{ост}} = \sum_u (y_u - \sum_{i=1}^k b_i x_{iu})^2 \rightarrow \min. \quad (\text{III.96})$$

и квадратичной нормы — суммы квадратов отклонений Δ_u^2

$$\sum_{u=1}^N \Delta_u^2 = \sum_{u=1}^N (y_u - \hat{y}_u)^2 \rightarrow \min.$$

(III.92)

Вторая норма предпочтительнее (строгие доказательства приведены в теории математической статистики); при этом можно отметить, что, во-первых, операция возведения в квадрат делает все *отклонения положительными* (ср. III.91), во-вторых, при возведении в квадрат большим отклонениям придается *большая степень влияния* на результат, в третьих, форма (III.92) позволяет перейти к такой базовой, хорошо изученной в математике и

Здесь и далее знак \sum_u обозначает суммирование по всем опытам.

Для нахождения минимума функции $SS_{\text{ост}}$ частные производные по всем неизвестным приравниваются нулю (см. гл. II), и получается система так называемых *нормальных уравнений* (общим числом $k + 1$). Для однофакторной модели ($x_0 \equiv +1$) неизвестны две оценки — b_0 и b_1 , следовательно,

$$\left. \begin{aligned} \frac{\partial SS_{\text{ост}}}{\partial b_0} &= -2 \sum_u (y_u - b_0 x_0 - b_1 x_{1u}) = 0; \\ \frac{\partial SS_{\text{ост}}}{\partial b_1} &= -2 \sum_u (y_u - b_0 x_0 - b_1 x_{1u}) x_{1u} = 0. \end{aligned} \right\} \quad (\text{III.97})$$

Алгебраические преобразования

$$\left. \begin{aligned} \sum_u y_u - \sum_u b_0 x_0 - \sum_u b_1 x_{1u} &= 0, \\ \sum_u y_u x_{1u} - \sum_u b_0 x_0 x_{1u} - \sum_u b_1 x_{1u}^2 &= 0 \end{aligned} \right\} \quad (\text{III.98})$$

с учетом $\sum_u (x_0 \equiv +1) = N$ приводят систему (III.97) к виду

$$\left. \begin{aligned} N b_0 + \left(\sum_u x_{1u} \right) b_1 &= \sum_u y_u; \\ \left(\sum_u x_{1u} \right) b_0 + \left(\sum_u x_{1u}^2 \right) b_1 &= \sum_u x_{1u} y_u. \end{aligned} \right\} \quad (\text{III.99})$$

По аналогии для двухфакторной модели можно получить:

$$\left. \begin{aligned} N b_0 + \left(\sum_u x_{1u} \right) b_1 + \left(\sum_u x_{2u} \right) b_2 &= \sum_u y_u; \\ \left(\sum_u x_{1u} \right) b_0 + \left(\sum_u x_{1u}^2 \right) b_1 + \left(\sum_u x_{1u} x_{2u} \right) b_2 &= \sum_u x_{1u} y_u; \\ \left(\sum_u x_{2u} \right) b_0 + \left(\sum_u x_{1u} x_{2u} \right) b_1 + \left(\sum_u x_{2u}^2 \right) b_2 &= \sum_u x_{2u} y_u. \end{aligned} \right\} \quad (\text{III.100})$$

Для упрощения записи нормальных уравнений удобно использовать *условные обозначения сумм* всех *известных* из эксперимента *элементов* (x_{iu} — априорно, y_u — апостериорно, см. п. III.1):

$$\left. \begin{aligned} N = \sum_u x_0 &= (00); \quad \sum_u x_{iu} = (i0) = (0i); \quad \sum_u x_{iu}^2 = (ii); \\ \sum_u x_{iu} x_{ju} &= (ij) = (ji) \quad (\text{при } i < j); \\ \sum_u y_u &= (0Y); \quad \sum_u x_{iu} y_u = (iY). \end{aligned} \right\} \quad (\text{III.101})$$

Тогда для однофакторной модели система нормальных уравнений будет

$$\left. \begin{aligned} (00)b_0 + (01)b_1 &= (0Y); \\ (10)b_0 + (11)b_1 &= (1Y), \end{aligned} \right\} \quad (\text{III.102})$$

а для k -факторной —

$$\begin{aligned} (00)b_0 + (01)b_1 + \dots + (0K)b_k &= (0Y); \\ (\ddot{K}\ddot{O})b_0 + (\ddot{K}\ddot{I})b_1 + \dots + (\ddot{K}\ddot{K})b_k &= (\ddot{K}\ddot{Y}). \end{aligned} \quad (\text{III.103})$$

Пример III.24. При разработке гипсового композита исследовалось влияние на плотность ρ , кг/м³, в сухом состоянии введения вспученного перлитового песка в количестве $0 \leq Y_1 = C_{п.п} \leq 10\%$ от массы гипса [6] при формовании изделий из технологической смеси нормальной густоты (по Суттарду). При гипотезе линейного снижения ρ в зависимости от нормализованного фактора $x_1 = (x_1 - 5)/5$ нужно найти две оценки МНК в модели $\rho = b_0 + b_1 x_1$ по результатам пяти опытов, представленных в табл. III.19; там же дан расчет вспомогательных величин для составления нормальных уравнений.

Таблица III.19. Результаты пяти опытов при исследовании $\rho = \varphi(C_{п.п})$ и расчет вспомогательных величин для оценки b_0 и b_1 по МНК

Номер опыта	Количество перлита $C_{п.п.} \%$	x_{1u}	$y_u = \rho$	x_{1u}^2	$x_{1u}y_u$	$\hat{y}_u = \hat{\rho}_u$	Δ_u	Δ_u^2
1	0	-1	1228	1	-1228	1227	+1	1
2	2,5	-0,5	1136	0,25	-568	1160	-24	576
3	5	0	1120	0	0	1094	+26	676
4	7,5	+0,5	1044	0,25	522	1028	+16	256
5	10	+1	942	1	942	961	-19	361
(00) = 5	—	(01) = 0	(0Y) = 5470	(11) = 2,5	(1Y) = -332	—	$\sum \Delta_u = 0$	$SS_{\text{ост}} = 1870$

Нормальные уравнения (III.102) с учетом данных табл. III.19 принимают вид

$$\begin{aligned} 5b_0 + 0 \cdot b_1 &= 5470; \\ 0 \cdot b_0 + 2,5 b_1 &= -332 \end{aligned} \quad (\text{III.104})$$

и имеют решение $b_0 = 5470/5 = 1094$ и $b_1 = -332/2,5 = -132,8$, что дает линейную модель

$$\hat{\rho} = 1094 - 132,8x_1, \quad (\text{III.105})$$

показанную на рис. III.12 в виде единственной прямой.

Можно было бы «для наглядности» подставить x_1 в (III.105) и получить модель в виде $\hat{\rho} = 1226,8 - 53,1 C_{п.п.}$, но этого делать, строго говоря, нельзя [4, 6], поскольку ухудшаются свойства модели и условия ее интерпретации.

По модели (III.105) определены расчетные значения $\hat{\rho}$, а также соответствующие отклонения Δ_u и их квадраты Δ_u^2 (табл. III.19). Сумма квадратов $SS_{\text{ост}} = 1870$ является минимально возможной для линейной модели. Остаточная дисперсия $s_{\text{ост}} = 1870/(5 - 2) = 623$ имеет число степеней свободы $f_{\text{ост}} = 3$ (по 5 опытам оценено 2 коэффициента). Оценка среднеквадратичного отклонения $s\{\rho\} = 25$ кг/м³, т. е. $(25/1094) \cdot 100 = 2,3\%$ по отношению к отклику в центре факторного пространства.

В системе нормальных уравнений можно выделить три упорядоченных множества, рассматриваемых далее для двух- и k -факторных моделей.

Неизвестные оценки коэффициентов модели образуют вектор \mathbf{b} размером $(k + 1) \times 1$:

$$\mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}; \quad \mathbf{b} = \begin{pmatrix} b_0 \\ \dots \\ b_i \\ \dots \\ b_k \end{pmatrix}. \quad (\text{III.106})$$

Известные априори суммы (00) , $(i0)$, (ij) и (ii) , определяемые только уровнями факторов x_i в информационной табл. III.1, образуют квадратичную матрицу \mathbf{M} размером $(k + 1) \times (k + 1)$:

$$\mathbf{M} = \begin{pmatrix} (00) & (01) & \dots & (0j) & \dots & (0K) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ (i0) & (i1) & \dots & (ij) & \dots & (iK) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ (K0) & (K1) & \dots & (Kj) & \dots & (KK) \end{pmatrix}. \quad (\text{III.107})$$

Матрица (III.107) называется *матрицей моментов плана* (ГОСТ 24026—80); ее можно получить из матрицы плана \mathbf{x} , если последнюю умножить слева на транспонированную матрицу \mathbf{x}^T (см. прил. I.1):

$$\mathbf{M} = \mathbf{x}^T \mathbf{x}. \quad (\text{III.108})$$

С целью исследования свойств и обобщения результатов для матриц различных планов \mathbf{x} матрицу (III.107) целесообразно нормировать по числу опытов N , получая *информационную матрицу* \mathbf{M}_N (ГОСТ 24026—80):

$$\mathbf{M}_N = \frac{1}{N} \mathbf{M} = \frac{1}{N} \mathbf{x}^T \mathbf{x}. \quad (\text{III.109})$$

Известные апостериори суммы $(0Y)$ и (iY) , включающие результаты эксперимента, образуют вектор \mathbf{y}_M размером $(k + 1) \times 1$, являющийся произведением \mathbf{x}^T на вектор наблюдений \mathbf{y} :

$$\mathbf{y}_M = \begin{pmatrix} (0Y) \\ \dots \\ (iY) \\ \dots \\ (KY) \end{pmatrix}; \quad \mathbf{y}_M = \begin{pmatrix} (0Y) \\ \dots \\ (iY) \\ \dots \\ (KY) \end{pmatrix}; \quad (\text{III.110})$$

$$\mathbf{y}_M = \mathbf{x}^T \mathbf{y}. \quad (\text{III.111})$$

Следовательно, в матричной форме система нормальных уравнений (III.103) записывается как (III.112), а с учетом матрицы \mathbf{x} — как (III.113):

$$\mathbf{M} \mathbf{b} = \mathbf{y}_M, \quad (\text{III.112})$$

$$\mathbf{x}^T \mathbf{x} \mathbf{b} = \mathbf{x}^T \mathbf{y}, \quad (\text{III.113})$$

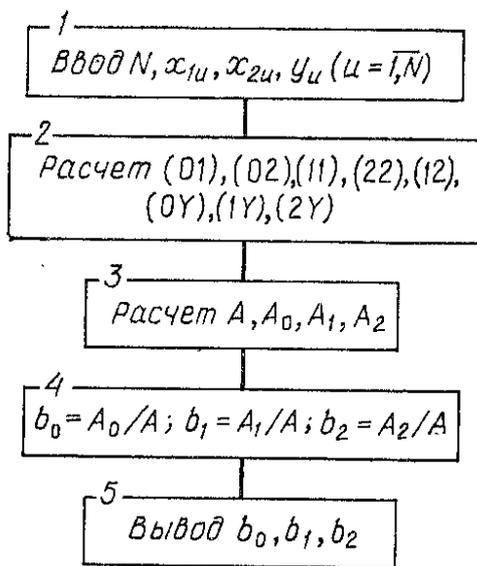


Рис. III. 13. Блок-схема алгоритма расчета оценок коэффициентов двухфакторной линейной модели

из нее определяется вектор оценок коэффициентов

$$\begin{aligned} \mathbf{b} &= \mathbf{M}^{-1} \mathbf{y}_M = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y} = \\ &= N^{-1} \mathbf{M}_N^{-1} \mathbf{x}^T \mathbf{y} = \mathbf{D} \mathbf{x}^T \mathbf{y} = \mathbf{L} \mathbf{y}. \quad (\text{III.114}) \end{aligned}$$

В решении (III.114) выделены две новые независимые от результатов эксперимента матрицы: так называемая **D-матрица**, или *ковариационная*, элементы которой оценивают (см. п. III. 8 и далее) статистические характеристики модели, и так называемая **L-матрица**, элементы которой позволяют [4, 58] рассчитывать оценки b_i без обращения матриц по простым формулам (см. п. III.11—III.13) с помощью микрокалькуляторов:

$$\mathbf{D} = \mathbf{M}^{-1} = (\mathbf{x}^T \mathbf{x})^{-1}; \quad (\text{III.115})$$

$$\mathbf{L} = \mathbf{D} \mathbf{x}^T = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T. \quad (\text{III.116})$$

Неизвестные оценки в системе нормальных уравнений (III.103) можно рассчитать через определитель $\det \mathbf{M} = A$ матрицы моментов и определители $\det \mathbf{M}_i = A_i$ матриц \mathbf{M}_i , образуемых путем замены i -го столбца на вектор-столбец \mathbf{Y}_M (III.110):

$$b_i = \det \mathbf{M}_i / \det \mathbf{M} = A_i / A. \quad (\text{III.117})$$

Таким образом, для однофакторной модели при использовании правила Крамера (см. прил. I.1) получаются простые формулы:

$$A = \begin{vmatrix} (00) & (10) \\ (10) & (11) \end{vmatrix}; \quad A_0 = \begin{vmatrix} (0Y) & (10) \\ (1Y) & (11) \end{vmatrix}; \quad A_1 = \begin{vmatrix} (00) & (0Y) \\ (10) & (1Y) \end{vmatrix}; \quad (\text{III.118})$$

$$\left. \begin{aligned} b_0 &= A_0/A = [(0Y)(11) - (1Y)(10)] / [(00)(11) - (10)^2]; \\ b_1 &= A_1/A = [(00)(1Y) - (10)(0Y)] / [(00)(11) - (10)^2]. \end{aligned} \right\} \quad (\text{III.119})$$

Аналогично получаются простые формулы и для двухфакторной модели (III.89). Блок-схема алгоритма расчета оценок b_0 , b_1 и b_2 приведена на рис. III.13.

Пример III.25. В гипсовый композит кроме перлита (пр. III.24) введен кремнегель (отход суперфосфатного производства) в количестве $0 \leq X_2 \leq 10\%$ от массы гипса и поставлено 3 дополнительных опыта. Результаты 8 опытов показаны в табл. III.20, где приведены и вспомогательные величины, необходимые для построения двухфакторной модели.

По данным табл. III.20 рассчитываются четыре определителя:

$$\begin{aligned} \det \mathbf{M} = A &= \begin{vmatrix} (00) & (01) & (02) \\ (10) & (11) & (12) \\ (20) & (21) & (22) \end{vmatrix} = \begin{vmatrix} 8 & 1 & -3 \\ 1 & 5,5 & 0 \\ -3 & 0 & 7 \end{vmatrix} = 8 \cdot 5,5 \cdot 7 + 0 \cdot 1 \cdot (-3) + \\ &+ 0 \cdot 1 \cdot (-3) - (-3) \cdot (-3) \cdot 5,5 - 0 \cdot 0 \cdot 8 - 1 \cdot 1 \cdot 7 = 252; \\ A_0 &= \begin{vmatrix} (0Y) & (01) & (02) \\ (1Y) & (11) & (12) \\ (2Y) & (21) & (22) \end{vmatrix} = \begin{vmatrix} 7794 & 1 & -3 \\ 238 & 5,5 & 0 \\ -3942 & 0 & 7 \end{vmatrix} = 233\,360; \end{aligned}$$

Таблица III.20. Результаты восьми опытов при исследовании $\rho = \varphi(C_{п.п}, C_{кг})$, и расчет вспомогательных величин для оценки b_0 , b_1 и b_2 по МНК

Номер опыта	Рецептура, %		x_{1u}	x_{2u}	$y_u = \rho$, кг/м ³	$x_{1u}x_{2u}$	x_{1u}^2	x_{2u}^2	$x_{1u}y_u$	$x_{2u}y_u$	$\hat{y}_u = \hat{\rho}$	Δ_u^2
	$C_{п.п}$	$C_{кг}$										
1	0	0	-1	-1	1228	+1	1	1	-1228	-1218	1228	100
2	2,5	0	-0,5	-1	1136	+0,5	0,25	1	-568	-1136	1155	361
3	5	0	0	-1	1120	0	0	1	0	-1120	1093	729
4	7,5	0	+0,5	-1	1044	-0,5	0,25	1	522	-1044	1031	169
5	10	0	+1	-1	942	-1	1	1	942	-942	968	676
6	0	10	-1	+1	877	-1	1	1	-877	877	888	121
7	10	5	+1	0	796	0	1	0	796	0	803	49
8	10	10	+1	+1	651	+1	1	1	651	651	638	169
(00) = 8	—	—	(01) = 1	(02) = = -3	(0Y) = = 7794	(12) = 0	(11) = = 5,5	(22) = 7	(1Y) = = 238	(2Y) = = -3942	—	$SS_{ост} =$ = 2374

$$A_1 = \begin{vmatrix} 8 & 7794 & -3 \\ 1 & 238 & 0 \\ -3 & -3942 & 7 \end{vmatrix} = -31\,546; \quad A_2 = \begin{vmatrix} 8 & 1 & 7794 \\ 1 & 5,5 & 238 \\ -3 & 0 & -3942 \end{vmatrix} = -41\,619.$$

По ним рассчитываются три оценки (например, $b_1 = -31\,546/252 = -125$) и строится двухфакторная модель

$$\hat{\rho} = 928 - 125x_1 = 165x_2, \quad (\text{III.120})$$

для которой расчетные значения $\hat{\rho}$ в экспериментальных точках и остаточная сумма квадратов $SS_{\text{ост}} = 2374$ приведены в табл. III.20. Остаточная дисперсия $SS_{\text{ост}} = 2374/(8 - 3) = 475$ при $f_{\text{ост}} = 5$. Ошибка $s_{\text{ост}} = 21,8$ МПа, что по абсолютному значению несколько ниже, чем в пр. III.24, но относительная ошибка осталась практически неизменной $= 2,3\%$.

Ковариационная матрица $\mathbf{D} = \mathbf{x}^T \mathbf{x}^{-1}$ (III.115) определяется только матрицей независимых переменных \mathbf{x} , т. е. планом эксперимента (см. табл. III.1), поэтому априорный анализ ее свойств лежит в основе синтеза этих планов (см. п. III.14).

Элементы c_{ij} ковариационной матрицы определяются как отношение алгебраических дополнений A_{ij} (см. прил. I.1) элементов (ij) в матрице моментов $\mathbf{M} = \mathbf{x}\mathbf{x}^T$ (III.108) к определителю этой матрицы:

$$\mathbf{D} = \begin{pmatrix} c_{00} & \dots & c_{0i} & \dots & c_{0j} & \dots & c_{0k} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ c_{i0} & \dots & c_{ii} & \dots & c_{ij} & \dots & c_{ik} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ c_{k0} & \dots & c_{ki} & \dots & c_{kj} & \dots & c_{kk} \end{pmatrix}, \quad (\text{III.121})$$

$$c_{ij} = A_{ij}/\det \mathbf{M} = A_{ij}/A. \quad (\text{III.122})$$

Для однофакторной модели с учетом (III.118) матрица \mathbf{D} имеет вид

$$\mathbf{D} = \begin{pmatrix} c_{00} & c_{01} \\ c_{10} & c_{11} \end{pmatrix} = \begin{pmatrix} (11)/A & -(01)/A \\ -(10)/A & (00)/A \end{pmatrix}. \quad (\text{III.123})$$

Матрица \mathbf{D} квадратная, ее размер $(k + 1) \times (k + 1)$; она симметрична (см. прил. I.1). По элементам c_{ii} и c_{ij} можно вычислить оценки коэффициентов b_i по формуле

$$b_i = \sum_{j=0}^k c_{ij} (jY), \quad i = \overline{0, k}, \quad (\text{III.124})$$

которая для однофакторной модели дает результат, тождественный (III.119):

$$\begin{aligned} b_0 &= c_{00}(0Y) + c_{01}(1Y); \\ b_1 &= c_{10}(0Y) + c_{11}(1Y). \end{aligned} \quad (\text{III.125})$$

Из (III.121) — (III.125) следует важный инженерный вывод о том, что процедура определения оценок b_i очень сильно упрощается, если все *внедиагональные элементы* $c_{ij} = 0$, т. е. матрица \mathbf{D} будет *диагональной*. Для этого исходная матрица-план \mathbf{x} должна быть *ортогональной* (прил. I.1). Одним из условий этого является равенство нулю сумм (ij) в матрице моментов \mathbf{M} (III.107).

Последнее, в частности, достигается за счет симметричного расположения опытных точек в нормализованном факторном пространстве x .

Пример III.26. Матрицы M для условий пр. III.24 и III.25 имеют вид и определители соответственно:

$$M = \begin{pmatrix} 5 & 0 \\ 0 & 2,5 \end{pmatrix}, \quad \det M = A = 12,5; \quad (\text{III.126})$$

$$M = \begin{pmatrix} 8 & 1 & -3 \\ 1 & 5,5 & 0 \\ -3 & 0 & 7 \end{pmatrix}, \quad \det M = A = 251,5. \quad (\text{III.127})$$

Согласно (III.121), ковариационные матрицы для (III.126) — (III.127) соответственно

$$D = \begin{pmatrix} 2,5/12,5 & 0 \\ 0 & 5/12,5 \end{pmatrix} = \begin{pmatrix} 0,2 & 0 \\ 0 & 0,4 \end{pmatrix}; \quad (\text{III.128})$$

$$D = \begin{pmatrix} 0,1531 & -0,0278 & 0,0656 \\ -0,0278 & 0,1869 & -0,0119 \\ 0,0656 & -0,0119 & 0,1710 \end{pmatrix}. \quad (\text{III.129})$$

Пять уровней нормализованного фактора x_1 в пр. III.24 симметричны относительно центра эксперимента, поэтому матрица D (III.128) диагональна, что приводит к элементарным формулам, дающим тождественные (III.105) результаты:

$$\left. \begin{aligned} b_0 &= c_{00}(0Y) = (11)(0Y)/A = 0,2 \cdot 5470 = 1094; \\ b_1 &= c_{11}(1Y) = (00)(1Y)/A = 0,4 \cdot (-332) = -133 \end{aligned} \right\} \quad (\text{III.130})$$

Матрица D (III.129) позволяет по (III.124) рассчитать b_0 , b_1 и b_2 , тождественные результату в пр. III.25:

$$\begin{aligned} b_0 &= 0,1531 \cdot 7794 - 0,0278 \cdot 238 + 0,0656(-3942) = 928; \\ b_1 &= -0,0278 \cdot 7794 + 0,1869 \cdot 238 - 0,0119(-3952) = -125; \\ b_2 &= 0,0656 \cdot 7794 - 0,0119 \cdot 238 + 0,1710(-3942) = -165. \end{aligned}$$

Элементы D -матриц рассчитываются на ЭВМ по типовым программам линейной алгебры. Для стандартных планов эксперимента x значения c_{ij} приводятся в каталогах [58 и др].

III.8. Основные этапы регрессионного анализа линейных моделей

Метод наименьших квадратов (МНК) как *алгебраический инструмент* построения линейных моделей (III.88) позволяет вычислить точечные оценки коэффициентов b_i ($i = 0, k$) с точностью до любой значащей цифры. Однако эти оценки $b_i = \Theta_i^*$ являются *случайными* величинами, интервал распределения которых с известным риском α накрывает значения истинных коэффициентов $\beta_i = \Theta$. Таким образом, к оценкам b_i следует применить вышеизложенный (см. п. III.3) аппарат проверки статистических гипотез. Следовательно, построение ЭС-модели можно считать законченным, а саму модель использовать для принятия инженерных решений только после того, как алгебраическая процедура вычисления оценок и невязок Δ_u будет дополнена *статистическим анализом* как отдельных коэффициентов, так и модели в целом.

В экспериментах исследователь меняет факторы целенаправленно и старается удерживать их на выбранном уровне, следовательно, левая часть информационной табл. III.1, образующая план x , детерминирована. При этом стохастичность системы (см. п. III.1) находит отражение только в ее откликах — результаты эксперимента y_u суть случайные величины, распределенные по одномерному закону. Использование в такой ситуации корреляционного анализа не корректно, поскольку Y и x не образуют $(k + 1)$ -мерной нормально распределенной совокупности (для Y и x_1 , в частности, не образуется колоколообразная поверхность — рис. III.8, а), следовательно, коэффициенты корреляции $r\{x_i, Y\}$ и соответствующий им риск α не могут иметь предполагаемого смысла.

Для проверки гипотез о форме и степени влияния x на Y необходимо использовать идеи и методы регрессионного анализа [11 и др.]. Их применение будет корректным, если правдоподобны предположения, во-первых, о взаимонезависимости результатов v -х опытов, во-вторых, о нормальности распределения частных результатов при любом наборе x , в-третьих, о постоянстве дисперсии σ_y^2 во всей области эксперимента или об изменении ее по известной функции $D\{Y\} = \sigma_y^2 h^2$ (ср. с предпосылками корреляционного анализа в п. III.4).

Регрессионный анализ целесообразно начинать с проверки гипотез о равенстве нулю отдельных истинных коэффициентов β_i , образующих вектор β с первоначальным размером $k + 1$. Если такая нуль-гипотеза будет принята, то с риском α можно исключить незначимый эффект $b_i \rightarrow 0$ из модели, считая его появление обусловленным ошибками эксперимента. При этом форма модели упрощается, а главное, закономерно сокращается содержащийся в ней «информационный шум».

Вектор β равен математическому ожиданию вектора оценок b (III.106), т. е. $\beta = E\{b\}$; матрицу математического ожидания разности $(b - \beta)$ можно записать так:

$$\begin{aligned}
 E\{(b - \beta)(b - \beta)^T\} &= E\left\{\begin{pmatrix} b_0 - \beta_0 \\ b_1 - \beta_1 \\ \vdots \\ b_k - \beta_k \end{pmatrix} (b_0 - \beta_0, b_1 - \beta_1, \dots, b_k - \beta_k)\right\} = \\
 &= E\begin{pmatrix} (b_0 - \beta_0)^2 & (b_0 - \beta_0)(b_1 - \beta_1) & \dots & (b_0 - \beta_0)(b_k - \beta_k) \\ (b_1 - \beta_1)(b_0 - \beta_0) & (b_1 - \beta_1)^2 & \dots & (b_1 - \beta_1)(b_k - \beta_k) \\ (b_2 - \beta_2)(b_0 - \beta_0) & (b_2 - \beta_2)(b_1 - \beta_1) & \dots & (b_2 - \beta_2)(b_k - \beta_k) \\ \vdots & \vdots & \ddots & \vdots \\ (b_k - \beta_k)(b_0 - \beta_0) & (b_k - \beta_k)(b_1 - \beta_1) & \dots & (b_k - \beta_k)^2 \end{pmatrix} = \\
 &= \begin{pmatrix} \sigma^2\{b_0\} & \text{cov}\{b_0 b_1\} & \dots & \text{cov}\{b_0 b_k\} \\ \text{cov}\{b_1 b_0\} & \sigma^2\{b_1\} & \dots & \text{cov}\{b_1 b_k\} \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}\{b_k b_0\} & \text{cov}\{b_k b_1\} & \dots & \sigma^2\{b_k\} \end{pmatrix} =
 \end{aligned}$$

$$= \begin{pmatrix} c_{00} & c_{01} & \dots & c_{0k} \\ c_{10} & c_{11} & \dots & c_{1k} \\ \dots & \dots & \dots & \dots \\ c_{k0} & c_{k1} & \dots & c_{kk} \end{pmatrix} \sigma_3^2 = \mathbf{D}\sigma_3^2. \quad (\text{III.131})$$

Из (III.131) следует, что при известной оценке дисперсии эксперимента s_3^2 *диагональный* элемент c_{ii} задает значения *дисперсии* оценки b_i :

$$s^2 \{b_i\} = c_{ii} s_3^2, \quad (\text{III.132})$$

а *внедиагональный* c_{ij} — *ковариацию* пары коэффициентов b_i и b_j , которая позволяет вычислить *парный коэффициент корреляции* оценок

$$\rho \{b_i b_j\} = \text{cov} \{b_i b_j\} / (\sigma \{b_i\} \sigma \{b_j\}) = c_{ij} / \sqrt{c_{ii} c_{jj}}. \quad (\text{III.133})$$

Оценки b_i и b_j *независимы* ($\rho \{b_{ij}\} = 0$), если *внедиагональные* элементы c_{ij} равны *нулю*.

Нуль-гипотеза $H_0: \beta_i = 0$ будет отклонена, а оценка b_i оставлена как *значимая* в модели в том случае, если *индивидуальные доверительные интервалы*

$$\left. \begin{aligned} b_i - t_\alpha s \{b_i\} \leq \beta_i < b_i + t_\alpha s \{b_i\}, \\ b_i - t_\alpha s_3 \sqrt{c_{ii}} \leq \beta_i < b_i + t_\alpha s_3 \sqrt{c_{ii}} \end{aligned} \right\} \quad (\text{III.134})$$

не накрывают нуль; последнее можно сформулировать как *превышение* с риском α оценкой b_i *критического уровня*

$$|b_i| > b_{i,\text{кр}} = t_\alpha s_3 \sqrt{c_{ii}}. \quad (\text{III.135})$$

Из (III.135) следует, что для значимых оценок *фактический уровень t -критерия* должен быть выше критического

$$t_{\text{факт}} = \frac{|b_i|}{s_3 \sqrt{c_{ii}}} > t_{\alpha,\text{кр}}. \quad (\text{III.136})$$

Выше критического должен быть и *уровень «гауссовской точности»* [6] коэффициентов $h\{b_i\}$, если они значимы с риском α :

$$|h \{b_i\}| = \frac{|b_i|}{\sqrt{c_{ii}}} > h_{\text{кр}} = (t_\alpha s_3)_{\text{кр}}. \quad (\text{III.137})$$

В выражениях (III.134) — (III.137) фигурирует уровень риска α ; он будет соответствовать заданному значению только при определении интервалов существования β_i для независимых оценок ($\rho \{b_i b_j\} = 0$). Поэтому такие незначимые оценки можно исключать без пересчета оставшихся b_j .

В остальных случаях ($\rho \{b_i b_j\} \neq 0$) заданное значение α не будет соответствовать истинному риску и возможна лишь сравнительная оценка значимости оценок коэффициентов по значению $t_{\text{факт}}$ (III. 136) или по гауссовской точности оценок $h \{b_i\}$ (III.

137). Сравнительный анализ $h\{b_i\}$, если и не позволяет оценить точно риск α , то допускает весьма эффективную *очистку* модели от *лишних оценок* при пошаговом последовательном регрессионном анализе (см. III.9). Для всех $k + 1$ коэффициентов b_i , связанных между собой, корректно можно оценить лишь совместную *доверительную* (при $p = 1 - \alpha$) *область*, которая ограничена гиперэллипсом рассеяния оценок параметров. Это понятие активно используется далее при анализе оптимальности планов эксперимента (см. п. III.10).

Пример III.27. На основании ковариационных матриц, определенных в пр. III.26, можно проверить гипотезу о равенстве нулю истинных значений параметров моделей плотности гипсового композита (III.105) и (III.120). При этом учитывается информация [6] о том, что ошибка эксперимента $s_g = 16$ кг/м³ при $f_g = 50$; по прил. II.5 $t_{0,1} = 1,676$, следовательно, критическая гауссовская точность $h_{кр} = (ts_g)_{кр} = 1,676 \cdot 16 = 26,8$.

Для однофакторной модели (III.105), у которой $\rho\{b_0 b_1\} = 0$,

$$\left. \begin{aligned} h\{b_0\} &= b_0/\sqrt{c_{00}} = 1094/\sqrt{0,2} = 2447 \gg h_{кр} = 26,8, \\ h\{b_1\} &= b_1/\sqrt{c_{11}} = -132/\sqrt{0,4} = 210 \gg h_{кр} = 26,8, \end{aligned} \right\} \quad (\text{III.138})$$

следовательно, все оценки значимы, а истинные параметры имеют корректно оцениваемые при $\alpha = 0,1$ доверительные интервалы

$$\begin{aligned} P\{1094 - 1,676 \cdot 16\sqrt{0,2} \leq \beta_0 \leq 1094 + 1,676 \cdot 16\sqrt{0,2}\} &= 1 - 0,1; \\ P\{1082 \leq \beta_0 \leq 1106\} &= 0,8; \\ P\{-150 \leq \beta_1 \leq -116\} &= 0,9. \end{aligned}$$

Для двухфакторной (трехпараметрической) модели (III.120) вычисленные, согласно (III.133) и (III.129), коэффициенты корреляции отличны от нуля:

$$\begin{aligned} r\{b_0 b_1\} &= -0,028/\sqrt{0,153 \cdot 0,187} = -0,166; \\ r\{b_0 b_2\} &= +0,408; \quad r\{b_1 b_2\} = -0,067. \end{aligned}$$

Гауссовская точность коэффициентов

$$\begin{aligned} h\{b_0\} &= 928/\sqrt{0,153} = 2372 \gg h_{кр}; \\ h\{b_1\} &= 289 \gg h_{кр}, \quad h\{b_2\} = 399 \gg h_{кр} \end{aligned}$$

позволяет отклонить гипотезу $H_0: \beta_i = 0$ для всех параметров модели; сравнительный анализ $h\{b_i\}$ показывает, что влияние второго фактора, судя по модели (III.120), сильнее, чем первого.

На втором этапе регрессионного анализа необходимо для модели со всеми значимыми коэффициентами проверить *адекватность* (лат. *adaequatus* — приравненный) экспериментальным данным, по которым она построена. При этом используются минимизируемая МНК остаточная сумма квадратов $SS_{ост} = \sum_u \Delta_u^2$ (III.92), соответствующее ей число степеней свободы модели $f_{ост}$ (как разность между числом опытов и числом значимых коэффициентов модели!) и вычисленная по ним остаточная дисперсия $s_{ост}^2 = SS_{ост}/f_{ост}$.

Если модель построена по информационной таблице, в которой нет повторяющихся строк, то статистические характеристики не-

адекватности модели совпадают с остаточными величинами, т.е.

$$SS_{\text{на}} = SS_{\text{ост}}; f_{\text{на}} = f_{\text{ост}}; s_{\text{на}}^2 = SS_{\text{на}}/f_{\text{на}} = s_{\text{ост}}^2. \quad (\text{III.139})$$

Если повторяющиеся строки есть, то с помощью однофакторного дисперсионного анализа выделяется сумма квадратов $SS_{\text{э}}$, связанная с повторением опытов и имеющая число степеней свободы $f_{\text{э}}$, и соответствующие величины в (III.139) уменьшаются:

$$SS_{\text{на}} = SS_{\text{ост}} - SS_{\text{э}}; f_{\text{на}} = f_{\text{ост}} - f_{\text{э}}; \\ s_{\text{на}}^2 = SS_{\text{на}}/f_{\text{на}}. \quad (\text{III.140})$$

Проверка адекватности сводится к проверке нуль-гипотезы $H_0: \sigma_{\text{на}}^2 = \sigma_{\text{э}}^2$ о равенстве истинной дисперсии $\sigma_{\text{на}}^2$, обусловленной неадекватностью модели, и истинной дисперсии $\sigma_{\text{э}}^2$, соответствующей совокупности экспериментальных величин в любой i -й точке. Эта же гипотеза в инженерных терминах может быть записана так: точность результатов \hat{y} , рассчитанных по модели, в среднем по области эксперимента не хуже, чем точность экспериментальных данных. Гипотеза проверяется по F -критерию (III.70)

$$F_a = s_{\text{на}}^2/s_{\text{э}}^2. \quad (\text{III.141})$$

Если $F_a < F_T(f_{\text{на}}, f_{\text{э}})$ (обычно при $\alpha = 0,05$ — см. прил. II.6), то принимается гипотеза адекватности модели экспериментальным данным. Если $F_a > F_T$, то модель признается неадекватной и требуется дальнейшее исследование с целью уточнения описания поведения системы (выбор нелинейной модели, расширение информационной таблицы, преобразование факторов и откликов — см., в частности п. III.9).

Проверка адекватности по F_a -критерию (III.141) корректна со статистической точки зрения, однако к ее результатам следует относиться разумно с технико-экономических позиций, на которых учитывается содержательная сторона задачи. Возможны две граничных ситуации, в которых решение по (III.141) оказывается недостаточно полным. Во-первых, предназначенная для инженерных расчетов модель может оказаться неадекватной из-за малой ошибки опытов $s_{\text{э}}$ (прецизионные лабораторные измерения), но абсолютное значение $s_{\text{на}}$ удовлетворяет исследователя; тогда допустимо заменить $s_{\text{э.лаб}}^2$ на характерную для производства величину $s_{\text{э.пр}}^2 > s_{\text{э.лаб}}^2$ и вновь проверить адекватность. Во-вторых, полученная в условиях грубого эксперимента модель адекватна, но по значению $s_{\text{на}}$ не удовлетворяет исследователя, поэтому ее следует усовершенствовать для улучшения точности описания поведения системы.

На третьем этапе с помощью F -критерия проверяется и *информационная способность модели*. Однако задача в этом случае принципиально отлична от предыдущей, так как необходимо отклонить нуль-гипотезу о равенстве исходного рассеяния случайной величины (по отношению к общему среднему $\bar{Y}_{\text{общ}}$ всех N результатов)

и рассеяния, связанного с неадекватностью модели, т. е.
 $H_0: \sigma_{\text{общ}}^2\{Y\} = \sigma_{\text{на}}^2$. Если такая гипотеза по критерию $F_{\text{и}}$

$$F_{\text{и}} = \frac{s_{\text{общ}}^2\{Y\}}{s_{\text{на}}^2} \quad (\text{III.142})$$

будет отклонена ($F_{\text{и}} > F_{\text{T}}\{f_{\text{общ}}, f_{\text{на}}\}$) при $\alpha = 0,05$ (см. прил. II.6), то линейная модель описывает результаты *лучше*, чем простейшая модель-константа $\hat{Y} = \bar{Y}_{\text{общ}} = \text{const}$ при любом наборе x_i . В противном случае ($F_{\text{и}} \leq F_{\text{T}}$, следовательно, $\sigma_{\text{общ}}^2\{Y\} = \sigma_{\text{на}}^2$) линейная модель не имеет никакой информационной ценности. Количественно мера такой ценности может быть [4] оценена по показателю $\Theta_{\text{и}} = 100(\sqrt{F_{\text{и}}} - 1)$. Если $\Theta_{\text{и}} < 30 \dots 50$, то разумно такой моделью не пользоваться. Ряд авторов [28] предлагают еще более жесткий уровень информационной ценности модели для того, чтобы избежать применения сложных, но бессодержательных моделей.

Пример III.28. Однофакторная модель (III.105) по данным пр. III.24 характеризуется (рис. III.14): общим рассеянием пяти результатов $s_{\text{общ}}^2\{Y\} =$

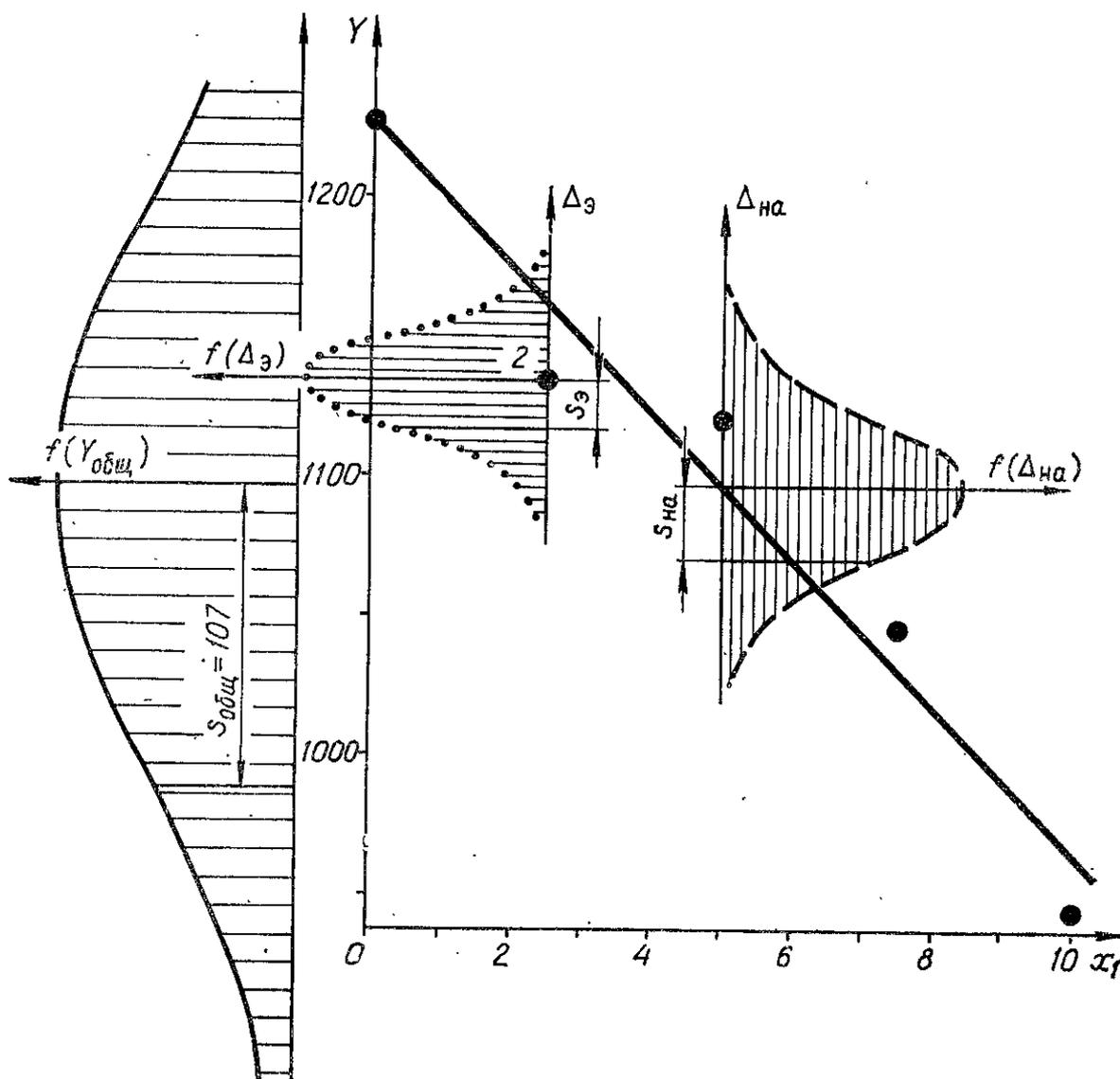


Рис. III. 14. Проверка адекватности и информационной способности модели по F -критерию

$$= \frac{1}{4} \left\{ \sum_u [y_u - (0Y)/5]^2 \right\} = 11\,500$$
 при $f_{\text{общ}} = 4$; остаточной дисперсией $s_{\text{ост}}^2 = 623$ при $f_{\text{ост}} = 23$, которая совпадает с $s_{\text{на}}^2$, поскольку в табл. III.19 нет повторяющихся опытов; ошибкой эксперимента $s_{\text{э}} = 16$ кг/м³ при $f_{\text{э}} = 50$ (пр. III.27); все характеристики рассеяния показаны (при гипотезе нормальности распределения ρ) на рис. III.14.

Модель адекватна эксперименту, поскольку $F_a = 623/16 = 2,43$ меньше чем $F_{0,05} = 2,8$ ($f_1 = 3, f_2 = 50$) — прил. II.6. Модель информативна, поскольку критерий $F_{\text{и}} = 11500/623 = 18,46$ больше чем $F_{0,05} = 9,12$ ($f_1 = 4, f_2 = 3$). Мера информативности $\Theta_{\text{и}} = 100(\sqrt{18,46} - 1) = 330\%$ позволяет сделать заключение о полезности модели.

На четвертом этапе регрессионного анализа оценивается *прогностическая* (греч. *pro* — вперед, *gnōsis* — значение) *способность* ЭС-модели.

Предсказанное значение отклика в точке с координатами \mathbf{x}_p k -мерного факторного пространства определяется после подстановки этих координат в линейное уравнение $\hat{Y} = \mathbf{x}^T \mathbf{b}$ (III.87), которое с учетом (III.114) явно зависит от ковариационной матрицы \mathbf{D} :

$$\hat{Y}_p = \mathbf{x}_p^T \mathbf{b} = \mathbf{x}_p^T \mathbf{D} \mathbf{x}^T \mathbf{y}. \quad (\text{III.143})$$

Дисперсия предсказанного значения $\sigma^2(\hat{Y})$ для линейной модели (III.87) находится как дисперсия функции случайных величин — оценок параметров b_i , которые определяются с дисперсиями $\sigma^2\{b_i\}$ и ковариациями $\text{cov}\{b_i b_j\}$ (III.131):

$$\begin{aligned} \sigma^2\{\hat{Y}\} &= \sum_{i=0}^k \left(\frac{\partial Y}{\partial b_i}\right)^2 \sigma^2\{b_i\} + 2 \sum_{i < j} \text{cov}\{b_i b_j\} \left(\frac{\partial Y}{\partial b_i}\right) \left(\frac{\partial Y}{\partial b_j}\right) = \\ &= \sigma^2\{b_0\} + \sum_{i=1}^k \sigma^2\{b_i\} x_i^2 + 2 \sum_{i < j} \text{cov}\{b_i b_j\} x_i x_j. \end{aligned} \quad (\text{III.144})$$

Выражение (III.144) в матричной форме — общая функция для оценки (в связи с заменой σ^2 на оценку s^2) дисперсии предсказания отклика

$$s^2\{\hat{Y}\} = \mathbf{x}_p^T \mathbf{D} \mathbf{x}_p s_{\text{э}}^2 = d s_{\text{э}}^2. \quad (\text{III.145})$$

Она содержит независимую от результатов эксперимента функцию

$$d = \mathbf{x}_p^T \mathbf{D} \mathbf{x}_p, \quad (\text{III.146})$$

свойства которой определяются матрицей плана \mathbf{x} , что позволяет анализировать ее априори и выбирать на ее основе те или иные планы эксперимента (см. п. III.10 — III.14).

Для оценки предсказания по линейной однофакторной модели (III.88) в точке с координатами $x_{0p} \equiv +1$ и x_{1p} с учетом (III.123) функция (III.145) конкретизируется:

$$\begin{aligned} s^2\{\hat{Y}\} &= (1, x_{1p}) \begin{pmatrix} (11)/A & -(01)/A \\ -(10)/A & (00)/A \end{pmatrix} \begin{pmatrix} 1 \\ x_{1p} \end{pmatrix} s_3^2 = \\ &= A^{-1} [(11) - 2(10)x_{1p} + (00)x_{1p}^2] s_3^2 = ds_3^2. \end{aligned} \quad (\text{III.147})$$

Функция d как мера точности прогноза в (III.147) — квадратичная парабола, поэтому дисперсия предсказания возрастает по мере удаления от центра эксперимента, причем скорость этого роста зависит от того, как был организован эксперимент, т. е. от свойств плана x .

Математическое ожидание предсказанного значения $E(\hat{Y})$ подчиняется t -распределению, поэтому доверительные интервалы определяются соотношениями типа (III.38):

$$P\{\hat{Y}_p - t\{\alpha/2, f_3\} s\{\hat{Y}_p\} \leq E\{\hat{Y}\} < \hat{Y}_p + t\{\alpha/2, f_3\} s\{\hat{Y}_p\}\} = 1 - \alpha; \quad (\text{III.148})$$

$$P\{\hat{Y}_p - t\{\alpha/2, f_3\} s_3 \sqrt{d} \leq E\{\hat{Y}\} < \hat{Y}_p + t\{\alpha/2, f_3\} s_3 \sqrt{d}\} = 1 - \alpha. \quad (\text{III.149})$$

Пример III.29. Функция $s^2\{\hat{Y}\}$ для однофакторной модели (III.105) с учетом результатов пр. III.24 записывается, согласно (III.147), как

$$s^2\{\hat{\rho}\} = (2,5 - 2 \cdot 0 \cdot x_{1p} + 5x_{1p}^2) 16^2/12,5 = 51,2 + 102,4x_{1p}^2.$$

График функции оценки квадратической ошибки предсказания $s\{\hat{\rho}\} = \sqrt{51,2 + 102,4 x_{1p}^2}$, совпадающей по размерности с ρ , кг/м³, показан на рис. III.15, а. По мере выхода за границы изученного факторного пространства $|x_1| \leq 1$ значение $s\{\hat{Y}\}$ резко возрастает.

На этом же рис. III.15, а показаны доверительные границы $\hat{\rho}$ при $\alpha = 0,1$ (см. прил. II.5, $f_3 = 50$, $t = 1,676$), построенные вдоль прямой, описываемой моделью III.105. Эти интервалы относятся к каждому фиксированному значению фактора x_1 . Доверительная зона для линии регрессии в целом строится по другим правилам [11, 23].

Пример III.30. Функция $s^2\{\hat{Y}\}$ для двухфакторной модели (III.120) может быть (с учетом результатов пр. III.26 и при условии $x_0 \equiv +1$) получена из (III.144) как

$$\begin{aligned} s^2\{\hat{\rho}\} &= s^2\{b_0\} x_0^3 + s^2\{b_1\} x_1^2 + s^2\{b_2\} x_2^2 + 2 \text{cov}\{b_0 b_1\} x_0 x_1 + \\ &+ 2 \text{cov}\{b_0 b_2\} x_0 x_2 + 2 \text{cov}\{b_1 b_2\} x_1 x_2 = s_3^2 (c_{00} + c_{11} x_1^2 + c_{22} x_2^2 + 2c_{01} x_1 + \\ &+ 2c_{02} x_2 + 2c_{12} x_1 x_2) = 16^2 [0,153 + 0,187 x_1^2 + 0,171 x_2^2 + 2(-0,028) x_1 + \\ &+ 2 \cdot 0,066 x_2 + 2(-0,012) x_1 x_2] = 39,2 + 47,9 x_1^2 + 43,8 x_2^2 - 14,3 x_1 + \\ &+ 33,8 x_2 - 6,1 x_1 x_2. \end{aligned} \quad (\text{III.150})$$

Если изолинии модели $\hat{\rho}$ (III.20) — параллельные прямые (рис. III.15, б), то изолинии $s^2\{\hat{\rho}\}$ — эллипсы с общим центром в точке $x_{1s} = 0,11$ и $x_{2s} = -0,35$, где $s_{\min}^2\{\hat{\rho}\} = 32$ (рис. III.15, в). Изолинии верхней границы $\hat{\rho}_{\alpha, в}$

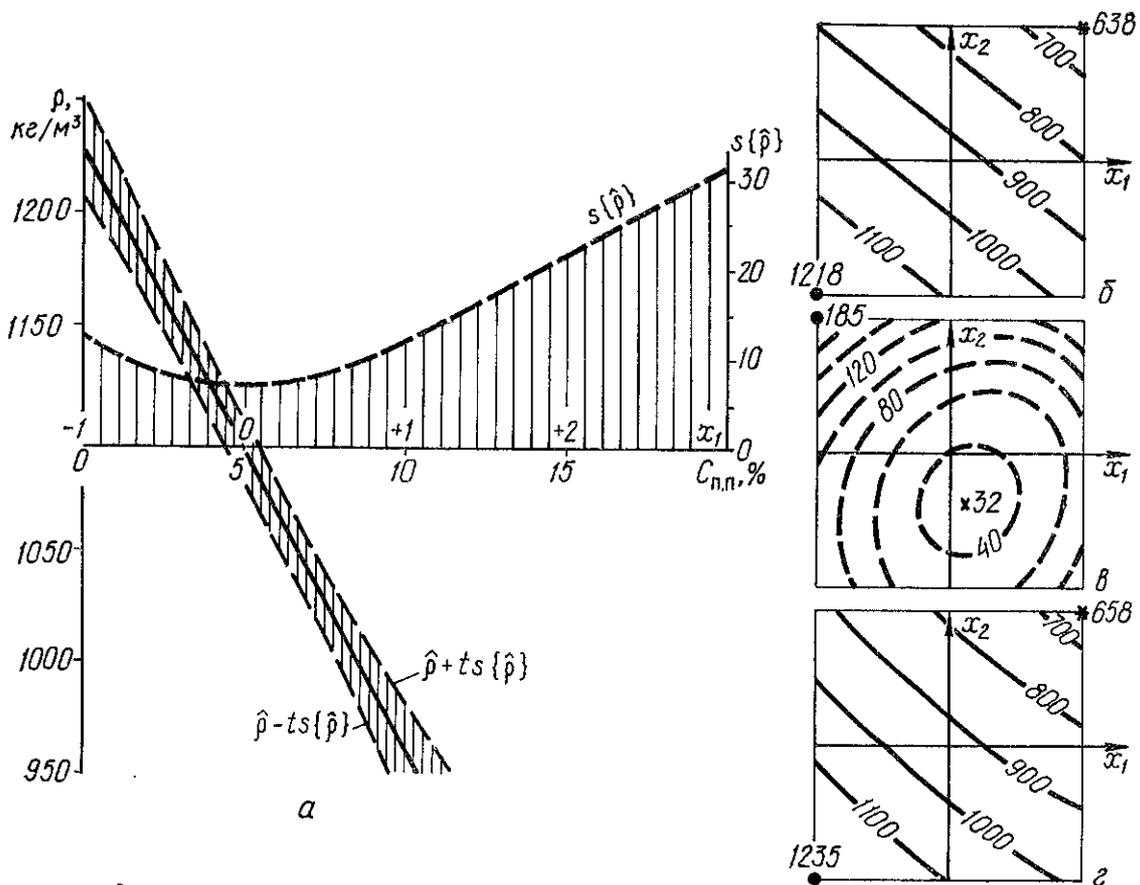


Рис. III. 15. Ошибка предсказания двухфакторной линейной модели:
 а — доверительный интервал для $\hat{\rho} = f(x_1)$ и график функции $s^2\{\hat{\rho}\} = f(x_1)$; б — изолинии ρ ;
 в — изолинии $s^2\{\hat{\rho}\}$; г — изолинии верхней границы $\hat{\rho}$

(рис. III.15, г) истинных значений ρ (с риском $\alpha = 0,1$ при $t = 1,676$) рассчитаны как

$$\hat{\rho}_{\alpha, в} = 928 - 125x_1 - 165x_2 + 1,676 \sqrt{s^2\{\hat{\rho}\}}. \quad (III.151)$$

По сравнению с рис. III.15, б одинаковые изолинии несколько изогнуты на краях области и сдвинуты к точке $x_1 = x_2 = +1$.

Основные положения регрессионного анализа линейных моделей правомочны и при анализе более сложных нелинейных моделей, хотя в последнем случае могут появляться и особенности, рассмотренные в п. III.9. Вышеизложенные положения регрессионного анализа образуют одну из основ теории планирования оптимальных экспериментов (см. п. III.10 и др.).

III.9. Построение и статистический анализ нелинейных моделей [линейных по параметрам]

Среди обширного множества нелинейных моделей, применяемых в науке и технике, следует выделить достаточно многочисленную группу моделей, которые за счет преобразования действующих на систему k факторов \mathbf{x} могут быть представлены в линейной форме:

$$\begin{aligned} Y &= \beta_1 z_1 + \dots + \beta_i z_i + \dots + \beta_L z_L = \\ &= \beta_1 f_1(\mathbf{x}) + \dots + \beta_i f_i(\mathbf{x}) + \dots + \beta_L f_L(\mathbf{x}) = \\ &= \sum_{i=1}^L \beta_i f_i(\mathbf{x}) = \beta^T \cdot \mathbf{f}(\mathbf{x}). \end{aligned} \quad (III.152)$$

Система известных *базисных функций* $z_i = f_i(x_1, x_2, \dots, x_k)$ не зависит от неизвестных параметров модели β_i . Таким образом, нелинейная по отношению к рецептурно-технологическим, структурным, эксплуатационным и другим факторам (действующим на систему и учитываемым в количестве k исследователем) модель (III.152) является линейной по $L \leq k$ параметрам β_i . Последнее позволяет не только применить для построения ЭС-моделей

$$\hat{Y} = b_1 z_1 + \dots + b_i z_i + \dots + b_L z_L = \mathbf{b}^T \cdot \mathbf{f}(\mathbf{x}) \quad (\text{III.153})$$

весь вышеизложенный для моделей первого порядка аппарат (МНК, регрессионный анализ и др.— п. III.7 — III.8), но и обобщить его. Модели первого порядка — частный случай модели (III.153), когда вектор базисных функций $\mathbf{f}(\mathbf{x})$ упрощается до вектора действующих факторов \mathbf{x} .

Процесс выбора базисных функций $z_i = f_i(\mathbf{x})$ может сопровождаться и процессом преобразования выходов системы по функции $V = \varphi(Y)$ для упрощения модели и вычислений при ее построении. Полный процесс замены исходной функции $Y = F(X)$ на линейную $V = \psi(Z)$ называется *линеаризацией функций*.

В табл. III.21 приведен ряд наиболее часто встречающихся однофакторных функций, приводимых к прямой. Здесь же указаны необходимые преобразования для Y , K_0 , K_1 , и X , а также области существования $Y = f(X)$, ее характерные точки и отношение стандартных ошибок функции и аргумента, необходимое для регрессионного анализа (см. п. III.8) линеаризованных функций $V = a_0 + a_1 Z$.

Необходимо отметить некоторые особенности результатов линеаризации функции $Y = F(X)$ в функцию $V = \psi(Z)$. Во-первых, если исходные факторы X_i изменяли уровни через равные интервалы, то интервалы изменения ΔZ_i не будут постоянными; следует нормализовать Z_i (а не X_i) с тем, чтобы $|z_i| \leq 1$. Во-вторых, замена Y на $V = \varphi(Y)$ приводит к тому, что по МНК минимизируется не сумма квадратов отклонений $(y_u - \hat{y}_u)^2$, а сумма квадратов отклонений $(v_u - \hat{v}_u)^2$, что не тождественно. В-третьих, замена Y на V приводит к изменению дисперсии выхода модели, поскольку из (III.144) ясно, что $\sigma^2 \{V\} / \sigma^2 \{Y\} = (\partial V / \partial Y)^2$. Расчет оценок ведется по алгоритмам, показанным на рис. III.13.

Пример III.31. При исследовании технологических свойств композиций из полиэфирной смолы и минеральных порошков [8, 13] определена с помощью ротационного вискозиметра [52] эффективная вязкость смеси η , Па · с, при разных скоростях ее деформации $\dot{\epsilon}$, с⁻¹.

Результаты эксперимента с композицией на мраморном порошке средней крупности (удельная поверхность $S = 100$ м²/кг) при четырех скоростях деформации $\dot{\epsilon} = 3, 9, 27, 81$ с⁻¹ приведены в табл. III.22 (строка 2). Необходимо выбрать наиболее точную модель из веера четырех конкурирующих моделей — линейной, логарифмической, показательной и степенной:

$$\eta = K_0 + K_1 \dot{\epsilon}; \quad \eta = K_0 + K_1 \ln \dot{\epsilon}; \quad (\text{III.154}); (\text{III.155})$$

$$\eta = K_0 e^{K_1 \dot{\epsilon}}; \quad \eta = K_0 \dot{\epsilon}^{K_1}. \quad (\text{III.156}); (\text{III.157})$$

Таблица III.21. Однофакторные функции $Y = F(X)$, приводимые к прямой $V = a_0 a_1 Z$

Характеристика функции	Функция			
	Степенная	Показательная	Логарифмическая	Гиперболическая
	$Y = K_0 X^{K_1}$	$Y = K_0 e^{K_1 X}$	$Y = K_0 + K_1 \ln X$	$Y = K_0 + K_1 \frac{1}{X}$
V	$\ln Y$	$\ln Y$	Y	Y
a_0	$\ln K_0$	$\ln K_0$	K_0	K_0
a_1	K_1	K_1	K_1	K_1
Z	$\ln X$	X	$\ln X$	X^{-1}
Область существования	$0 < X < \infty, Y > 0 (K_0 > 0)$	$-\infty < X < \infty, Y > 0 (K_0 > 0)$	$0 < X < \infty, -\infty < Y < \infty$	$-\infty < X < \infty, K_0 < Y < \infty (K_0 > 0, K_1 > 0)$
Координаты характерных точек	$C \{1, K_0\}$	$C \{0, K_0\}$	$C \{1, K_0\}$	$C \{0, K_0\}$
Отношение $s\{Y\}/s\{X\}$	$K_0 K_1 X^{K_1-1}$	$K_0 K_1 e^{K_1 X}$	$\frac{K_1}{X}$	$\frac{K_1}{X^2}$
Графики				

Расчет параметров линейной модели (III.154) на основе экспериментальной информации (строки 1 и 2 в табл. III.22) проводится по (II.119) с помощью элементов информационной матрицы (0Y), (01), (11) и (1Y), рассчитанных в строках 2 — 5. Линейную модель (рис. III.16, а) нельзя признать удовлетворительной, так как оцененная по $SS_{\text{ост}}$ (строка 7) ошибка неадекватности $s_{\text{на}} = \sqrt{358/(4-2)} = 13,4$ Па · с весьма велика (около 67 % к b_0). Есть и вторая причина, по которой модель (III.154) следует отклонить, — в области эксперимента (в окрестностях точки 4) она предсказывает отрицательные значения вязкости, что лишено физического смысла.

В расчете параметров логарифмической модели (III.155), приведенной на рис. III.16, б, использованы строки 2, 8 — 11 табл. III.22, показательной (рис. III.16, в) — строки 3, 4, 14 и 15; степенной (рис. III.16, г) — строки 8 — 10 и 20. Соответствующие остаточные суммы квадратов обязательно рассчитываются для вязкости η (строки 13, 19 и 23) даже в тех случаях, когда по МНК минимизировались отклонения $V = \ln \eta$ (строка 17).

Сравнительный анализ веера моделей $\eta = F(\epsilon)$ показывает, что степенная модель (III.157) имеет наилучшую точность (ср. строки 7, 13, 19 и 23)

Таблица III.22. Результаты эксперимента и расчета вспомогательных величин для оценки параметров веера моделей $\eta = F(\varepsilon)$

Входы и выходы веера моделей	Номер операции	Операция	Номера опытов				Суммы (00) = 4	
			1	2	3	4		
Исходные данные	1	$\varepsilon = X_u$	3	9	27	81	(0Y) = 49,39	
	2	$\eta = y_u$	33,2	11,2	3,75	1,24		
Модель	линейная	3	x_u	-1	-0,846	-0,538	+1	(01) _x = -1,384
		4	x_u^2	1	0,7157	0,2894	1	(11) _x = 3,0051
		5	$x_u y_u$	-33,2	-9,475	-2,018	1,24	(1Y) _x = -43,453
		6	$\hat{y} = \hat{\varepsilon}$	19,18	17,57	14,36	-1,7	SS _{ост} {Y} = 358
		7	Δ_u^2	196,56	40,20	112,57	8,64	
	логарифми- ческая	8	$\ln \varepsilon = Z_u$	1,099	2,197	3,296	4,394	(01) _z = 0 (11) _z = 2,2218 (1Y) _z = -34,44
		9	z_u	-1	-0,333	+0,333	+1	
		10	z_u^2	1	0,1109	0,1109	1	
		11	$z_u y_u$	-33,2	-3,73	+1,25	+1,24	

Модель

	12	$\hat{y}_u = \hat{\varepsilon}_u$	27,85	17,52	7,18	-3,15	
	13	$\Delta_u^2 \{Y\}$	28,62	39,94	11,76	19,27	$SS_{\text{ocr}} \{Y\} = 100$
показательная	14	$\ln \eta = v_u$	3,502	2,416	1,322	0,215	$(0V) = 7,455$
	15	$x_u v_u$	-3,502	-2,044	-0,711	0,215	$(1V)_x = -6,042$
	16	\hat{v}_u	2,761	2,55	2,128	0,019	
	17	$\Delta_u^2 \{V\}$	0,5991	0,018	0,6496	0,0384	$SS_{\text{ocr}} \{V\} = 1,305$
	18	$\hat{\eta} = e^{v_u}$	15,82	12,81	8,4	1,02	
	19	$\Delta_u^2 \{Y\}$	302,06	2,59	21,72	0,04	$SS_{\text{ocr}} \{Y\} = 326$
	20	$z_u v_u$	-3,502	-0,805	+0,441	+0,215	$(1V)_z = -3,651$
	степенная	21	\hat{v}_u	3,507	2,412	1,316	0,221
22		$\eta = e^{v_u}$	33,35	11,16	3,73	1,25	
23		$\Delta_u^2 \{Y\}$	0,02	0	0	0	$SS_{\text{ocr}} \{Y\} = 0,02$

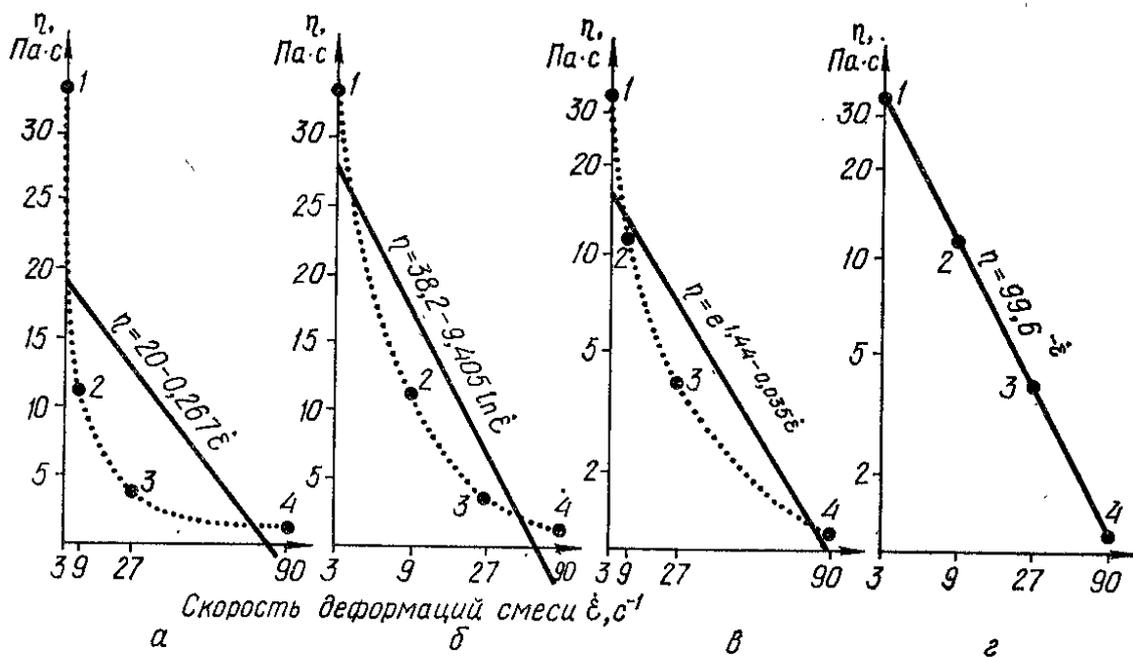


Рис. III.16. Векр четырех конкурирующих моделей для аппроксимации эмпирических точек 1 — 4:

а — линейная; б — логарифмическая; в — показательная; г — степенная

и обеспечивает физический смысл результатам моделирования ($\eta > 0$), совпадая по форме с известным в реологии уравнением Освальда — Вейля [52] для псевдопластичных жидкостей.

Однофакторные функции (табл. III.21) используются в качестве базисных функций в моделях (III.152), линейных по параметрам. Действительно, полиномиальная модель (III.4) порядка m для одного фактора будет получена при суммировании степенных базисных функций с возрастающим до m показателем степени

$$\begin{aligned} \tilde{Y} &= a_0 f_0(x_1) + a_1 f_1(x_1) + a_2 f_2(x_1) + \dots + a_m f_m(x_1) = \\ &= a_0 x_1^0 + a_1 x_1 + a_2 x_1^2 + \dots + a_m x_1^m. \end{aligned} \quad (\text{III.158})$$

В табл. III.23 показаны четыре достаточно распространенные в практике ЭС-моделирования функции, приводимые за счет линеаризации (по X и Y) к плоскости. Три из них приводятся к квадратичной параболе по X (вторая — при выходе $\ln Y$, а четвертая — при выходе Y^{-1}). Расчет оценок b_0 , b_1 и b_2 ведется по алгоритму на рис. III.13.

Пример III.32. Эффективная вязкость смеси (пр. III.31) при постоянной скорости деформаций $\dot{\epsilon}$ должна возрастать с увеличением степени наполнения X (отношение масс наполнителя и полимера Н/П). Это подтверждают результаты эксперимента, приведенные в табл. III.24. Предполагается, что зависимость $\eta = F(\text{Н/П})$ может быть описана с помощью экспоненциальной функции (табл. III. 24).

Параметры модели

$$\hat{Y} = \ln \hat{\eta} = 4,102 + 1,037x_1 - 0,608x_1^2 \quad (\text{III.159})$$

Таблица III.23. Однофакторные функции $Y = F(X)$, приводимые к плоскости $V = a_0 + a_1 Z_1 + a_2 Z_2$

Характеристика функции	Функция			
	Степенная	Показательная	Экспоненциальная	Гиперболическая
	$Y = K_0 + K_1 X + K_2 X^2$	$Y = K_0 X^{K_1} e^{K_2 X}$	$Y = e^{K_0 + K_1 X + K_2 X^2}$	$Y = \frac{1}{K_0 + K_1 X + K_2 X^2}$
A	Y	ln Y	ln Y	1/Y
a_0	K_0	ln K_0	K_0	K_0
a_1	K_1	K_1	K_1	K_1
Z_1	X	ln X	X	X
a_2	K_2	K_2	K_2	K_2
Z_2	X^2	X	X^2	X^2
Некоторые особенности функции	$-\infty < X < \infty,$ $-\infty < Y < \infty,$ $ K_{2.a} < K_{2.b} $	$X > 0,$ $Y > 0 (K_0 > 0),$ $ K_{1.a} < K_{1.b} $	$-\infty < X < \infty,$ $Y > 0,$ $ K_{1.a} > K_{1.b} $	$-\infty < X < \infty,$ $-\infty < Y < \infty,$ $\Delta_0 = -K_1/K_2,$ $\Delta^2 = \Delta_0^2 - K_0/K_2$
Графики				

рассчитываются по формулам, приведенным в пр. III.25 с учетом (02) = (01); эта же модель в натуральных переменных имеет вид

$$\eta = \exp\{-9,77 + 4,51(N/\Pi) - 2,43 (N/\Pi)^2\}$$

при $1,5 \leq N/\Pi \leq 2,5$. (III.160)

Ошибка неадекватности модели (III.159) $s_{на} = \sqrt{7059 \cdot 10^{-6} / (5 - 3)} = 0,059$ Па · с, или около 6 % по всей области эксперимента, а модель (III.160) — около 3 Па · с. Для интерпретации модель (III.160) менее удобна, чем в нормализованных факторах.

ЭС-модели с k факторами в виде полиномов (III.4) или (III.11) являются моделями, линейными по параметрам. Линеаризация происходит за счет введения базисных функций $f(x) = x_i^2$, $f(x) = x_i x_j$, $f(x) = x_i^3$ и т. п. Общее число эффектов $L\{b\}$ для полиномов порядка $m \leq 3$ при $k \leq 12$ показано в табл. III.25.

Таблица III.24. Результаты эксперимента и расчет вспомогательных величин для оценки параметров модели

$$\eta = \exp \left\{ K_0 + K_1 \frac{H}{\Pi} + K_2 \left(\frac{H}{\Pi} \right)^2 \right\}$$

Номер опыта	$H/\Pi = X$	$x_1 = \frac{X-2}{0,5}$	$x_2 = x_1^2$	$x_3^2 = x_1 x_2$	$x_4^2 = x_2^2$	η , Па · с	$Y = \ln \eta$	$x_1 Y$	$x_2 Y$	\hat{Y}	$\hat{\eta}$, Па · с
1	1,5	-1	1	-1	1	12	2,485	-2,485	2,485	2,457	12
2	1,75	-0,5	0,25	-0,125	0,0625	29	3,367	-1,684	0,842	3,432	31
3	2,0	0	0	0	0	62	4,127	0	0	4,102	60
4	2,25	+0,5	0,25	+0,125	0,0625	90	4,500	2,250	1,125	4,468	87
5	2,5	+1	1	+1	1	91	4,511	4,511	4,511	4,531	33
(00)	—	(01)	(02) = = (11)	(12)	(22)	—	(0Y)	(1Y)	(2Y) = = (11Y)	—	—
5	—	0	2,50	0	2,125	—	18,990	2,592	8,963	—	—

Таблица III.25. Число эффектов в линейных, квадратичных и кубических моделях

Число факторов k	1	2	3	4	5	6	7	8	9	10	11	12
--------------------	---	---	---	---	---	---	---	---	---	----	----	----

Линейная модель

Всего эффектов $L_{л.м} = 1 + k$	2	3	4	5	6	7	8	9	10	11	12	13
-------------------------------------	---	---	---	---	---	---	---	---	----	----	----	----

Неполная квадратичная модель

Число взаимодействий b_{ij} $L_{ij} = 0,5k(1 - k)$	0	1	3	6	10	15	21	28	36	45	55	66
Всего эффектов $L_{н.к.м} = 1 + k + 0,5k(1 - k)$	2	4	7	11	16	22	29	37	46	56	68	79

Квадратичная модель

Число квадратичных эффектов b_{ii} $L_{ii} = k$	1	2	3	4	5	6	7	8	9	10	11	12
Всего эффектов $L_{к.м} = 0,5(k + 1) \times (k + 2)$	3	6	10	15	21	28	36	45	55	66	79	91

Число факторов k	1	2	3	4	5	6	7	8	9	10	11	12
--------------------	---	---	---	---	---	---	---	---	---	----	----	----

Неполная кубическая модель

Число взаимодействий с квадратичным эффектом b_{ij} $L_{ij} = k(k-1)$	0	2	6	12	20	30	42	56	72	90		
Всего эффектов $L_{н.кб.м}^{ij}$	3	8	16	27	41	58	78	101				
Число взаимодействий тройных b_{ijk} $L_{ijk} = \frac{1}{6} k(k-1) \times (k-2)$	0	0	1	4	10	20	35	56	84			
Всего эффектов $L_{н.кб.м}^{ijk}$	3	6	11	19	31	48	71	101				
Всего эффектов $L_{н.кб.м}$	3	8	17	31	51	78						

Кубическая модель

Число эффектов b_{iii} $L_{iii} = k$	1	2	3	4	5	6	7	8	9	10	11	12
Всего эффектов $L_{кб.м}$	4	10	20	35	56	84						

Если число эффектов превышает $L\{b\} = 100$ (такие ситуации в табл. III.25 опущены), то полиномиальные ЭС-модели, вообще говоря, теряют инженерный смысл, во-первых, из-за трудностей интерпретации, во-вторых, из-за вычислительных трудностей оценки вектора (III.114), в-третьих, из-за сложности реализации экспериментов, состоящих из сотен различных опытов. Как правило, в задачах материаловедения и технологии трудно реализовать в едином исследовательском цикле число опытов более 60, поэтому в табл. III.25 выделена предельная граница числа эффектов $L \simeq \simeq 60$. Чем больше факторов хочет исследователь включить в модель, тем он вынужден сильнее упрощать ее форму. Если технолог собрал и глубоко проанализировал обширную априорную информацию об объекте исследования, то он может обоснованно включить в модель лишь некоторые из сложных эффектов (см. п. III.14), увеличив одновременно число факторов k .

Для облегчения интерпретации и решения других инженерных задач по многофакторным полиномиальным ЭС-моделям необходимо удалить из нее все эффекты с оценками, близкими к нулю.

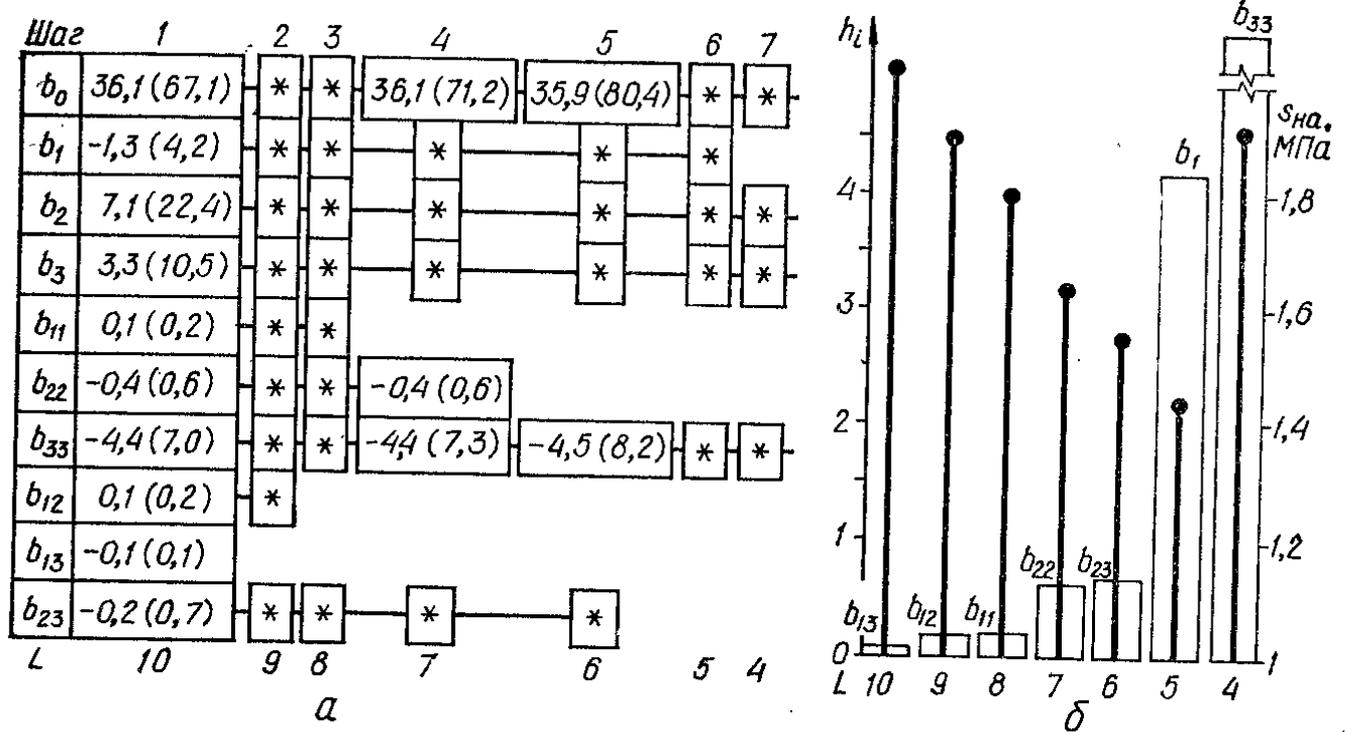


Рис. III. 17. Последовательный регрессионный анализ (ПРА):

a — изменение структуры модели; b — изменение минимальной гауссовской точности $h \{b_{\tau}\}$ (отрезки прямой) и ошибки неадекватности $s_{на}$ (столбчатая диаграмма)

Степень близости к нулю с заданным риском α определяется в ходе последовательного регрессионного анализа (ПРА) на каждом цикле которого [6]:

а) для модели с L эффектами вычисляются ковариационная матрица D , оценки коэффициентов $b_i (i = 0, L)$ дисперсия неадекватности $s_{на.L}^2$ модели, критерий F_α и гауссовская точность $h_i = b_i / \sqrt{c_{ii}}$ (III.137) каждой оценки коэффициентов;

б) выбирается наименьшая гауссовская точность $h_{i,min}$ и, если она меньше критического уровня $h_{кр} = t_\alpha s_\varepsilon$ (III.137), то соответствующий эффект исключается;

в) процедура возвращается к п. «а», но рассматривается модель с уменьшенным на единицу числом эффектов $L - 1$;

г) процедура прекращается в том случае, если $h_{i,min} > h_{кр}$ или если отношение дисперсий $F_\alpha = s_{на}^2 / s_\varepsilon^2$ возрастает выше критического уровня при $\alpha = 0,05$.

Дисперсию предсказанного значения отклика $s^2 \{\hat{Y}\}$ и другие характеристики точности модели (см. п. III.8) следует определять и анализировать после окончания ПРА по конечной, «очищенной» модели.

Пример III.33. При исследовании влияния на предел прочности бетона R , МПа, расходов активной минеральной добавки ($0 \leq X_1 \leq 30$ % от массы цемента), цемента ($250 \leq X_2 \leq 450$ кг/м³) и поверхностно-активной добавки ($0 \leq X_3 \leq 0,5$ % от массы цемента) получена для нормализованных переменных полная квадратичная модель с $L = 10$ членами. Оценки коэффициентов приведены на рис. III.17, a в столбце «Шаг 1»; рядом в скобках указана их гауссовская точность h_i . Критическое значение $h_{кр} = 2,13$ ($s_\varepsilon = 1,3$ МПа при $f_9 = \infty$, $t_{0,10} = 1,645$ — см. прил. II.5). При ПРА в со-

ответствии со значениями h_i удаляются $b_{13}, b_{12}, b_{11}, \dots$. Эта процедура показана на рис. III.17, а. Структура матрицы D такова, что линейные эффекты и эффекты взаимодействия оцениваются независимо, поэтому эти незначимые эффекты удаляются без пересчета остающихся оценок (на рис. III.17, а ячейка со «звездочкой» означает, что в ней должно быть записано значение эффекта на предшествующем шаге). Остановить ПРА целесообразно при получении модели с 5 эффектами

$$R = 35,9 - 1,3x_1 + 7,2x_2 + 3,3x_3 - 4,5x_3^2, \quad (\text{III.161})$$

для которой минимальная неадекватность $s_{\text{на}} = 1,43$ МПа; это соответствует критерию $F_a = 1,43^2/1,3^2 = 1,21 < F_{0,05} = 1,83$ (при $f_1 = 15 - 5 = 10$ и $f_2 = \infty$ — см. прил. II.6), свидетельствующему об адекватности модели. Если бы была бы удалена оценка b_1 , для которой $h_1 = 4,2 > h_{\text{кр}} = 2,13$, то модель стала бы неадекватной, так как $F_a = 1,91^2/1,3^2 = 1,16 > F_{0,05} = 1,79$ (при $f_1 = 15 - 4 = 11$ и $f_2 = \infty$). Модель (III.161) с заданным риском пригодна для решения инженерных задач.

В задачах интерпретации многофакторных ЭС-моделей весьма полезным является ПРА при *переменной возрастающей степени риска* α [4, 6, 23]. Сначала задается весьма малый уровень $\alpha \ll 0,001$, и оставшееся небольшое число эффектов интерпретируется как наиболее достоверное описание явления. По мере увеличения α растет число значимых эффектов; как следствие увеличивается количество выводов по модели, но достоверность их снижается вплоть до осторожных «равновозможных» выводов при $\alpha = 0,5$ (что, в частности, допустимо на поисковых этапах работ).

III.10. Основные принципы оптимального планирования экспериментов

Эксперимент как активное воздействие на объект с целью получения необходимого эффекта [4, 52] требует на свою реализацию некоторого расхода ресурсов: материальных (оборудование, приборы, образцы и т. п.), временных (период подготовки и проведения эксперимента, интерпретации его результатов, выработки рекомендаций по их использованию и т. п.), трудовых (привлечение исполнителей от подсобных рабочих до специалистов по методологии и организации исследований) и др.

Исходя из сформулированной во «Введении» простейшей классификации задач оптимизации, эксперимент может быть *оптимизирован* по двум граничным направлениям. Во-первых, если исследователь может израсходовать на эксперимент весь имеющийся в наличии ресурс, то естественно его стремление к тому, чтобы за это получить *максимально возможный объем* новой информации о поведении, структуре, надежности и других сторонах исследуемого объекта. Во-вторых, если исследователю задан уровень необходимого эффекта или объем информации о поведении объекта, то естественно его стремление к тому, чтобы достигнуть результата при *минимальном расходе ресурсов* (уменьшая количество опытных установок и образцов, сокращая сроки работы, не привлекая специалистов определенного уровня и т. п.).

Решение этих задач (или их компромиссного варианта) *оптимизации эксперимента*, вообще говоря, традиционно в истории науки и техники. Их постановка всегда приводила исследователей к необходимости как-то планировать свою экспериментальную деятельность с учетом уже накопленного опыта. В настоящее время, с одной стороны, в связи с требованиями ускоренного научно-технического прогресса в условиях хозрасчета резко возросла актуальность оптимального планирования эксперимента. С другой стороны, создание *математической теории эксперимента* (МТЭ) позволило рассматривать и решать проблемы такого планирования через *оптимальность математических моделей*, описывающих объекты в соответствии с целями исследования, проектирования и управления ими.

На основании изложенного в предыдущих параграфах гл. III описания методов оценки коэффициентов моделей, линейных по параметрам, регрессионного канала ЭС-моделей и элементов их инженерной интерпретации можно сделать ряд важных выводов, определяющих некоторые основные принципы оптимального планирования эксперимента.

Во-первых, из системы нормальных уравнений (III.103) для определения L неизвестных оценок b_i в модели (III.152), линейной по параметрам, следует, что необходимо и достаточно, чтобы матрица x (см. табл. III.1—III.3) содержала L разных строк. Следовательно, *минимальное число опытов в плане* $N_{\min} = L$ (так называемый насыщенный план), а *избыточность опытов* $N - L$ полезна не с алгебраической, а с *статистической* точки зрения.

Во-вторых, *ковариационная* матрица D (III.121) определяет не только численные значения вектора V (III.114) неизвестных оценок коэффициентов модели, но и *характеристики точности* модели как в целом, так и отдельных ее параметров. Диагональные элементы матрицы c_{ii} оценивают дисперсии оценок каждого коэффициента $s^2\{b_i\}$, следовательно, точность оценок будет различной, если эти элементы не равны между собой. Внедиагональные элементы c_{ij} определяют корреляцию (III.133) оценок коэффициентов модели и если $\rho\{b_i b_j\} \neq 0$, то работа исследователя с моделью усложнится. Так, трудно оценить риск при построении доверительных интервалов (III.134), включающих истинные значения отдельных параметров β . Кроме того, исключение (добавление) некоторого эффекта $b_i f_i(x)$ в ЭС-модель для уточнения формы связи между откликом системы и факторами x приводит к пересчету всех оставшихся оценок коэффициентов. Это не только усложняет вычислительные процедуры, но и, главное, не позволяет надежно управлять системой, ориентируясь на соотношение между эффектами в модели.

В-третьих, *ковариационная* матрица $D = (x^T x)^{-1}$ (III.115) не зависит от результатов эксперимента и, следовательно, может быть *исследована до его реализации*. Результаты такого анализа могут служить основой для *конструирования* матрицы плана x некоторым *оптимальным* (с точки зрения целей всего исследования системы)

образом до проведения опытов. Априори можно исследовать и меру точности предсказания выхода $d = \mathbf{x}^T \mathbf{D} \mathbf{x}$ (III.146), связанную только с расположением опытных точек, использовать результаты анализа для повышения точности модели и т. п.

Вследствие этих трех выводов, например, становится естественным желание построить такую модель, у которой все оценки коэффициентов независимы друг от друга. Формально, следуя (III.133), это положение записывается как $\rho \{b_i, b_j\} = c_{ij} / \sqrt{c_{ii} c_{jj}} = 0$, для чего необходимо, чтобы $c_{ij} = 0$. Таким образом, ковариационная матрица размером $(k+1) \times (k+1)$ должна быть диагональной:

$$\mathbf{D} = \begin{pmatrix} c_{00} & 0 & \dots & 0 \\ 0 & c_{11} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & c_{kk} \end{pmatrix}. \quad (\text{III.162})$$

Условие (III.162) выполняется (см. прил. I.1) только для тех матриц \mathbf{x} размером $(k+1) \times N$, которые являются ортогональными; для них сумма почленных произведений двух векторов-столбцов матрицы равна нулю, т. е.

$$\sum_{u=1}^N x_{iu} x_{ju} = (ij) = 0 \quad (i < j, u = \overline{1, N}). \quad (\text{III.163})$$

Таким образом, в (III.162) — (III.163) записан критерий оптимальности, который называется ортогональностью плана и тождественен высказанному пожеланию экспериментатора.

Исследование матриц \mathbf{D} , мер точности d и других статистических характеристик с целью конструирования матриц плана \mathbf{x} и есть одно из основных направлений оптимального алгоритмизированного планирования эксперимента. Оптимальность оценивается по большому числу критериев, каждый из которых приобретает ту или иную ценность для экспериментатора в зависимости от целей его исследований. Эти критерии в абстрактной математической форме обобщают ту массу пожеланий, которые экспериментатор выдвигает к результатам своих исследований на профессионально-логическом уровне («Нужно как можно точнее оценить координаты максимальной прочности», «Влияние всех факторов должно оцениваться с одинаковой точностью» и т. п.). Возможно, что требования к плану эксперимента будут противоречивы, однако МТЭ [30, 40, 45, 46 и др) позволяет построить планы, достаточно хорошие с позиций разных критериев (планы, близкие к оптимальным). Ниже из большого числа критериев проанализированы те, которые имеют наибольшее распространение в прикладных областях, в частности в материаловедении и технологии.

Принятие принципа оптимальности плана эксперимента и введение четких (с математической точки зрения) критериев оптимальности, являющихся, как правило, некоторыми функционалами плана, требует уточнения ряда базовых определений МТЭ. Пла-

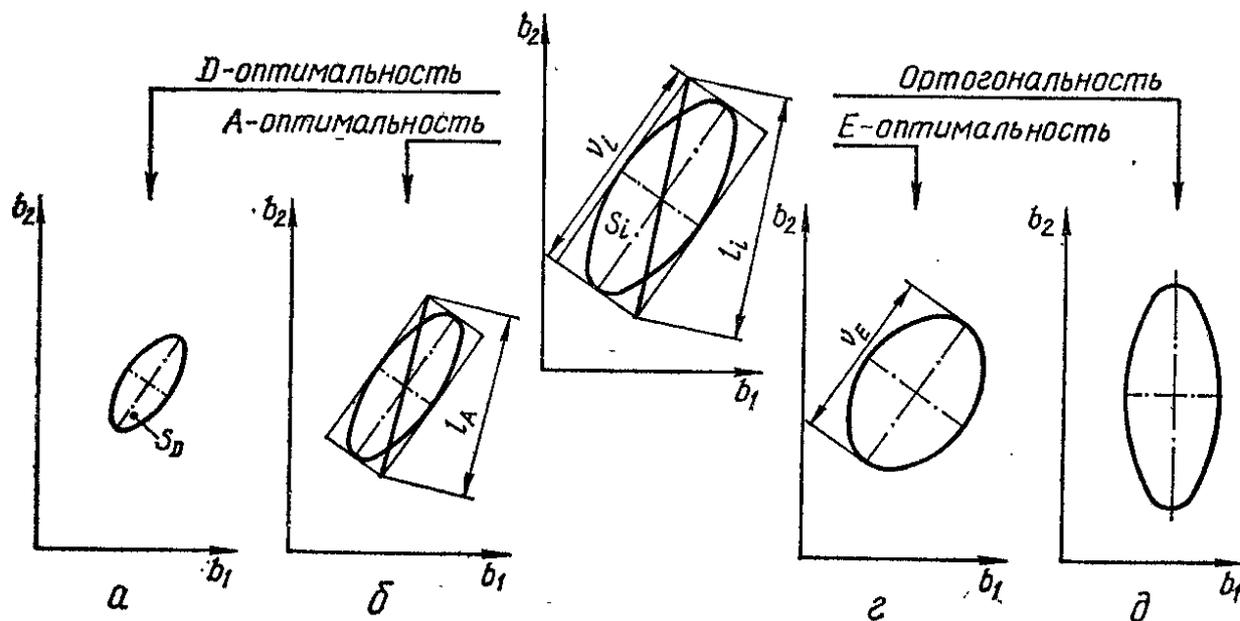


Рис. III.18. Взаимосвязи критериев оптимальности с эллипсом рассеяния:
 a — D -оптимальность; $б$ — A -оптимальность; $в$ — параметры эллипса; $г$ — E -оптимальность;
 $д$ — ортогональность

ном эксперимента E в общем случае является множество из N точек факторного пространства Ω_x с координатами x_u , в которых распределены N_m наблюдений за величиной Y :

$$E = \left\{ \begin{matrix} x_1, x_2, \dots, x_u, \dots, x_N \\ m_1, m_2, \dots, m_u, \dots, m_N \end{matrix} \right\}; \quad m_u \geq 1; \quad \sum_u m_u = N_m. \quad (\text{III.164})$$

Разные точки (верхняя строка в (III.164)) образуют *спектр плана* (если $m_u = \text{const}$, то понятие «план» на практике совпадает с понятием «спектр плана»; при $m_u = 1$ оба понятия тождественны — см. информационную таблицу в п. III.1). Для решения задач сравнительного анализа разных планов по критериям оптимальности целесообразно вместо формы (III.164) использовать *нормированные планы* ε , в которых абсолютное число дублирующих опытов (параллельных испытаний) m_u заменено на их относительное число $p_u = m_u/N_m$:

$$\varepsilon = \left\{ \begin{matrix} x_1, x_2, \dots, x_u, \dots, x_N \\ p_1, p_2, \dots, p_u, \dots, p_N \end{matrix} \right\} = \left\{ x_u \right\}_{u=1}^N; \quad \sum_u p_u = 1. \quad (\text{III.165})$$

Понятие «нормированный план» используется в тех случаях, когда поиск оптимального решения ведется при не фиксированном заранее общем объеме эксперимента N_m .

Первая группа критериев связана с *минимизацией дисперсий оценок коэффициентов* модели [4, 46, 58 и др.], которые описываются ковариационной матрицей D . Все эти критерии иллюстрируются на рис. III.18 эллипсами рассеяния при $\alpha = \text{const}$ для двух параметров — b_1 и b_2 (без потери общности).

Обобщенная дисперсия оценок коэффициентов (как случайных

величин) пропорциональна объему эллипсоида рассеяния (см. площадь S_t эллипса рассеяния двух случайных величин Y_1 и Y_2 на рис. III.8 и III.18, которая зависит от относительных ошибок s_1/\bar{Y}_1 и s_2/\bar{Y}_2).

План ε^* называется *D-оптимальным* (англ. *determinant* — определитель), если на множестве нормированных планов ε в факторной области Ω_x он минимизирует определитель $\det M_N^{-1}$ ковариационной матрицы (по ГОСТ 24026 — 80 критерии записываются как функционалы от матрицы, обратной информационной, $M_N^{-1} = D_N$). Это соответствует минимизации эллипсоида рассеяния оценок коэффициентов ЭС-модели (S_D на рис. III.18, а). Критерий *D-оптимальности* наиболее тесно связан с центральными идеями математической статистики (с теорией эффективных оценок, см. п. III.3) и является одним из наиболее распространенных в МТЭ.

Средняя дисперсия оценок коэффициентов пропорциональна длине диагонали l_A (рис. III.18, б) многомерного прямоугольника, описанного вокруг эллипса рассеяния. План ε^* называется *A-оптимальным* (англ. *average variance* — средняя дисперсия), если на множестве планов ε в Ω_x он минимизирует след $\text{tr} \{M_N^{-1}\}$ ковариационной матрицы $M_N^{-1} = D_N$ (см. прил. I.1), а следовательно, вышеуказанную диагональ.

План ε^* называется *E-оптимальным* (англ. *eigen value* — собственное значение), если он минимизирует максимальный главный диаметр ν_E (рис. III.18, в) эллипса рассеяния, что требует минимизации собственного значения (прил. I.1) ковариационной матрицы $\lambda(M^{-1})$.

План ε^* ортогонален, если $\rho\{b_i b_j\} = 0$. В данном случае: функционал от M^{-1} не минимизируется; D должна быть диагональной (III.162); степень приближения к ортогональному плану может быть оценена [46], в частности, по максимальному коэффициенту корреляции $\rho_{\max}(b_i b_j)$. Для ортогональных планов главные оси эллипса по направлению совпадают (рис. III.18, д) с координатными осями пространства параметров.

С оценками коэффициентов модели связаны и другие, реже употребляемые, критерии оптимальности планов: минимакс дисперсии оценок $\min \max(c_{ii})$, постоянство дисперсий оценок $c_{ii} = \text{const}$, минимум суммы относительных ошибок $\min \sum_i s\{b_i\}/b_i$, максимум суммы гауссовской точности оценок $\max \sum_i h\{b_i\}$ и др. [4, 9, 46].

Вторая группа критериев связана с *дисперсией предсказанных значений отклика* [4, 46, 58 и др.], т. е. с точностью модели в целом. Они базируются на мере точности d (III.146).

План *Q-оптимален*, если он минимизирует среднюю дисперсию оценки выхода $\bar{s}^2(Y)$; это обеспечивается при минимуме средней меры точности \bar{d} , определяемой при интегрировании $d = \psi(x)$, по всей области эксперимента Ω_x .

План *G-оптимален* (англ. *general variance* — общая дисперсия), если на множестве планов минимизируется максимальное в области Ω_x значение оценки дисперсии модели $(\max d)_{\min}$.

План *ротатабелен* (англ. *rotation* — вращение), если дисперсия постоянна на равных расстояниях $\rho = (\sum_{i=1}^k x_i^2)^{1/2}$ от центра, что обеспечивается при $d = \psi(\rho)$. Такие планы необходимы, когда у исследователя нет информации о предпочтительном направлении оптимизации от центра эксперимента, а все направления равновозможны.

План приближается к *униформному*, когда в пределах области эксперимента Ω_x характеристика d стремится к const . Такое свойство плана достигается повторением опытов в определенных точках, в частности в центре плана.

Третья группа критериев оптимальности (называемых иногда [46] критериями желательности) направлена на *облегчение* процедур *эксперимента* и *построения модели*.

Минимизация числа опытов N — приближение его к числу оцениваемых параметров L . В пределе $N = L$ и план становится *насыщенным* (для проверки адекватности начальной модели числа степеней свободы не остается). Такие планы полезны на поисковых этапах исследования, когда можно получить хотя и не очень точное и полное представление об объекте, но зато при минимуме расхода ресурсов. Следует отметить, что МТЭ предусматривает и такие варианты исследований, когда $N < L$ (сверхнасыщенные планы [46, 48 и др.]). Они значительно менее информативны, однако полезны в условиях дефицита ресурсов.

Простота вычислений оценок коэффициентов оказывается важным достоинством плана при отсутствии ЭВМ.

Композиционность плана — возможность использовать точки предыдущего этапа эксперимента на последующем этапе, если модель первого этапа оказалась неудовлетворительной для описания системы (в частности, неадекватной) — см. п. III.12.

При реализации опытов по тому или иному оптимальному плану *порядок* их реализации должен быть *рандомизирован* (англ. *random* — наугад, случайный, беспорядочный). Использование принципа случайного выбора строки матрицы плана (например, по таблице равномерно распределенных случайных чисел — см. прил. II.16) для проведения очередного опыта позволяет снять систематическое влияние неучтенных факторов; это является одним из принципов оптимизации эксперимента.

Критерии оптимальности позволяют, во-первых, выбрать из каталогов типовых планов эксперимента именно тот план, который при заданных ресурсах даст наилучшую модель с точки зрения конечных целей исследования, во-вторых, при отсутствии типовых планов синтезировать на ЭВМ (см. п. III.14) специальные для данной постановки задачи оптимальные планы E^* .

III.11. Планы полного и дробного факторного эксперимента для построения линейных и неполных квадратичных моделей

Линейные модели (III.87) применимы в двух случаях: 1) когда из результатов профессионально-логического анализа известно априори, что факторы во всей исследуемой области факторного пространства Ω_x влияют на отклик линейно (весьма редкий случай в материаловедении и технологии); 2) когда в отношении малой части Ω_x можно допустить (в частности, на поисковом этапе исследования), что участок криволинейной поверхности аппроксимируется плоскостью (отрезок кривой аппроксимируется прямой).

Неполные квадратичные модели, содержащие эффект $b_{ij}x_i x_j$, также применимы только в том случае, когда выдвигается гипотеза *линейности* влияния *каждого* отдельного фактора (см. рис. III.4), поэтому их практическое значение не намного больше линейных, хотя выявление значимых эффектов взаимодействия факторов становится, как правило, стимулятором углубленного анализа механизма явления.

Для построения k -факторных линейных и неполных квадратичных моделей лучшими (оптимальными по многим критериям) являются двухуровневые факторные планы эксперимента; в них отсутствуют точки, кроме тех, которые соответствуют -1 (нижний уровень фактора $x_{i,\min}$) и $+1$ (верхний уровень фактора $x_{i,\max}$). С геометрической точки зрения это означает, что опытные точки расположены только в вершинах гиперкуба. Число вершин $N_k = 2^k$: при $k = 2$ — квадрат с $N_k = 4$; при $k = 3$ — куб с $N_k = 8$; при $k = 4$ — гиперкуб с $N_k = 16$ и т. п.

Эксперимент по плану, образуемому полным перебором возможных сочетаний для k факторов на Q уровнях и содержащий $N = Q^k$ опытов, называется *полным факторным экспериментом* (ПФЭ). Для построения рассматриваемых в данном разделе моделей достаточен ПФЭ 2^k , который не только *симметричен* ($\sum_u x_i = 0$) и *ортogonalен* (III.163), но и в силу отсутствия точек с координатами, отличными от $|x_i| = 1$, обладает свойством

$$\sum_{u=1}^N x_{iu}^2 = N, \quad (\text{III.166})$$

вследствие которого все *диагональные* элементы ковариационной матрицы c_{ii} равны между собой и

$$c_{ii} = N^{-1}, \quad (\text{III.167})$$

что приводит формулы (II.123) и (III.132) к простейшим соотношениям:

$$b_i = (iY)/N = \left(\sum_{u=1}^N x_{iu} y_u \right) / N; \quad (\text{III.168})$$

$$s^2 \{b_i\} = s_y^2 / N.$$

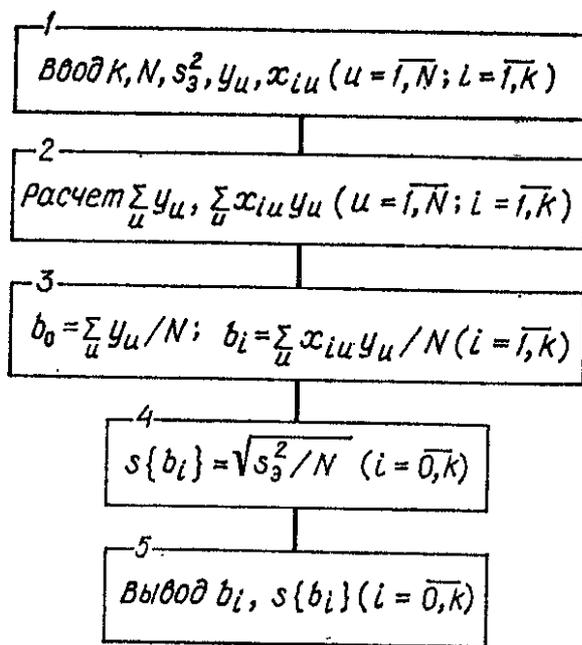
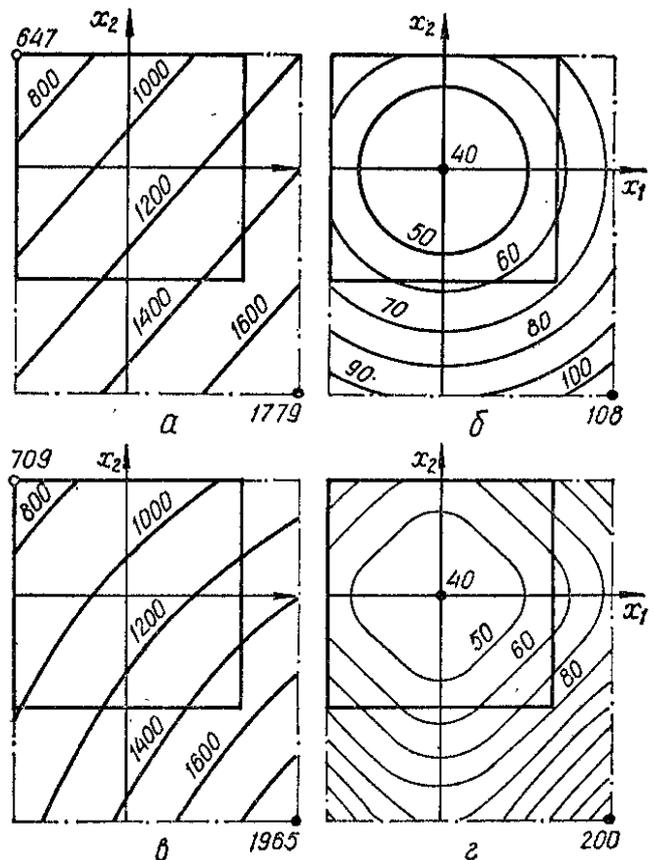


Рис. III.19. Блок-схема алгоритма расчета оценок модели по результатам ПФЭ

Рис. III.20. Изолинии (в области $-1 \leq x_1 \leq +1,5, -2 \leq x_2 \leq +1$):

a — двухфакторной линейной модели; *b* — ее ошибок предсказания; *в* — неполной квадратичной модели, *г* — ее ошибок предсказания



План полного факторного эксперимента 2^k не только ортогонален — он оптимален в широком смысле, т. е. соответствует *D*-, *A*-, *E*-, *G*-оптимальности, а для линейных моделей еще и ротатабелен.

Блок-схема алгоритма расчета оценок коэффициентов ЭС-моделей при использовании результатов ПФЭ показана на рис. III.19.

Пример III.34. При исследовании ячеистого бетона в условиях пр. III.4 был определен модуль упругости E . План и результаты эксперимента ПФЭ 2^2 представлены в табл. III.26 (средняя ошибка эксперимента принята $s_y = 80$ МПа при $f_y = \infty$). Рассмотрена гипотеза о начальном описании объекта неполной квадратичной моделью.

Число опытов $N = 2^2 = 4$ (столбец 1); оно совпадает с суммой по строкам фиктивной переменной $(00) = 4$. План (столбцы 3 и 4) — полный перебор из $Q = 2$ по $k = 2$ — симметричен и ортогонален (столбец 5). Требование (III.166) выполняется (столбцы 6 и 7). Сумма $(0Y)$ результатов опытов (столбец 8) при умножении на 0,25 в соответствии с (III.168) дает значение оценки свободного члена $b_0 = 4244 \cdot 0,25 = 1061$. В столбцах 9 — 11 экспериментальные данные умножены на значения x_1, x_2 и $x_1 x_2$ (соответственно), а произведения суммированы. После расчетов по (III.168) получена начальная модель

$$\hat{E}_H = 1061 + 220x_1 - 194x_2 - 62x_1 x_2. \quad (\text{III.169})$$

Ошибка оценок коэффициентов по (III.168) равна $s\{b_l\} = \sqrt{80^2/4} = 40$ МПа; критический уровень по (III.135) $b_{кр} = t_{0,1}s\{b_l\} = 65,8$ ($t_{0,1} = 1,645$ — см. прил. II.5), следовательно, истинный коэффициент взаимодействия факторов $\beta_{12} = 0$ и эффект $b_{12} = -62$ следует удалить как незначимый. При этом остающиеся оценки конечной модели

$$\hat{E} = 1061 + 220x_1 - 194x_2 \quad (\text{III.170})$$

не пересчитываются.

Таблица III.26. Полный двухфакторный двухуровневый эксперимент (ПФЭ 2²) и его результаты

Номер опыта	Фиктивная переменная $x_0 \equiv +1$	Планирование		$x_1 x_2$	x_1^2	x_2^2	Результат $Y = E$, МПа	$x_1 Y$	$x_2 Y$	$x_1 x_2 Y$	\hat{E} , МПа	Кодовое обозначение строки
		x_1	x_2									
1	+1	-1	-1	+1	+1	+1	973	-973	-973	973	1035	(1)
2	+1	+1	-1	-1	+1	+1	1537	+1537	-1537	-1537	1475	a
3	+1	-1	+1	-1	+1	+1	709	-709	+709	-709	647	b
4	+1	+1	+1	+1	+1	+1	1025	+1025	+1025	+1025	1087	ab
N	$\sum_u x_{0u}$	$\sum_u x_{1u}$	$\sum_u x_{2u}$	$\sum_u x_{1u} x_{2u}$	$\sum_u x_{1u}^2$	$\sum_u x_{2u}^2$	$\sum_u y_u$	$\sum_u x_{1u} y_u$	$\sum_u x_{2u} y_u$	$\sum_u x_{1u} x_{2u} y_u$	—	—
(00)	(00)	(01)	(02)	(12)	(11)	(22)	(0Y)	(1Y)	(2Y)	(12Y)	—	—
4	4	0	0	0	4	4	4244	880	-776	-248	—	—

Изолинии модели (III.170) показаны на рис. III.20, а (общий вид см. рис. III.3). При этом для упрощения интерпретации результатов и сравнительного анализа границы анализируемого факторного пространства расширены вправо до $x_1 = +1,5$ и вниз до $x_2 = -2$. Расчетные значения для опытных точек приведены в столбце 12 табл. III.26; дисперсия неадекватности, возникающая в связи с удалением незначимого эффекта, $s_{\text{на}}^2 = \sum_{u=1} (E_u - \hat{E}_u)^2 / f_{\text{на}} = 15\,376(4-3) = 15\,376$. Критерий (III.141) $F_a = s_{\text{на}}^2 / s_{\text{э}}^2 = 15\,376 / 80^2 = 2,4$, что меньше критического $F_{\alpha} \{ \alpha = 0,05, f_1 = 1, f_2 = \infty \} = 3,84$ (см. прил. II.6), поэтому модель (III.170) адекватно описывает результаты эксперимента.

Дисперсия предсказания, согласно (III.144), описывается моделью

$$\begin{aligned} s^2 \{ \hat{E} \} &= s_{\text{э}}^2 (c_{00} + c_{11}x_1^2 + c_{22}x_2^2) = \frac{s_{\text{э}}^2}{N} (1 + x_1^2 + x_2^2) = \\ &= \frac{80^2}{4} (1 + x_1^2 + x_2^2) = 1600 (1 + \rho^2). \end{aligned} \quad (\text{III.171})$$

Изолинии ошибки $s \{ \hat{E} \}$ показаны на рис. III.20, б. Это концентрические окружности с радиусом $\rho = \sqrt{x_1^2 + x_2^2}$, поскольку план ПФЭ 2^2 для линейной модели ротатабелен.

Следует отметить, что если уровень риска повысить до $\alpha = 0,2$, то критерий $t_{0,2}$ снизится до 1,282, и эффект взаимодействия будет оставлен. Тогда конечная модель совпадает с начальной (III.169); для проверки ее адекватности степени свободы отсутствуют. Модель (III.169) отображает гиперболоид, поэтому изолинии искривлены (рис. III.20, в), а экстремальные значения $\hat{E}_{\text{min}} = 709$ МПа и $\hat{E}_{\text{max}} = 1965$ МПа отличаются от полученных для плоскости (рис. 20, а). Значительные изменения претерпевает и дисперсия предсказанного значения, которая теперь определяется как

$$s^2 \{ \hat{E} \} = \frac{s_{\text{э}}^2}{N} (1 + x_1^2 + x_2^2 + x_1^2 x_2^2) = 1600 [(1 + x_1^2) (1 + x_2^2)]. \quad (\text{III.172})$$

План не ротатабелен; изолинии (рис. III.20, в) сплющены вдоль диагоналей квадрата к центру. Ошибка предсказания возрастает за границей области эксперимента значительно быстрее, чем по (III.171).

Допустима запись матриц в табл. III.26: «+» вместо «+1» и «-» вместо «-1». Кроме того, существует сокращенная запись строк в кодовых обозначениях (см. столбец 13 в табл. III.26): например, буква *a* обозначает, что x_1 находится на верхнем уровне; запись *ab* показывает, что на верхних уровнях находятся обе переменные; символ (1) обозначает, что все x_i находятся на нижних уровнях.

Для трех факторов матрица планирования 2^3 получается из матрицы 2^2 при повторении ее дважды: один раз при значении x_3 на нижнем уровне, второй раз — на верхнем. Это формально равносильно умножению кодовой записи матрицы один раз на (1), другой — на *c*: (1), *a*, *b*, *ab*, *c*, *ac*, *bc*, *abc*.

Схема построения полных факторных экспериментов из матрицы 2^3 показана на рис. III.21. Продолжая схему далее, легко образовать ПФЭ для любого числа факторов.

С ростом числа факторов число опытов растет по показательной функции $N = 2^k$. В то же время в (см. табл. III.25) число ли-

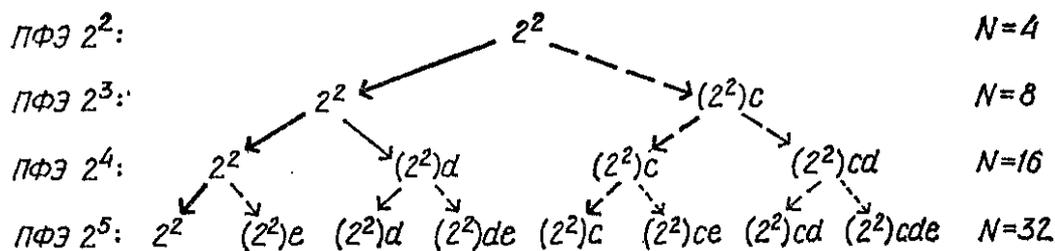


Рис. III. 21. Схема построения ПФЭ 2^5

нейных эффектов растет по зависимости $L_{л.м} = 1 + k$, а число эффектов в неполной квадратичной модели — по параболической зависимости $L_{н.к.м} = 0,5 (2 + k + k^2)$. Таким образом, начиная с $k = 3$ быстро растет количество лишних экспериментов. Рост оценивается коэффициентом избыточности $e_N = N/L$. Так, для $k = 5$ при линейной модели $e_N = 32/6 = 5,33$, а при неполной квадратичной — $e_N = 32/16 = 2$.

Алгоритмизированное планирование эксперимента получило широкое распространение после разработки метода дробных реплик, позволяющего ограничиться некоторой частью ПФЭ. В случае дробного факторного эксперимента (ДФЭ) нужно так разделить матрицу ПФЭ, чтобы e_N стремилось к 1, что минимизирует расход ресурсов.

Основной идеей метода дробных реплик является построение ортогональных планов, в которых эффекты высших порядков (появление которых маловероятно в данной задаче) смешиваются с какими-то новыми эффектами. Так, если в трехфакторном эксперименте есть основания априори полагать, что эффектов взаимодействия нет ($\beta_{12} = \beta_{13} = \beta_{23} = 0$), достаточно поставить 4 опыта (см. табл. III.25). Поскольку $\beta_{12} = 0$, то в матрице независимых переменных ПФЭ 2^3 (табл. III.26) можно принять $x_1 x_2 = x_3$ и получить план для трех факторов x_1, x_2 и x_3 , задаваемый левой половиной табл. III.27.

Таблица III.27. Полу реплики ПФЭ 2^3 (планирование типа 2^{3-1})

Первая полу реплика $2^{3-1} (x_1 x_2 = x_3)$					Вторая полу реплика $2^{3-1} (x_1 x_2 = x_3)$						
Базисные функции				Код строк	y_i	Базисные функции				Код строк	y_i
$f_0 = x_0 = +$	$f_1 = x_1$	$f_2 = x_2$	$f_3 = x_1 x_2 = x_3$			$f_4 = x_0 = +$	$f_5 = x_1 = x_2 x_3$	$f_6 = x_2 = x_1 x_3$	$f_7 = x_3 = x_1 x_2$		
+	+	+	+	c	y_1	+	+	+	+	(1)	y_5
+	+	-	-	a	y_2	+	+	-	-	ac	y_6
+	-	+	-	b	y_3	+	-	+	+	bc	y_7
+	-	-	+	abc	y_4	+	-	-	-	ab	y_8

Если же в действительности окажется, что в данной задаче $\beta_3 = 0$, то коэффициент b_3 , оцененный по такому плану, будет оценкой

одновременно и параметра β_3 и параметра β_{12} , что символически записывается как $b_3 \rightarrow \beta_3 + \beta_{12}$. В этой ситуации b_3 является *смешанной оценкой*, относящейся к нескольким отличным от нуля эффектам. В плане столбцы смешанных эффектов одинаковы, поэтому без дополнительных гипотез и (или) экспериментов делать выводы о влиянии на Y факторов (в частности, x_3 или взаимодействия x_1x_2) *нельзя!*

В первой полуреплике 2^{3-1} (табл. III.27) смешанной оценкой является не только коэффициент b_3 , но и коэффициенты $b_1 \rightarrow \beta_1 + \beta_{23}$ и $b_2 \rightarrow \beta_2 + \beta_{13}$. Если возникают сомнения в том, что $\beta_{ij} = 0$, то нужно поставить еще четыре опыта при $x_3 = -x_1x_2$. В результате получается вторая полуреплика 2^{3-1} , которая вместе с первой дает ПФЭ 2^3 .

Расщепление ПФЭ на полуреплики при сохранении их ортогональности нельзя выполнять произвольно. В первую реплику отбираются строки с нечетным числом букв (выполняются требования $x_3 = x_1x_2$), во вторую — с четным, причем код (1) всегда считается четным. Реплики, содержащие только четные или нечетные комбинации букв, называются *главными*.

Соотношения $x_3 = x_1x_2$ или $x_3 = -x_1x_2$, задающие дробную реплику в плане 2^{3-1} (табл. III.27), называются *генерирующим соотношением*. Оно показывает, какие взаимодействия заменены новыми факторами при построении дробной реплики. Вторым необходимым понятием является *определяющий контраст* — соотношение, задающее элементы столбца матрицы планирования, соответствующего фиктивной переменной $x_0 = +1$. В рассматриваемом трехфакторном плане определяющий контраст будет получен, если обе части генерирующего соотношения $x_3 = x_1x_2$ умножить на x_3 (величина $x_3^2 = +1$):

$$x_3^2 = x_1x_2x_3 = 1. \quad (\text{III.177})$$

Определяющий контраст дает возможность установить, какие оценки являются смешанными. Для этого обе его части последовательно умножаются на каждый из факторов x_i . Например, для фактора x_1 (учитывая, что $x_i^2 = 1$) будет получено:

$$\left. \begin{aligned} x_1x_2x_3 &= 1; \\ x_1(x_1x_2x_3) &= x_1 \cdot 1; \\ x_1^2x_2x_3 &= x_1; \\ x_2x_3 &= x_1, \end{aligned} \right\} \quad (\text{III.178})$$

следовательно, оценка $b_1 \rightarrow \beta_1 + \beta_{23}$. Аналогично определяются $b_2 \rightarrow \beta_2 + \beta_{13}$ и $b_3 \rightarrow \beta_3 + \beta_{12}$; это соответствует ранее сказанному о свойствах первой полуреплики 2^{3-1} в табл. III.27. В прил. II.11 приведены наиболее часто встречающиеся планы ДФЭ и указаны смешанные оценки. Следует отметить, что кроме реплик типа 2^{k-l} , называемых регулярными репликами, могут быть взяты от ПФЭ и нерегулярные реплики с числом опытов $N = \sum 2^{k-m_i}$, например $3/4$ от 2^k и т. п. [46, 48 и др.].

Для построения *линейных* моделей большой размерности ($k \geq 8$) применение регулярных планов ДФЭ может оказаться недостаточно эффективным по избыточности e_N . Так, для $k = 18$ ближайшая дробная реплика $2^{18-18} = 32$ имеет $e_N = 1,68$. Однако класс насыщенных ортогональных планов расширен за счет конструирования специальных матриц. При $(k + 1)$ кратном 4, т. е. для $k = 7, 11, 15, \dots, 199$ (кроме $k = 91$) при построении линейных моделей используются планы Плакетта — Бермана PB_i [4] и др., для которых в строках табл. III.28 приведены первые столбцы матриц x .

Таблица III. 28. Первые столбцы ортогональных планов PB_i при различных k

$k = 7$	+ + + -	+ - -			
$k = 11$	+ + - +	+ + - -	- + -		
$k = 15$	+ + + +	- + - +	+ - - +	- - - -	
$k = 19$	+ + - -	+ + + +	- + - +	- - - -	+ + -
$k = 23$	+ + + +	+ - + -	+ + - -	+ + - -	+ - + -

Порядок построения x : а) выписать в столбец для фактора x_1 строку из табл. III.28 размерности k ; б) сдвинув на один элемент вверх (верхний элемент первого столбца теперь станет нижним), записать уровни варьирования фактора x_2 ; в) продолжить процедуру до получения квадратной матрицы $k \times k$; г) дописать внизу $(k + 1)$ -ю строку. Если размерность задачи $(k + 1)$ не кратна 4, то добавляются «фиктивные» переменные в количестве, необходимом для применения ближайшего плана PB_i [4].

Пример III. 35. План PB_i использован [4] для оценки влияния на конечный результат контроля качества гипса нарушений семи стандартизованных условий:

- x_1 — неверное определение нормальной плотности гипсового теста ((НГ $\pm \pm 2$) %, верхняя граница совмещается с «+1», а нижняя — с «-1»);
- x_2 — изменение в порядке загрузки компонентов («+1» — гипс в воду; «-1» — наоборот);
- x_3 — продолжительность загрузки $\tau_3 = 30 \pm 15$ с;
- x_4 — продолжительность смешивания $\tau_{см} = 30 \pm 10$ с;
- x_5 — вид смешивания («+1» — механическое, «-1» — ручное);
- x_6 — вид заливки в формы («-1» — параллельно, «+1» — последовательно);
- x_7 — время твердения до испытания $\tau_{ТВ} = (60 \pm 10)$ мин.

Факторы $x_2, x_5,$ и x_6 варьируют на двух дискретных уровнях, поэтому не принципиально, какой из них совмещается с $x_i = +1$, а какой — с $x_i = -1$. Для этих факторов понятие «линейная функция» не имеет смысла, а оценки коэффициентов соответствуют «полускачку» выхода системы за счет перехода от «-1» к «+1». Результаты эксперимента по плану PB_7 показаны в табл. III.29; строка плана из табл. III.28 обведена.

Элементы матрицы моментов $(0Y), \dots, (7Y)$, полученные (см. табл. III.28) суммированием произведений $x_{iu}R_{iu}$, позволяют найти все оценки $b_{i_s}^* = (iY)/8$ и построить начальную модель

$$\hat{R} = 6,01 - 0,54x_1 + 0,06x_2 + 0,21x_3 + 0,24x_4 + 0,36x_5 - 0,06x_6 + 0,14x_7.$$

Ошибка эксперимента $s_3 = 0,1$ МПа при $f_3 = 14$; для риска $\alpha = 0,05$ $t_{\alpha} = 1,767$ (см. прил. II.5), критический уровень коэффициентов определяется

Таблица III.29. Результаты эксперимента по определению прочности гипса по плану РВ₇

Номер опыта	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	R_u , МПа
1	+	+	+	+	-	+	-	-	5,8
2	+	+	+	-	+	-	-	+	5,4
3	+	+	-	+	-	-	+	+	5,1
4	+	-	+	-	-	+	+	+	6,6
5	+	+	-	-	+	+	+	-	5,6
6	+	-	-	+	+	+	-	+	7,5
7	+	-	+	+	+	-	+	-	6,5
8	+	-	-	-	-	-	-	-	5,6
(00)	(0Y)	(1Y)	(2Y)	(3Y)	(4Y)	(5Y)	(6Y)	(7Y)	
8	48,1	-4,3	0,5	+1,7	+1,9	+2,9	-0,5	+1,1	

по (III.135) как $b_{кр} = 1,761 \cdot 0,1 / \sqrt{8} = 0,062$. Оценки b_2 и b_8 принимаются равными нулю, и конечная модель имеет вид

$$\hat{R} = 6,01 - 0,54x_1 + 0,21x_3 + 0,24x_4 + 0,36x_5 + 0,14x_7.$$

Для оценки степени влияния нарушений стандартных условий ранжируются удвоенные абсолютные значения оценок $2|b_i|$, интерпретируемые, например для b_5 , как «переход от ручного смешивания компонентов к механическому», а для b_7 — как «флуктуация времени твердения от 50 до 70 мин» и т. п.:

$$|2b_1| > |2b_5| > |2b_4| > |2b_3| > |2b_7|.$$

Из ряда следует, в частности, что кроме точной оценки нормальной густоты гипсового теста необходимо особенно тщательно выполнять операции по его приготовлению.

В целом проблему получения оптимальных планов для построения линейных моделей можно считать решенной.

III.12. Планы для построения квадратичных моделей

Многофакторные полиномиальные модели (III.11) второго порядка позволяют решить абсолютное большинство инженерных задач (см. п. III.15, III.16) в материаловедении и технологии, поэтому эти ЭС-модели можно считать базовыми для инженерной практики в данной отрасли.

Полиномиальные k -факторные ЭС-модели второго порядка описывают поверхности, образованные *квадратичными параболоми*, которые в частных случаях могут вырождаться в прямые. Известно (см. табл. III.25), что для построения каждой однофакторной квадратичной параболы (III.12) необходима информация об Y как минимум при трех разных уровнях этого фактора. Если же последовательно ввести в исследование второй фактор, также действующий на Y нелинейно, то следует построить не одну параболу, а семейство как минимум из трех парабол, на что будет $3 \times 3 = 9$ опытов, соответствующих ПФЭ 3^2 .

Пример III.36. Изменение предела прочности мелкозернистого бетона R , МПа, под влиянием изменения концентраций добавок суперпластификатора $D_{\text{сп}} = X_1 = 0,4 \pm 0,4 \%$ и электролита $D_{\text{э}} = X_2 = 0,6 \pm 0,6 \%$ (от массы цемента) описывается [22] как поверхность эллиптического параболоида ($|x_i| \leq 1$)

$$\hat{R} = 18,1 - 3,8x_1^2 + 0,4x_2 - 3,6x_2^2 + 2x_1x_2, \quad (\text{III.179})$$

показанного на рис. III.22

Если изменять $D_{\text{сп}}$ при трех фиксированных уровнях $D_{\text{э}}$, то будут получены три параболы:

а) при $D_{\text{э}} = 0$ (нет добавки электролита; $x_2 = -1$) через точки 1—2—3 проходит парабол

$$\hat{R}\{x_2 = -1\} = 14,1 - 2x_1 - 3,8x_1^2$$

с максимумом прочности $R_{\text{max}} = 14,36$ МПа в точке $x_1 = -0,263$ или $D_{\text{сп}} = 0,3$;

б) при $D_{\text{э}} = 0,6 \%$ ($x_2 = 0$) через точки 4—5—6 проходит парабол

$$\hat{R}\{x_2 = 0\} = 18,1 - 3,8x_1^2$$

с максимумом $R_{\text{max}} = 18,1$ МПа при $x_1 = 0$ или $D_{\text{сп}} = 0,4 \%$;

в) при $D_{\text{э}} = 1,2 \%$ ($x_2 = +1$) через точки 7—8—9 проходит парабол

$$\hat{R}\{x_2 = +1\} = 14,9 + 2x_1 - 3,8x_1^2$$

с максимумом $R_{\text{max}} = 15,16$ МПа при $x_2 = +0,263$ или $D_{\text{сп}} = 0,5$.

Как видно, оптимальная концентрация добавки $D_{\text{сп}}$ существенно изменяется в зависимости от принятой концентрации добавки $D_{\text{э}}$, что обусловлено *неаддитивностью* (лат. *additio* — прибавление) действия на систему этих факторов; это отражено в модели (III.179) эффектом взаимодействия $b_{12} = 2$.

По тем же 9 точкам при $x_1 = \text{const}$ можно построить другое семейство парабол:

$$\begin{cases} \hat{R}\{x_1 = -1\} = 14,3 - 1,6x_2 - 3,6x_2^2; & x_{2, \text{opt}} = -0,222; \\ \hat{R}\{x_1 = 0\} = 18,1 + 0,2x_2 - 3,6x_2^2; & x_{2, \text{opt}} = +0,056; \\ \hat{R}\{x_1 = +1\} = 14,3 + 2,4x_2 - 3,6x_2^2; & x_{2, \text{opt}} = +0,333. \end{cases}$$

Следует отметить, что квадратичные параболы будут получены в любом сечении поверхности (III.179). Так, через три точки 1—5—9, лежащие на диагонали факторного квадрата, проходит полученная из (III.179) при $x_2 = x_1$ парабол

$$\hat{R}\{x_2 = x_1\} = 18,1 + 0,4x_1 - 5,4x_1^2$$

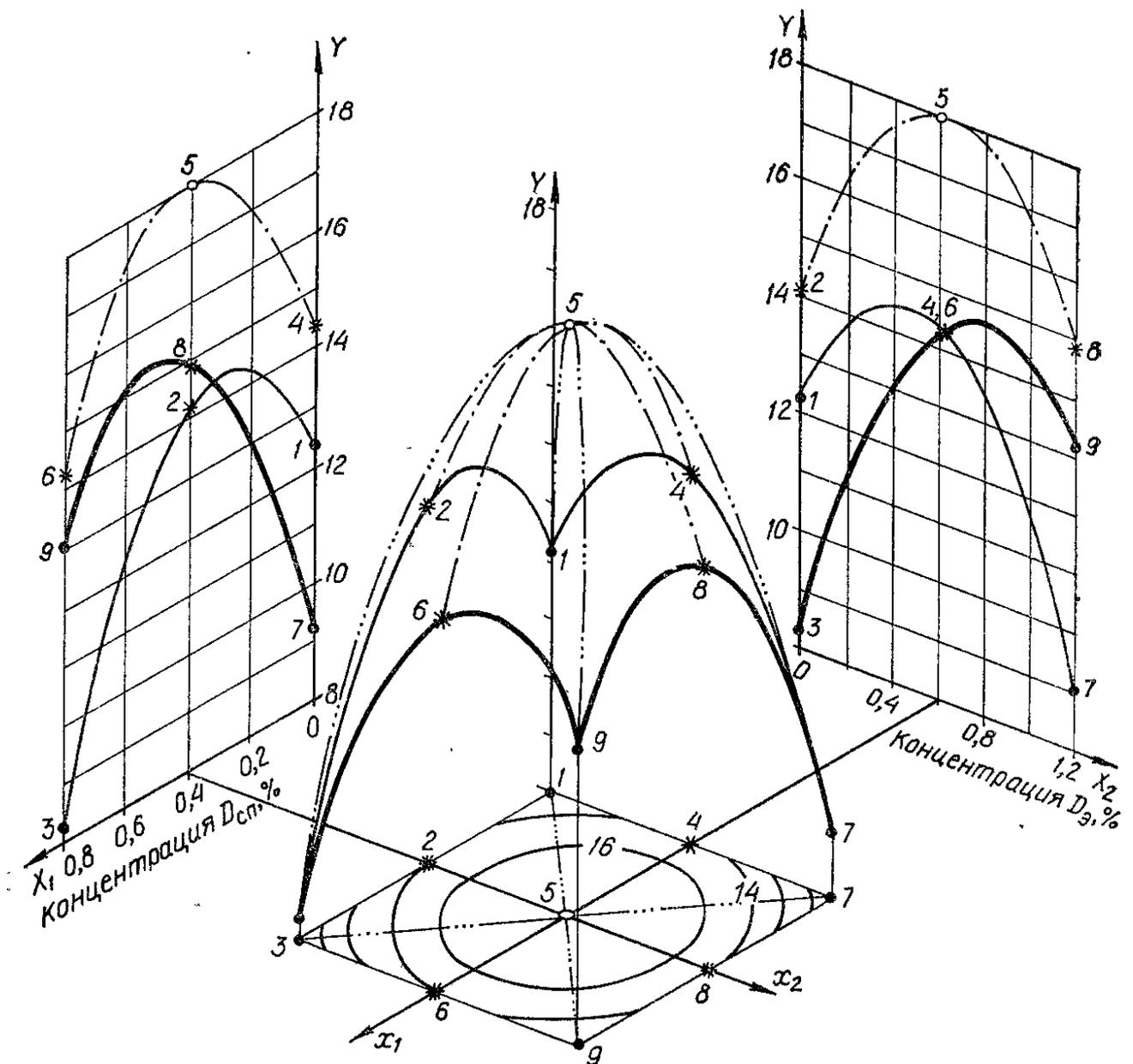


Рис. III. 22. Геометрический образ поверхности отклика в виде эллиптического параболоида и проекции его основных криволинейных образующих на три координатные плоскости

Для оценки 6 коэффициентов число опытов $N = 9$ — избыточно. Три опыта можно было бы сократить.

Точки плана ПФЭ 3^2 совпадают (рис. II.22) с 4 вершинами квадрата и 5 точками-центроидами (4 середины ребер квадрата и центр тяжести на пересечении диагоналей). План, приведенный в столбцах 2 и 3 табл. III.30, симметричен (суммы в столбцах 2 и 3 равны нулю), но не ортогонален для квадратичной модели, поскольку, в частности, сумма произведений двух базисных функций в столбцах 5 и 6 $\sum_u x_{1u}^2 x_{2u}^2 = 4$, т. е. условие (III.163) не соблюдается.

Не выполняется и условие (III.165), характерное для ПФЭ 2^k , так как $\sum_u x_{1u}^2 = 6$ (столбцы 5 и 6), в то время как $N = \sum_u x_{0u} = 9$ (столбец 4).

Таблица 111.30. План ПФЭ 3^2 , результаты эксперимента и расчет вспомогательных величин для оценки коэффициентов b_0 , b_i , b_{ii} и b_{12}

Номер опыта	Матрица базисных функций						Рецептура, %		Результаты эксперимента $Y = R$, МПа	Расчет сумм (ijY) , $j \geq i > 0$				
	План		x_0	x_1^2	x_2^2	x_1x_2	$D_{\text{сп}}$	D_9		x_1Y	x_2Y	x_1^2Y	x_2^2Y	x_1x_2Y
	x_1	x_2												
1	-1	-1	+1	1	1	+1	0	0	12,3	-12,3	-12,3	12,3	12,3	+12,3
2	0	-1	+1	0	1	0	0,4	0	14,1	0	-14,1	0	14,1	0
3	+1	-1	+1	1	1	-1	0,8	0	8,3	+8,3	-8,3	8,3	8,3	-8,3
4	-1	0	+1	1	0	0	0	0,6	14,3	-14,3	0	14,3	0	0
5	0	0	+1	0	0	0	0,4	0,6	18,1	0	0	0	0	0
6	+1	0	+1	1	0	0	0,8	0,6	14,3	+14,3	0	14,3	0	0
7	-1	+1	+1	1	1	-1	0	1,2	9,1	-9,1	+9,1	9,1	9,1	-9,1
8	0	+1	+1	0	1	0	0,4	1,2	14,9	0	+14,9	0	14,9	0
9	+1	-1	+1	1	1	+1	0,8	1,2	13,1	+13,1	+13,1	13,1	13,1	+13,1
	$(01) = 0$	$(02) = 0$	$(00) = 9$	$(11) = 6$	$(12) = 6$	$(22) = 0$	—		$(0Y) = 108,5$	$(1Y) = 0$	$(2Y) = 2,4$	$(11Y) = 71,4$	$(22Y) = 71,8$	$(12Y) = 8$

Матрица моментов плана M , образованная, согласно (III.107) из матрицы базисных функций в табл. III.30, имеет блочно-диагональную структуру:

$$M = \begin{pmatrix} (00) & (011) & (022) & (01) & (02) & (012) \\ (110) & (1111) & (1122) & (111) & (112) & (1112) \\ (220) & (2211) & (2222) & (221) & (222) & (2212) \\ (10) & (111) & (122) & (11) & (12) & (112) \\ (20) & (211) & (222) & (21) & (22) & (212) \\ (120) & (1211) & (1222) & (121) & (122) & (1212) \end{pmatrix} = \begin{pmatrix} 9 & 6 & 6 & | & 0 & 0 & 0 \\ 6 & 6 & 4 & \cdot & 0 & 0 & 0 \\ 6 & 4 & 6 & | & 0 & 0 & 0 \\ \dots & \dots & \dots & \cdot & \dots & \dots & \dots \\ 0 & 0 & 0 & | & 6 & 0 & 0 \\ 0 & 0 & 0 & \cdot & 0 & 0 & 0 \\ 0 & 0 & 0 & | & 0 & 0 & 6 \end{pmatrix}. \quad (\text{III.180})$$

Информационная матрица M_N , полученная из (III.180) согласно (III.109), и ковариационная матрица D , полученная из (III.180) согласно (III.115), также блочно-диагональны:

$$M_N = \frac{1}{N} M = \begin{pmatrix} 1 & 2/3 & 2/3 & | & & & \\ 2/3 & 2/3 & 4/9 & \cdot & & & 0 \\ 2/3 & 4/9 & 2/3 & | & & & \\ \dots & \dots & \dots & \cdot & \dots & \dots & \dots \\ & & & & 2/3 & 0 & 0 \\ 0 & & & \cdot & 0 & 2/3 & 0 \\ & & & & 0 & 0 & 2/3 \end{pmatrix}; \quad (\text{III.181})$$

$$D = \begin{pmatrix} 5/9 & -1/3 & -1/3 & | & & & \\ -1/3 & 1/2 & 0 & \cdot & & & 0 \\ -1/3 & 0 & 1/2 & | & & & \\ \dots & \dots & \dots & \cdot & \dots & \dots & \dots \\ & & & & 1/6 & 0 & 0 \\ 0 & & & \cdot & 0 & 1/6 & 0 \\ & & & & 0 & 0 & 1/4 \end{pmatrix} = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}. \quad (\text{III.182})$$

Блочно-диагональная структура ковариационных матриц симметричных планов позволяет, используя элементы матриц

$$D_1 = \left. \begin{pmatrix} T_1 & T_2 & \dots & T_2 \\ T_2 & T_{11} & \dots & T_5 \\ \cdot & \cdot & \cdot & \cdot \\ T_2 & T_5 & \dots & T_{11} \end{pmatrix} \right\} \begin{array}{l} \text{— для оценки } b_0, \\ \text{— для оценки } b_{ii}; \end{array} \quad (\text{III.183})$$

$$D_2 = \left(\begin{array}{ccccc} T_3 & 0 & \dots & 0 & 0 \\ 0 & T_3 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & T_6 & 0 \\ 0 & 0 & \dots & 0 & T_6 \end{array} \right) \left. \begin{array}{l} \text{— для оценки } b_i, \\ \text{— для оценки } b_{ij}; \end{array} \right\} \quad (\text{III.184})$$

построить достаточно простые формулы для расчета на микрокалькуляторах оценок коэффициентов регрессии:

$$\left. \begin{array}{l} b_0 = T_1(0 Y) + T_2 \sum_{i=1}^k (i i Y); \\ b_{ii} = T_4(i i Y) + T_5 \sum_{i=1}^k (i i Y) + T_2(0 Y); \quad T_4 = T_{11} - T_5; \\ b_i = T_3(i Y); \quad b_{ij} = T_6(i j Y), \end{array} \right\} \quad (\text{III.185})$$

а также их ошибок и ковариаций:

$$\left. \begin{array}{l} s\{b_0\} = s_3 \sqrt{T_1} = s_3 T_7; \quad s\{b_{ii}\} = s_3 \sqrt{T_{11}} = s_3 T_9; \\ s\{b_i\} = s_3 \sqrt{T_3} = s_3 T_8; \quad s\{b_{ij}\} = s_3 \sqrt{T_6} = s_3 T_{10}; \\ \text{cov}\{b_0 b_{ii}\} = s_3^2 T_2; \quad \text{cov}\{b_{ii} b_{jj}\} = s_3^2 T_5. \end{array} \right\} \quad (\text{III.186})$$

$$\text{cov}\{b_0 b_{ii}\} = s_3^2 T_2; \quad \text{cov}\{b_{ii} b_{jj}\} = s_3^2 T_5. \quad (\text{III.187})$$

Из (III.182) и (III.187) следует, что все оценки b_i и b_{ij} независимы и поэтому любая из них может быть исключена из модели без пересчета оставшихся значимых (с риском α) оценок.

Блок-схема алгоритма расчета оценок коэффициентов ЭС-моделей с использованием соотношений (III.185) — (III.187) показана на рис. III.23.

Пример III.37. План ПФЭ 3^2 и результаты эксперимента* по определению прочности бетона в условиях пр. III.36 приведены в табл. III.30; там же в столбцах 10 — 15 дан расчет вспомогательных сумм для оценки коэффициентов модели по формулам (III.185) — (III.186).

На основе ковариационной матрицы (III.182) конкретизируются общие формулы (III.185):

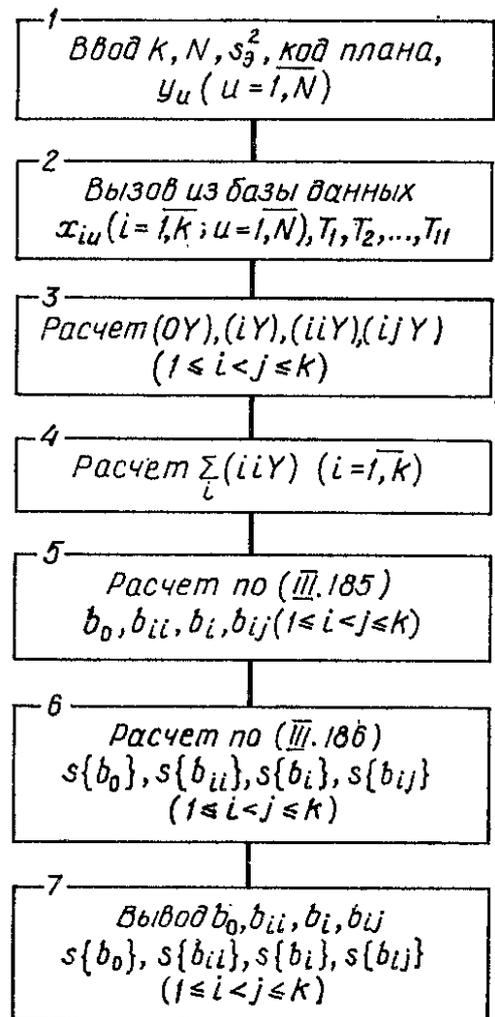


Рис. III. 23. Блок-схема алгоритма расчета оценок коэффициентов квадратичной ЭС-модели при реализации симметричного плана

* Здесь и далее в примерах (если специально не оговорено) «результаты эксперимента» точно совпадают с «теоретическими» величинами, чтобы для упрощения анализа остаточная дисперсия модели была $s_{\text{ост}}^2 = 0$.

$$\left. \begin{aligned} b_0 &= 5/9 (0Y) - 1/3 \sum_{i=1}^2 (iY); \\ b_{ii} &= 1/2 (iY) + 0 \sum_{i=1}^2 (iY) - 1/3 (0Y); \\ b_i &= 1/6 (iY); \quad b_{12} = 1/4 (12Y) \end{aligned} \right\} \quad (\text{III.188})$$

По (III.188) и данным табл. III.30 (столбцы 10 — 15) определяются оценки коэффициентов:

$$\begin{aligned} b_0 &= 5/9 \cdot 118,5 - 1/3(71,4 + 71,8) = 18,1; \\ b_{11} &= 1/2 \cdot 71,4 - 1/3 \cdot 118,5 = 3,8; \quad b_{22} = -3,6; \\ b_1 &= 0; \quad b_2 = 1/6 \cdot 2,4 = 0,4; \quad b_{12} = 2 \end{aligned}$$

начальной модели

$$\hat{R} = 18,1 - 3,8x_1^2 + 0,4x_2 - 3,6x_2^2 + 2x_1x_2. \quad (\text{III.189})$$

На основе (III.186) по (III.137) с учетом (III.182) рассчитывается гауссовская точность оценок коэффициентов

$$\begin{aligned} h\{b_{11}\} &= |-3,8| \sqrt{1/2} = 5,37; \quad h\{b_{22}\} = |-3,6| \sqrt{1/2} = 5,09; \\ h\{b_2\} &= 0,4/\sqrt{1/6} = 0,98; \quad h\{b_{12}\} = 2/\sqrt{1/4} = 4. \end{aligned}$$

При $s_9 = 0,5$ и $f_9 = \infty$ все значения больше, чем критический уровень $h_{кр} = 0,82$ ($t_{0,1} = 1,645$ — см. прил. II.5), поэтому модель (III.189) может быть использована для инженерного анализа. Общий вид поверхности отклика показан на рис. III.22.

Для построения двухфакторной квадратичной модели с 6 коэффициентами применение плана ПФЭ с 9 точками дает «перерасход» 3 опытов. На рис. III.24 показан взятый из каталога [58] ряд планов с меньшим числом точек. В табл. III.31 приведены некоторые статистические характеристики этих планов (при $s_9 = 1$), в том числе их относительная *эффективность* e , представляющая частное от деления критерия оптимальности (см. п. III.10) оптимального плана (III.164) на этот критерий данного плана.

План № 1. Вышерассмотренный план ПФЭ 3^2 имеет хорошую эффективность по критериям D , A , E и Q , но достаточно высокую избыточность опытов $e_N = N/L_{к.м}\{b\} = 1,5$. Он композиционен к плану ПФЭ 2^2 (к точкам 1 — 3 — 7 — 9 добавляются точки 2 — 4 — 5 — 6 — 8); его целесообразно применять всегда, когда *нет жестких ограничений* на ресурсы.

План № 2. Восемиточечный композиционный симметричный план B_3 (называемый «планом Бокса» по фамилии автора) за счет сокращения одной точки в центре квадрата ухудшил свою эффективность по всем критериям относительно плана № 1.

План № 3. Семиточечный план достаточно экономичен ($e_N = 1,16$) и композиционен к плану ПФЭ 2^2 ; эффективность его весьма высока, однако он резко асимметричен (в правой верхней части квадрата сосредоточено 4 точки, а в левой нижней — лишь 2, включая центральную № 5).

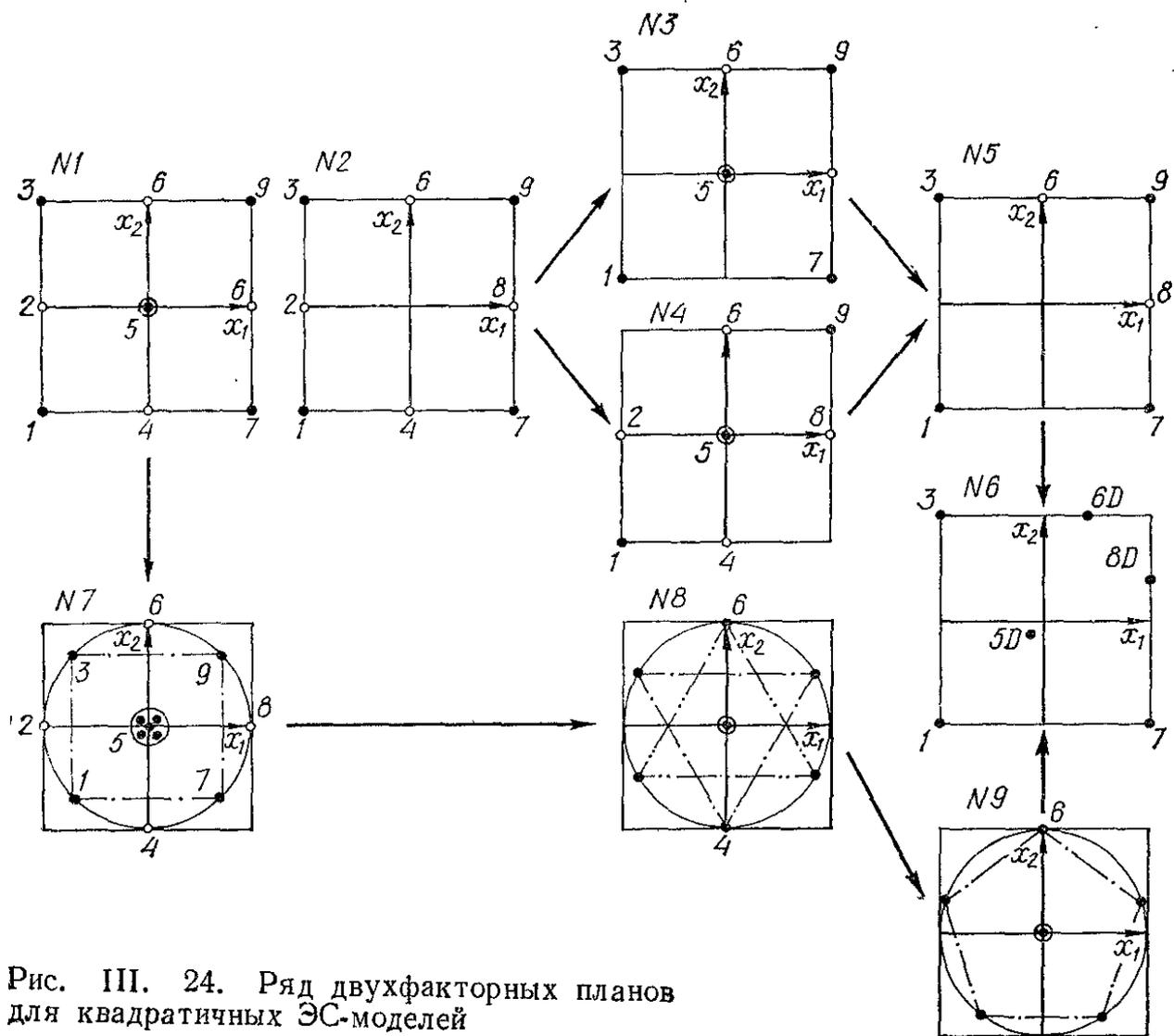


Рис. III. 24. Ряд двухфакторных планов для квадратичных ЭС-моделей

План № 4. В отличие от плана № 3 он не композиционен, но его точки симметричны относительно диагоналей. План имеет худшие (табл. III.31) показатели по A и E критериям, а также самую высокую ошибку в оценке коэффициентов модели — для эффекта взаимодействия $s\{b_{ij}\} = 1,58$.

Результат исключения (или замены) тех или иных точек из спектра плана ПФЭ 3^2 целесообразно анализировать по изолиниям среднеквадратичных ошибок предсказания $s\{\hat{Y}\}$ моделей, построенных по этим планам. На рис. III.25 показаны изолинии $s\{\hat{Y}\}$, рассчитанные по (III.144) для полной квадратичной модели, построенной по точкам планов № 1, 4, 5, 6 и 8 на рис. III.24 (границы анализируемого факторного пространства раздвинуты так же, как на рис. III.20).

Если для ПФЭ 3^2 (рис. III.25, а) в зоне эксперимента $s\{\hat{Y}\}$ изменится лишь в 1,5 раза и не превышает 0,9 в углах квадрата, то переход к плану Na_2 (с исключением двух опытов в противоположных вершинах) приводит к изменению $s\{\hat{Y}\}$ в 4,2 раза и увеличению $s\{\hat{Y}\}$ до 3,08 в исключенных точках. Из рис. III.25, б видно, что по диагонали с исключенными точками ошибка расчетов резко возрастает; прогноз поведения системы в таком направлении может привести технолога к неверным инженерным выводам.

Таблица III.31. Некоторые статистические характеристики двухфакторных планов на квадрате (№ 1—6) и на вписанной окружности (№ 7—9)

Номер плана	Название плана	Число точек	Эффективность					$s_{\max} \{\hat{Y}\}$	$\rho_{\max} = \rho \{b_{ii}\}$	$s_{\max} \{b\} = s \{b_{ii}\}$
			e_N	e_D	e_A	e_E	e_Q			
1	Полный факторный ПФЭ 3^2	9	1,50	0,99	0,97	0,75	0,95	0,90	-0,63	0,71
2	Бокса B_3	8	1,33	0,98	0,82	0,54	0,78	1,12	-0,76	0,87
3	Несимметричный композиционный	7	1,16	0,97	0,89	0,74	0,88	—	-0,49	0,95
4	Хартли Ha_2	7	1,16	0,80	0,60	0,37	0,72	3,08	-0,81	$s \{b_{ii}\} = 1,58$
5	Насыщенный композиционный	6	1	0,94	0,68	0,43	0,60	1,66	-0,73	1,22
6	Насыщенный точный D -оптимальный	6	1	0,95	0,84	0,67	0,85	1,36	-0,52	1,08
7*	Ротатабельный РЦКП-2 ($n_0 = 5$)	13	2,16	0,69	0,70	0,62	0,78	1,48 (0,79)	-0,58	0,76
8*	Симплекс-суммируемый	7	1,16	0,72	0,65	0,46	—	1,90 (1)	-0,81	1,22
9*	Ротатабельный D -оптимальный на пятиугольнике	6	1	0,73	0,67	0,49	—	2,05 (1)	-0,78	1,26

* Ротатабельные планы; значения $s_{\max} \{\hat{Y}\}$ в скобках соответствуют ограничению сферой $\rho = 1$.

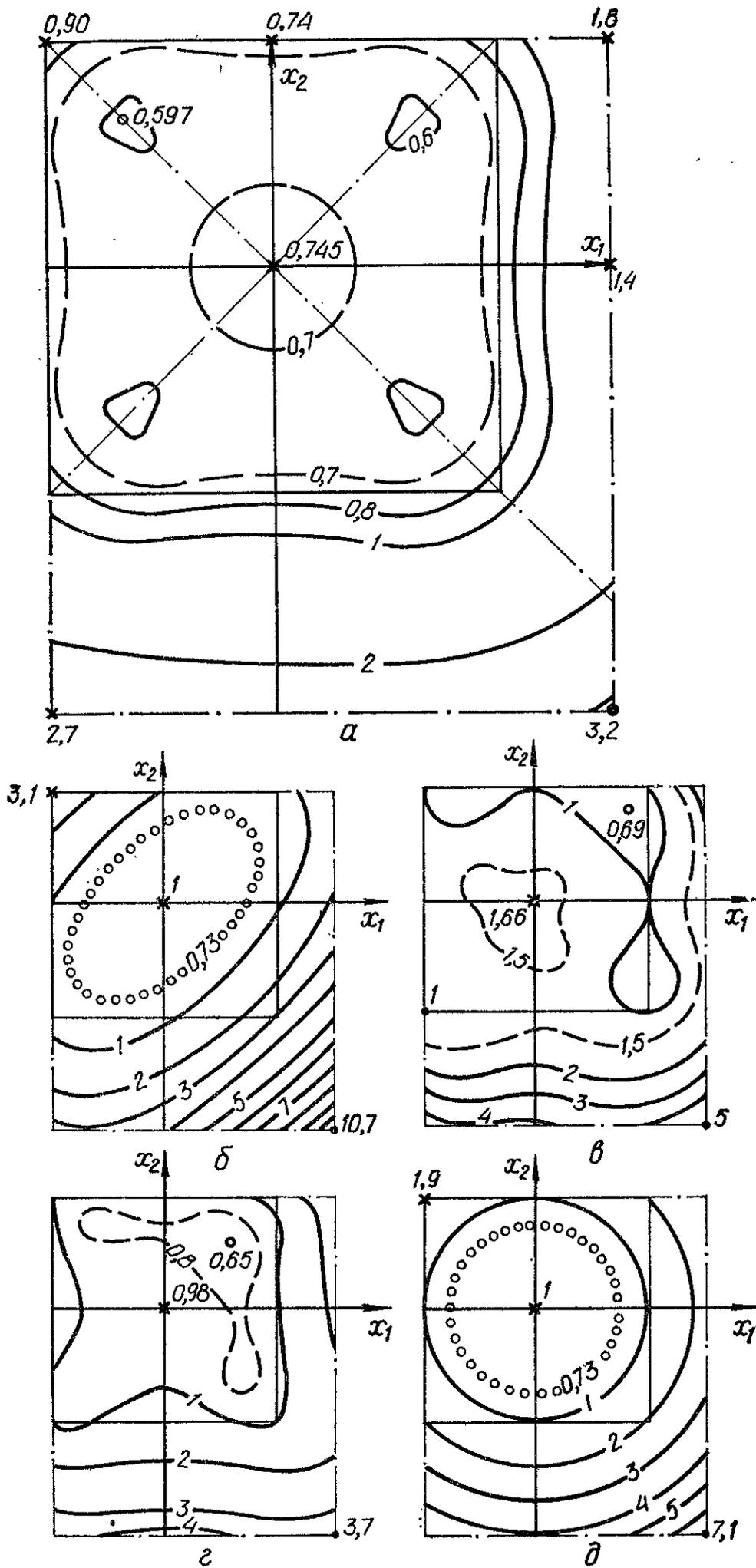


Рис. III 25 Изолинии ошибок предсказания для планов, представленных на рис. III. 24:

a — № 1; *б* — № 4; *в* — № 5; *г* — № 6; *д* — № 8

План № 5. Насыщенный план с 6 точками, композиционный к ПФЭ 2². По большинству критериев (особенно по D) он значительно лучше Na_2 . Сравнение изолиний $s\{\hat{Y}\}$ на рис. III.25, б и III.25, в показывает, что несмотря на то, что $s_{\min}\{\hat{Y}\} < 0,73$ локализуется в малой области (ср. с изолинией для плана Na_2) и средняя дисперсия предсказания (судя по Q -критерию в табл. III.31) несколько возрастает, максимальная ошибка $s_{\max}\{\hat{Y}\} = 1,66$ в центре плана № 5 вдвое меньше чем $s_{\max}\{\hat{Y}\} = 3,08$ в углах для плана Na_2 .

Многие недостатки планов № 2 — 5 можно уменьшить, если уйти от обязательного расположения опытов в точках-центроидах квадрата. Координаты новых точек планов, отвечающих заданным критериям оптимальности, определяются в процессе синтеза на ЭВМ оптимальных планов E^* (см. п. III.14).

План № 6. Насыщенный план, синтезированный на ЭВМ и оптимальный по D -критерию среди всех возможных планов с 6 точками. Все показатели эффективности у него лучше, чем у плана № 5 в точках-центроидах; изолинии $s\{\hat{Y}\}$ на рис. III.25, г для этого плана показывают большую однородность поля, чем для предыдущего, причем снижены и $s_{\min}\{\hat{Y}\} = 0,65$ и особенно $s_{\max}\{\hat{Y}\} = 1,36$. Этот план рекомендуется применять в условиях жестких ограничений на ресурсы.

Ковариационная матрица D этого плана не имеет блочно-диагональной структуры (III.182), поэтому формулы (III.185) не применимы и расчет оценок коэффициентов ведется по формуле (III.116) на основе L -матрицы [58]

$$L = 10^{-3} \begin{pmatrix} \text{№ 1} & \text{№ 3} & \text{№ 5} & \text{№ 6} & \text{№ 7} & \text{№ 8} & \text{— номер} \\ & & & & & & \text{опыта} \\ \left. \begin{array}{l} -46 & -3 & 979 & 37 & -3 & 37 \\ 174 & 296 & -576 & -512 & -61 & 678 \\ 174 & -61 & -576 & 678 & 296 & -512 \\ -198 & 268 & -172 & 204 & 232 & 204 \\ -198 & 232 & -172 & 204 & -268 & 204 \\ +302 & -268 & -172 & 204 & -268 & 204 \end{array} \right\} & \left. \begin{array}{l} \text{— для } b_0 \\ \text{— для } b_{11} \\ \text{— для } b_{22} \\ \text{— для } b_1 \\ \text{— для } b_2 \\ \text{— для } b_{12} \end{array} \right\} \cdot \end{pmatrix} \quad (III.190)$$

Пример III.38. В условия пр. III.36—37 для построения модели влияния добавок на прочность бетона по плану № 6 в табл. III.32 приведены координаты точек плана и результаты опытов. Расчет оценок коэффициентов в табл. III.32 заключается в умножении результатов опыта в данной точке на соответствующий элемент L -матрицы (III.190) и построением суммирования произведений.

На рис. III.24 и в табл. III.31 приведены и три ротатабельных плана № 7 — 9 на сфере $\rho = 1$ (в данном случае на круге, вписанном в квадрат). Как видно из рис. III.25, д, изолинии ошибки предсказания $s\{\hat{Y}\}$ для таких планов — концентрические окруж-

Таблица III.32. Построение квадратичной двухфакторной модели по насыщенному D -оптимальному плану № 6 с помощью L -матрицы

Номер опыта		№ 1	№ 3	№ 5	№ 6	№ 7	№ 8	Результаты расчета	
План эксперимента	нормализованный	x_1	-1	-1	-0,1	+1	+1	+0,4	—
		x_2	-1	+1	-0,1	+0,4	-1	+1	—
	натуральный, %	$D_{сп}$	0	0	0,36	0,8	0,8	0,56	—
		D_9	0	1,2	0,54	0,84	0	1,2	—
Результаты опытов R_u , МПа			12,3	9,1	18,0	15,2	8,3	14,7	—
Расчет оценок коэффициентов	b_0	-0,57	-0,03	17,62	0,56	-0,02	+0,54	18,1	
	b_{11}	2,14	2,69	-10,37	-7,73	-0,51	9,97	-3,81	
	b_{22}	2,14	-0,56	-10,37	10,24	2,46	-7,53	-3,62	
	b_1	2,44	-2,44	-3,1	3,1	1,93	3	0,05	
	b_2	-2,44	2,11	-3,1	3,1	-2,22	3	0,45	
	b_{12}	+3,71	-2,44	-3,1	3,1	-2,22	3	2,05	

Построенная в табл. III.32 модель

$$\hat{R} = 18,1 - 3,8x_1^2 + 0,4x_2 - 3,6x_2^2 + 2x_1x_2$$

совпадает с моделью (III.189) в пр. III.37.

ности. Для снижения $s\{\hat{Y}\}$ в центре факторного пространства, т. е. для оптимизации плана по критерию *униформности* (см. п. III.10), становится в *центральной* точке № 5 *больше опытов*, чем в точках на периферии (план № 5 при пяти точках в центре имеет $s\{\hat{Y}\} = 0,45$ вместо $s\{\hat{Y}\} = 1$ при одной точке). Характеристики планов на сфере (табл. III.31), как правило, существенно хуже, чем планов на квадратах (особенно по D -критерию и по значению $s\{\hat{Y}\}$ в углах квадрата), поэтому их следует применять только в задачах, где превалирующую роль играет равномерность ошибки предсказания по направлению от центра. В этом случае рекомендуется план № 8 (симплекс — суммируемый из двух треугольников с координатами точек основания $x_1 = \pm \sqrt{3/2}$ и $x_2 = \pm 1/2$).

Все вышеизложенные общие положения о планировании эксперимента при построении квадратичных, двухфакторных ЭС-моделей применимы и при большем числе факторов, хотя для каждой размерности k есть свои особенности. Следует отметить, что с увеличением числа факторов k особое значение, кроме D -оптимальности, приобретает критерий избыточности e_N , поэтому, как правило,

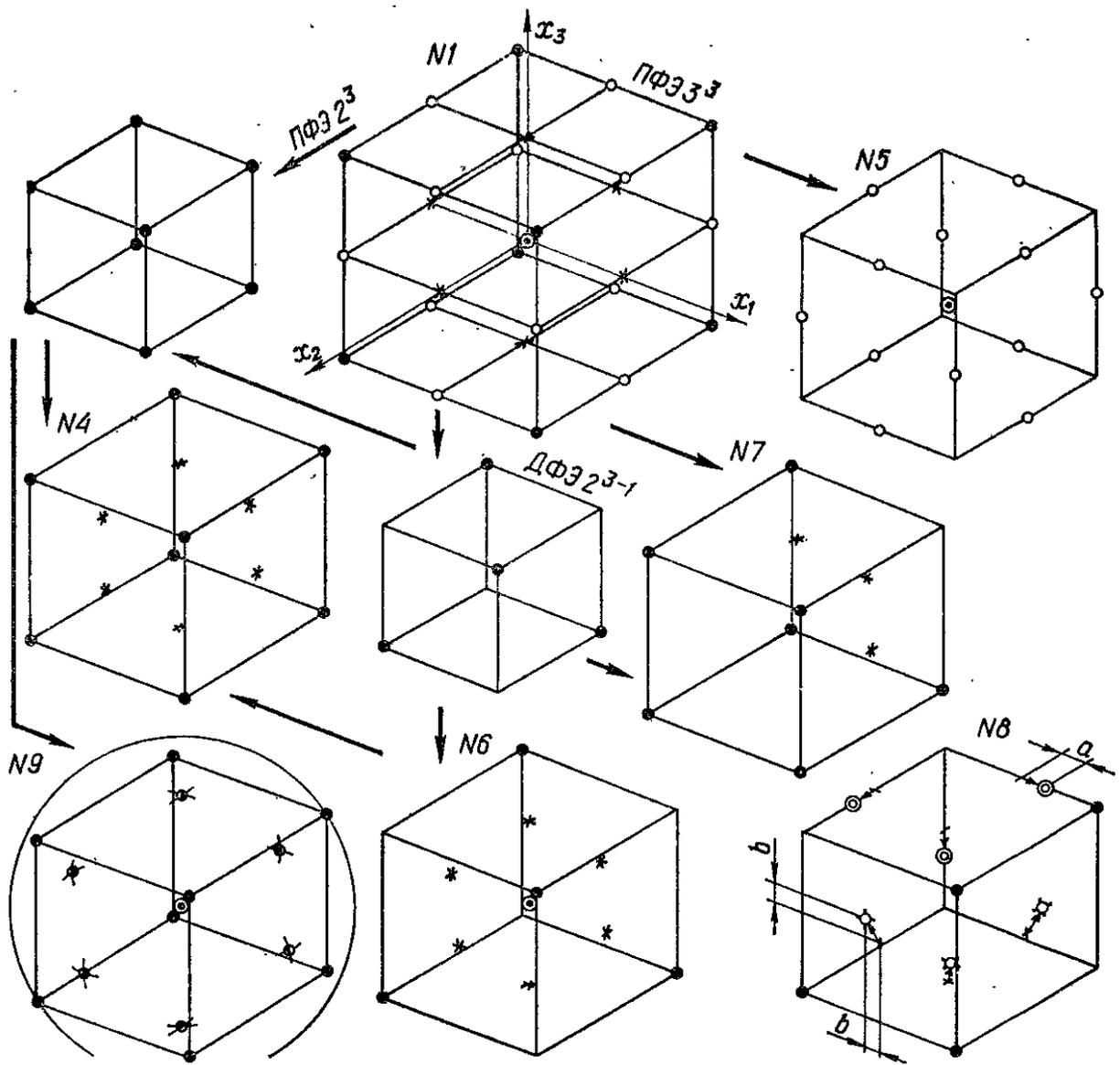


Рис. III. 26. Схема образования трехфакторных планов из 27 точек ПФЭ

применяются планы одновременно близкие к точным D -оптимальным и насыщенным.

Для трехфакторных моделей ряд планов [4, 58], в том числе связанных с ПФЭ 3^3 , показан на рис. III.26, а их статистические характеристики приведены в табл. III.33 (при $s_3 = 1$).

План № 1. План ПФЭ 3^3 имеет наилучшие характеристики (кроме величин $\rho\{b_{0b_{ii}}\}$, но его применение дает большой перерасход опытов для оценки 10 коэффициентов модели ($e_N = 2,7$).

Планы № 2 и 5. План Бокса — Бенкина $ВВ_3$ представляет собой выборку из 12 точек на серединах ребер куба плюс центральная точка (реализуемая три раза в плане № 2, отвечающем критерию равномерности). Эти 12 точек образуют три блока ПФЭ 2^2 ($\pm 1, \pm 1, 0; \pm 1, 0, \pm 1; 0, \pm 1, \pm 1$) на плоскостях, проходящих через центр куба. План ротатабелен, так как 12 точек лежат на поверхности сферы с радиусом $\rho = \sqrt{2}$. Планы по характеристикам (табл. III.33) весьма посредственны и их можно применять в условиях приоритета ротатабельности или при необходимости достройки плана 2^2 .

Планы № 3 и 4. План Бокса (с центральной точкой — план № 3 или без нее) имеет очень хорошие характеристики и композиционен

Таблица 111.33. Некоторые статистические характеристики трехфакторных планов на кубе

Номер плана	Название плана	Число точек N	Эффективность					$s_{\max} \{\hat{Y}\}$	$\rho_{\max} = \rho \{b_0 b_{ii}\}$	$s_{\max} \{b\} = s \{b_{ii}\}$ (кроме $s \{b_0\}$)
			e_N	e_D	e_A	e_E	e_Q			
1	ПФЭ 3^3	27	2,7	0,97	0,97	0,67	0,94	0,71	0,53	0,41
2	Некомпозиционный Бокса—Бенкина BB_3 ($n_0 = 3$)	15	1,5	0,88	0,94	0,74	0,96	1,18	0,55	0,52
3	Композиционный B_3 ($n_0 = 1$)	15	1,5	0,99	0,97	0,78	0,95	—	0,38	0,63
4	B_3 ($n_0 = 0$)	14	1,4	0,99	0,97	0,77	0,95	0,89	0,38	0,63
5	BB_3 ($n_0 = 1$)	13	1,3	0,90	0,82	0,47	0,83	1,25	0,75	0,66
6	Хартли Ha_3	11	1,1	0,78	0,72	0,63	0,7	—	$\rho \{b_k b_{ij}\} = 0,81$	$s \{b_{ij}\} = 0,87$
7	Рехтшафнера R_3	10	1	0,92	0,72	0,44	0,69	—	0,53	1
8	Насыщенный точный D -оптимальный	10	1	0,94	0,88	0,6	0,86	—	0,5	0,77
9	Ротатабельный на сфере РЦКП-3 ($n_0 = 6$)	20	2	0,52	0,52	0,5	0,59	2,32	0,52	$s \{b_{ij}\} = 1$

к ПФЭ 2^3 — к вершинам куба прибавлены 6 точек в центре граней. Его можно рекомендовать к применению во всех случаях, когда к планам не предъявляются какие-либо особые требования.

План № 6. Весьма экономичный план Хартли H_3 , композиционный к полуреплике ПФЭ 2^3 , имеет низкую эффективность и высокую закоррелированность линейных эффектов и эффектов взаимодействия.

План № 7. Насыщенный план Рехтшафнера R_3 также композиционен к ДФЭ 2^{3-1} (добавлены три вершины и три точки на середине граней, образующих ту вершину, где эксперимент не ставится). План D -оптимален при невысокой корреляции $\rho\{b_0 b_{ii}\}$, но по остальным критериям эффективность не высока. Поскольку план имеет простую L -матрицу, то его можно рекомендовать в тех случаях, когда у исследователя нет оперативного доступа к ЭВМ.

План № 8. Насыщенный точный D -оптимальный план, у которого эффективность выше, чем у планов № 6 и 7. План рекомендуется во всех случаях, когда возможен расчет оценок коэффициентов на ЭВМ, а также когда нет сложностей с установкой уровней факторов при эксперименте в точках, отличных от центроидов («сдвижки» по ребрам $a = 0,3$, по граням $b = 0,2$).

Для сравнения с планами на кубе приведен и план на сфере РЦКП (на рис. III.26 для наглядности сфера описана вокруг куба, отчего относительные показатели плана e_i в табл. III.33 изменяются незначительно), применение которого ограничено задачами, в которых, во-первых, принцип ротатабельности превалирует над другими требованиями к модели, во-вторых, на первом этапе обязательен эксперимент ПФЭ 2^3 .

Пример III.39. Исследовалось изменение предела прочности на сжатие R , МПа, мелкозернистых бетонов из равноподвижных смесей, твердевших в условиях ТВО [6]. Три основных рецептурно-технологических фактора изменялись по плану № 3 (табл. III.33 и рис. III.26), содержащему 15 точек, уровни которых показаны в табл. III.34 (столбцы «Эффекты линейные»). Отношение П/Ц = $X_1 = (2,5 \pm 0,5)/1$ характеризует расход основного материального ресурса — цемента (нормализация проведена аналогично пр. III.2, т. е. П/Ц = 3 совмещено с $x_1 = -1$); чем больше П/Ц (меньше x_1), тем экономичнее рецептура. Концентрация суперпластификатора $C_{сп} = X_2 = 0,35 \pm \pm 0,35, \%$ от массы цемента — главный регулирующий фактор. Температура $T = X_3 = 60 \pm 20$ °С при заданной длительности ТВО характеризует энергоемкость этого технологического передела; этот фактор необходимо минимизировать.

После реализации эксперимента по приведенному в табл. III.34 плану (с рандомизацией порядка опытов) по средним результатам $Y_n = \bar{R}$ получена модель

$$\begin{aligned} \hat{R} = & 19,5 + 5,5x_1 + 1,9x_1^2 + 2,0x_1x_2 + 1,5x_1x_3 + \\ & + 0,8x_2 - 1,8x_2^2 + 0,4x_2x_3 + \\ & + 6,8x_3 + 0,6x_3^2 \end{aligned} \quad (\text{III.191})$$

Оценки коэффициентов в (III.191) могут быть рассчитаны с помощью микрокалькуляторов по формулам (III.185), причем удобно пользоваться

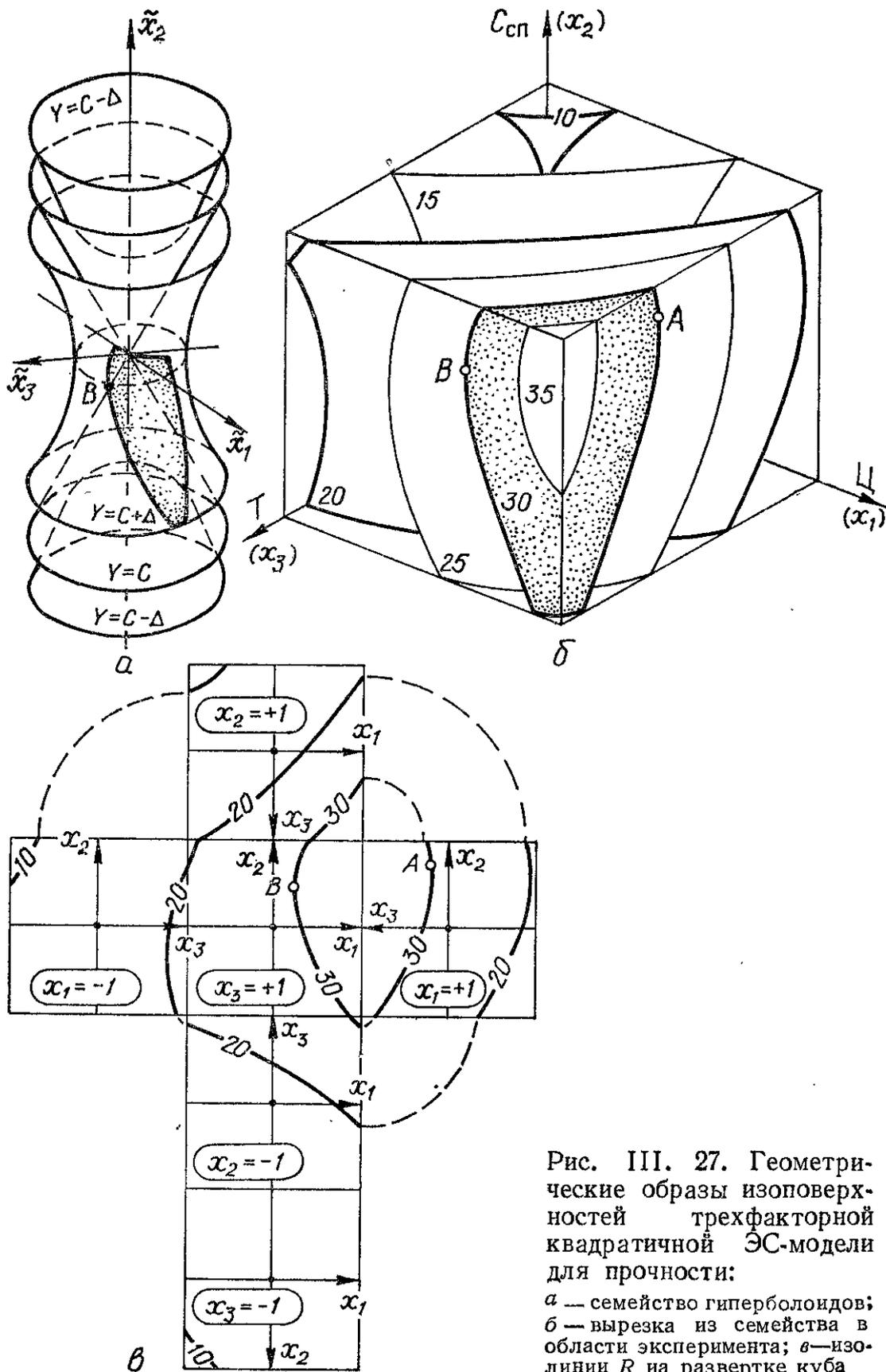


Рис. III. 27. Геометрические образы изоповерхностей трехфакторной квадратичной ЭС-модели для прочности:

a — семейство гиперболоидов;
б — вырезка из семейства в области эксперимента; *в* — изолинии *R* на развертке куба

так называемым [4] бланком-алгоритмом, фрагментом которого является табл. III.34. На нем сначала подсчитываются десять сумм произведений $(0Y), \dots, (23Y)$, представляющих собой элементы матрицы моментов плана *M*. Далее с использованием коэффициентов *T* (см. прил. II.12) рассчитываются промежуточные величины ρ_i и оценки коэффициентов. Все эффекты модели (III.191) значимы при $\alpha = 0,1$ ($s_{\text{э}} = 0,63$ МПа, $f_{\text{э}} = 30$). Расчетные значения отклика в точках плана показаны в правой части табл. III.34. Модель адекватна, поскольку $F_a = (s_{\text{ост}}^2 = s_{\text{на}}^2 / s_{\text{э}}^2 = 0,02 / 0,63^2 < 1$.

Таблица 111.34. Фрагмент бланка-алгоритма для построения трехфакторной

Номер		План в натуральных переменных			$Y_3 = \bar{R},$ МПа	$x_0 \equiv$ $\equiv +1$	Расчетная		
опыта	реализации	$X_1 =$ $= П/Ц$	$X_2 = С,$ %	$X_3 = Т,$ °С			Эф		
							линейные		
					x_1	x_2	x_3		
1	8	2	0,7	80	37,3	+	+	+	+
2	10	2	0,7	40	19,9	+	+	+	-
3	9	2	0	80	30,9	+	+	-	+
4	2	3	0,7	80	19,1	+	-	+	+
5	7	3	0	40	10,9	+	-	-	-
6	4	3	0	80	20,9	+	-	-	+
7	12	3	0,7	40	7,9	+	-	+	-
8	3	2	0	40	15,1	+	+	-	-
9	13	2	0,35	60	26,9	+	+	0	
10	6	3	0,35	60	15,9	+	-		0
11	11	2,5	0,7	60	18,6	+		+	
12	15	2,5	0	60	16,8	+	0	-	
13	5	2,5	0,35	80	26,8	+		0	+
14	14	2,5	0,35	40	13,4	+		0	-
15	1	2,5	0,35	60	19,5	+	0	0	0
						(0Y)	(1Y)	(2Y)	(3Y)
						299,9	+55,4	+8,2	+67,8
$\rho_1 = -1/9 (0Y) = 33,32$						$b_l = 0,1 (iY)$			
$\rho_2 = 13/45 (0Y) = 86,64$						b_1	b_2	b_3	
$b_0 = \rho_2 + \rho_3 = 19,48$						+5,54	+0,82	+6,78	
						$\rho_6 = 0,5 (iiY)$			
						$b_{ii} = \rho_5 + \rho_6$			

квадратичной ЭС-модели по плану В₃ (n₀ = 1)

матрица фекты						Расчетные результаты		
квадратичные			взаимодействия			$\hat{Y} = \hat{R}$, МПа	Δ_u	Δ_u^2
x_1^2	x_2^2	x_3^2	x_1x_2	x_1x_3	x_2x_3			
+	+	+	+	+	+	37,2	+0,1	0,01
+	+	+	+	-	-	19,8	+0,1	0,01
+	+	+	-	+	-	30,8	+0,1	0,01
+	+	+	-	-	+	19,2	-0,1	0,01
+	+	+	+	+	+	11,0	-0,1	0,01
+	+	+	+	-	-	20,8	+0,1	0,01
+	+	+	-	+	-	7,8	+0,1	0,01
+	+	+	-	-	+	15,0	+0,1	0,01
+	0					26,9	0	0
+						15,9	0	0
	+	0				18,5	+0,1	0,01
	+		0	0	0	16,9	-0,1	0,01
0		+				26,8	0	0
0	0	+				13,3	+0,1	0
0	0	0	0	0	0	19,5	0	0
(11Y)	(22Y)	(33Y)	(12Y)	(13Y)	(23Y)			
204,8	197,4	202,2	+16	+12	+2,8			
$\sum (iiY) = 604,4$			$b_{ij} = 1/8 (ijY)$					
$\rho_3 = \rho_4 = -1/9 \sum (iiY)$			b_{12}	b_{13}	b_{23}			
$\rho_3 = \rho_4 = -67,16$			+1,99	+1,46	+0,41			
$\rho_8 = \rho_1 + \rho_4 = -100,47$								
102,4	98,7	101,1						
b_{11}	b_{22}	b_{33}						
+1,93	+0,63	-1,77						

$$f_{\text{ост}} = 15 - 10 = 5$$

$$SS_{\text{ост}} = 0,1;$$

$$s_{\text{ост}}^2 = 0,02;$$

$$s_{\text{ост}} = 0,14$$

Модель (III.191) описывает семейство поверхностей второго порядка, инварианты которых (см. прил. I.2)

$$I_1 = b_{11} + b_{22} + b_{33} = 1,9 - 1,8 + 0,6 = 0,7,$$

$$I_2 = 1,9(-1,8) + (-1,8) \cdot 0,6 + 1,9 \cdot 0,6 + 0,25(2^2 + 1,5^2 + 0,4^2) = -1,76,$$

$$I_3 = 1,9(-1,8) \cdot 0,6 + 0,25 \cdot 2 \cdot 1,5 \cdot 0,4 - 0,25 (1,9 \cdot 0,4^2 - 1,8 \cdot 1,5^2 + 0,6 \cdot 2^2) = -1,42$$

позволяют заключить, что это семейство гиперболоидов ($I_3 \neq 0$; $I_1 I_3 < 0$; $I_2 < 0$), показанное на рис. III.27, а. Ограничивающие факторное пространство плоскости вычлениают часть этого семейства, главные оси которого \bar{X}_i повернуты на некоторый угол по отношению к факторным осям x_i (на рис. III.27, б они условно смещены из центра на грани куба и поэтому взяты в скобки). Аксонометрическое изображение изоповерхностей $\bar{R} = \text{const}$ достаточно просто построить, пользуясь разверткой поверхности куба (рис. III.27, в), состоящей из шести квадратов $\{x_i x_k\}$ при $x_i = -1 \vee +1 (i = \bar{1}, \bar{3})$. Изоповерхности выпуклы в сторону пониженного расхода цемента (точка В) и уменьшенной температуры ТВО (точка А), что позволяет найти (см. п III 15, III.16) ряд эффективных инженерных решений для исследованного бетона.

Планы для квадратичных моделей при $k > 3$ построены по тем же принципам, которые использованы выше для планов $k \leq 3$. В каталогах [58] приведены матрицы планов, их D- и L-матрицы, а также статистические характеристики аналогично табл. III.31 и III.33. Процедуры планирования эксперимента по таким планам, а также расчета оценок коэффициентов рассмотрены достаточно подробно на примерах в работах [4, 18, 23, 48, 57 и др.]. Для $k = 3 \dots 7$ в прил. II.12, II.13 приведены те планы, которые можно считать лучшими как для учебных работ, так и для исследовательских в области материаловедения и технологии, если к планам предъявляются специальные требования, сформулированные в п. III.13 и III.14, посвященных планированию эксперимента в особых условиях, вытекающих из инженерной постановки задачи.

III.13. Моделирование свойств смесей

Используемые для получения строительных материалов и изделий сырьевые материалы (вяжущие, заполнители и наполнители, химические добавки и т. д.) могут представлять собой смеси из q различных веществ, минералов, зерен разной крупности и т. п. Их состав (рецептура) задается концентрациями компонентов в виде массовых, объемных или мольных долей (процентов) v_i , причем

$$0 \leq v_i \leq 1, \quad \sum_{i=1}^q v_i = 1. \quad (\text{III.192})$$

Системы, свойства которых зависят только от соотношения компонентов $\mathbf{v} = (v_1, v_2, \dots, v_q)$ (III.192) и не зависят от количества смеси, а также от условий ее переработки и других факторов, называются *системами «состав—свойства»* или *«смесь — свойства»* и обозначаются MQ (аббревиатура от англ. *mixture* — смесь и *quality* — свойство, качество). Условие (III.192) определяет рецеп-

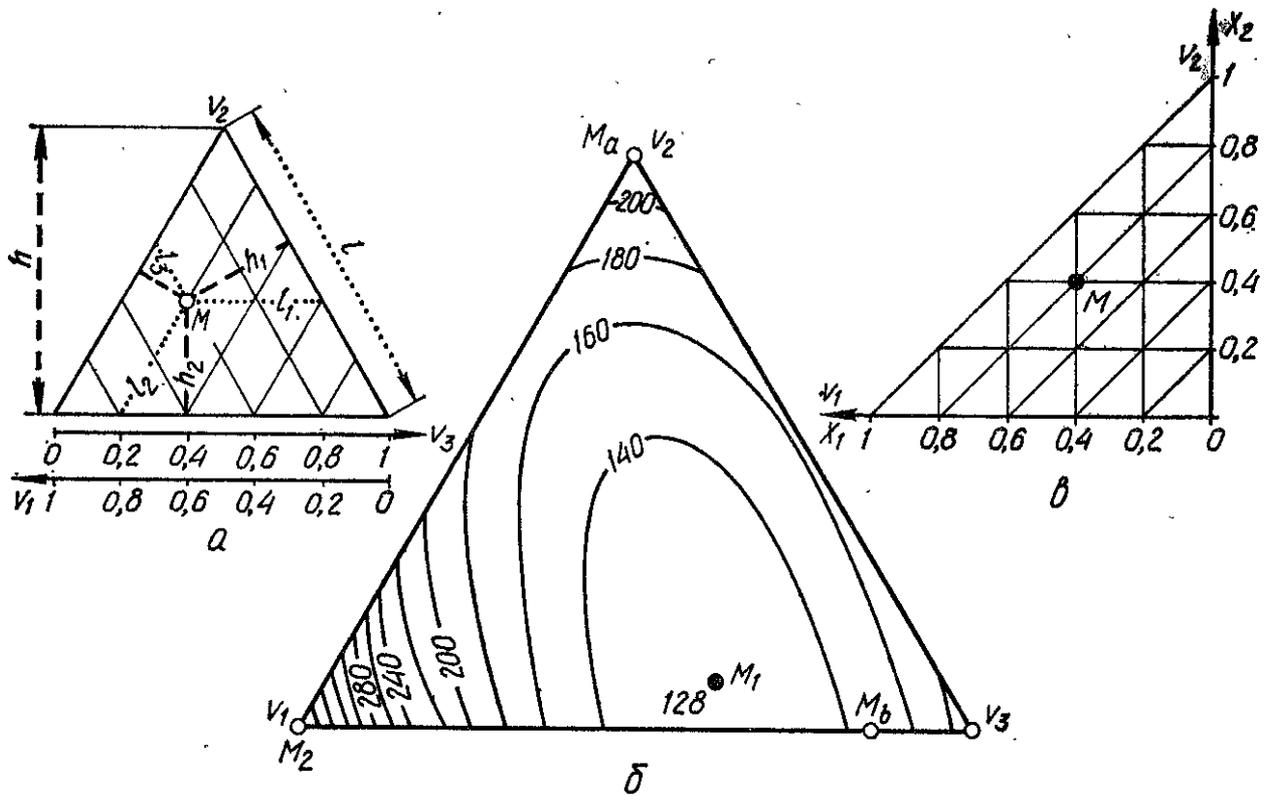


Рис. III. 28. Отображение на «смесевом» треугольнике:

a — зернового состава наполнителя; *b* — зависимости эффективной вязкости композиции от состава наполнителя; *c* — аналог «*a*» при исключении третьего компонента X_3

туру как систему с q линейно связанными элементами v_i , число степеней свободы которой равно $(q - 1)$. Ее геометрическим аналогом является $(q - 1)$ -мерный правильный симплекс (анг. *simplex* — простой), т. е. выпуклый многогранник, не имеющий диагональных сечений. Для двухкомпонентной системы это отрезок прямой, для $q = 3$ — равносторонний треугольник, для $q = 4$ — тетраэдр.

При изучении систем MQ факторы $X_i = v_i$ варьируются на симплексе. Зависимость $Y(\mathbf{v})$ может быть представлена как диаграмма «состав — свойство». Каждому j -му составу смеси \mathbf{v}_j , определяющему уровень свойства $Y(\mathbf{v}_j)$, соответствует точка симплекса: чистым веществам (или зернам одного размера, или порошку одного минерала) — вершины симплекса, двойным смесям — точки ребер, тройным системам — точки поля треугольника. Для любой точки M на треугольнике (рис. III.28) содержание i -го компонента $v_i = l_{im}/l$ пропорционально расстоянию l_{im} от M до стороны, противоположной i -й вершине: по перпендикуляру, если за единицу принята высота h_i , или по прямой, параллельной стороне, если за единицу принята сторона l . Свойство Y в зависимости от состава трехкомпонентной смеси представляют изолиниями в поле концентрационного треугольника.

При $q = 4$ возможно объемное изображение изоповерхностей свойства в теле тетраэдра, поверхность которого отображается разверткой из четырех треугольных диаграмм. В связи с удобством графической интерпретации результатов моделирования треугольная диаграмма является основным информационным элементом при

анализе многокомпонентных систем MQ . Поскольку технические свойства (в отличие, например, от температур фазовых переходов стекол, клинкеров, сплавов и т. п.) определяются, как правило, не только исходным составом смеси, но и условиями ее переработки в изделие, то при анализе этих свойств система и соответствующие диаграммы «состав — свойство» рассматриваются для фиксированных значений других рецептурно-технологических и эксплуатационных факторов.

Пример III.40. В качестве наполнителя для композиций на полиэфирной смоле (связующее для полимербетона) рассматриваются смеси трех фракций частиц мрамора [8, 13]. Дисперсионный состав наполнителя задается долями зерен: мелких v_1 (с удельной поверхностью $S_1 = 340 \text{ м}^2/\text{кг}$), средних v_2 ($S_2 = 100 \text{ м}^2/\text{кг}$) и крупных v_3 ($S_3 = 50 \text{ м}^2/\text{кг}$), причем $v_1 + v_2 + v_3 = 1$. Треугольник на рис. III.28 отображает все множество зерновых составов. Так, состав в точке M : 40 % мелких, 40 % средних и 20 % крупных частиц. На диаграмме (рис. III.28, б) изолинии отражают зависимость эффективной вязкости композиции η , Па · с, от зернового состава мраморного наполнителя при фиксированных значениях скорости деформации этих неньютоновских жидкостей ($\dot{\epsilon} = 1 \text{ с}^{-1}$), концентрации наполнителя (массовое отношение $N/P = 2,5$) и других факторов, определяющих ее свойства.

На гранулометрическом треугольнике одна и та же удельная поверхность образуется при разных зерновых составах, т. е. при разных распределениях частиц по размерам. Так, $S = 100 \text{ м}^2/\text{кг}$ соответствует и мононаполнителю средней фракции (вершина $M_a - v_2 = 1, v_1 = v_3 = 0$), и смеси мелких и средних частиц (точка M_b на нижней стороне треугольника, где $v_1 = 17,2 \%$, $v_3 = 82,8 \%$ и в соответствии с правилом пропорции $S = 0,172 \cdot 340 + 0,828 \times 50 = 100 \text{ м}^2/\text{кг}$). Однако $\eta\{M_b\} < \eta\{M_a\}$ почти в 1,5 раза за счет лучшего распределения размеров частиц дисперсной фазы, которое оказывается существенным для структурообразования в рассмотренных дисперсных системах [13]. Зона пониженной вязкости характерна для бинарных и тройных смесей частиц наполнителя.

Построение и анализ диаграмм «состав — свойство» осуществляются на основе ЭС-моделей зависимостей свойств от состава смеси. *Полный полином* (III.4) степени m от q смесевых факторов v_i в системе MQ не может быть использован в качестве формы таких моделей: в силу линейной связи факторов (III.192) члены такого полинома оказываются линейно зависимыми, т. е. не представляют систему базисных функций при разложении Y по C_{q+m}^m составляющим вектора $f^T(v) = (1, v_1, \dots, v_q, v_1v_2, \dots, v_1^2, \dots, v_q^2, \dots)$.

Возможно описание системы полиномом (III.4) от $k = q - 1$ факторов с C_{q+m-1}^m коэффициентами, если исключить один из компонентов:

$$v_q = 1 - \sum_{i=1}^{q-1} v_i, \quad (\text{III.193})$$

но при этом факторное пространство трансформируется так, что прямое влияние этого компонента не прослеживается. Так, для трех факторов v_i равносторонний треугольник (рис. III.28, а) превращается в прямоугольный, по катетам которого проходят только две факторные оси — X_1 и X_2 (рис. III.28, в), что и приведет к двухфакторной модели типа (III.179).

Поскольку желательнее, чтобы описание системы отражало воздействие на нее всех компонентов, в качестве моделей для систем MQ , как правило, используют предложенные Шеффе приведенные полиномы с числом коэффициентов C_{q+m-1}^m . Такой полином второй степени для тройной системы получается из обычного полинома

$$Y = a_0 + a_1v_1 + a_2v_2 + a_3v_3 + a_{12}v_1v_2 + a_{13}v_1v_3 + a_{23}v_2v_3 + a_{11}v_1^2 + a_{22}v_2^2 + a_{33}v_3^2 \quad (\text{III.194})$$

после вытекающих из (III.192) подстановок:

$$a_0 = a_0v_1 + a_0v_2 + a_0v_3; \quad (\text{III.195})$$

$$v_1^2 = v_1 - v_1v_2 - v_1v_3,$$

$$v_2^2 = v_2 - v_1v_2 - v_2v_3, \quad (\text{III.196})$$

$$v_3^2 = v_3 - v_1v_3 - v_2v_3;$$

при обозначении

$$A_i = a_0 + a_i + a_{ii}, \quad A_{ij} = a_{ij} - a_{ii} - a_{ij} \quad (\text{III.197})$$

получается приведенный полином

$$Y = A_1v_1 + A_2v_2 + A_3v_3 + A_{12}v_1v_2 + A_{13}v_1v_3 + A_{23}v_2v_3, \quad (\text{III.198})$$

в котором 6 коэффициентов A заменяют 10 коэффициентов полинома (III.194). Аналогично записывается приведенный полином второй степени от q переменных

$$Y = \sum_{1 \leq i \leq q} A_i v_i + \sum_{1 \leq i < j \leq q} A_{ij} v_i v_j. \quad (\text{III.199})$$

Коэффициенты приведенных полиномов имеют физический смысл: A_i равны величине Y (значению свойства) для «чистого» i -го компонента; нелинейную часть таких моделей называют синергизмом (греч. *sinergeia* — содружество), если смешивание компонентов вызывает увеличение отклика ($A_{ij} > 0$), или антагонизмом (греч. *antogōnisma* — борьба) при его уменьшении ($A_{ij} < 0$).

Пример III.41. Представленная в пр. III.40 диаграмма «гранулометрический состав — вязкость» (рис. III.28, б) описывается моделью

$$\hat{Y} = \ln \eta = 6,03v_1 + 5,4v_2 + 5,29v_3 - 2,06v_1v_2 - 3,01v_1v_3 - 1,04v_2v_3. \quad (\text{III.200})$$

Коэффициенты при линейных членах равны логарифму вязкости композиций с мононаполнителями (в вершинах концентрационного треугольника). Так, в точке M_2 вязкость композиции на мелком порошке $\eta \{v_1 = 1, v_2 = v_3 = 0\} = e^{6,03} = 412$ Па · с. Смешивание фракций, особенно мелкой и крупной, значительно снижает вязкость, что видно по отрицательным значениям коэффициентов при $v_i v_j$, и нашло отражение на диаграмме: минимум вязкости в точке M_1 ($v_1 = 0,35; v_2 = 0,09; v_3 = 0,56; \eta_{\min} = 128$ Па · с).

Поверхности отклика в многокомпонентных системах, как правило, имеют сложный характер. Редкие показатели (стоимостные критерии, плотности жидкостей и т. п.) адекватно описываются

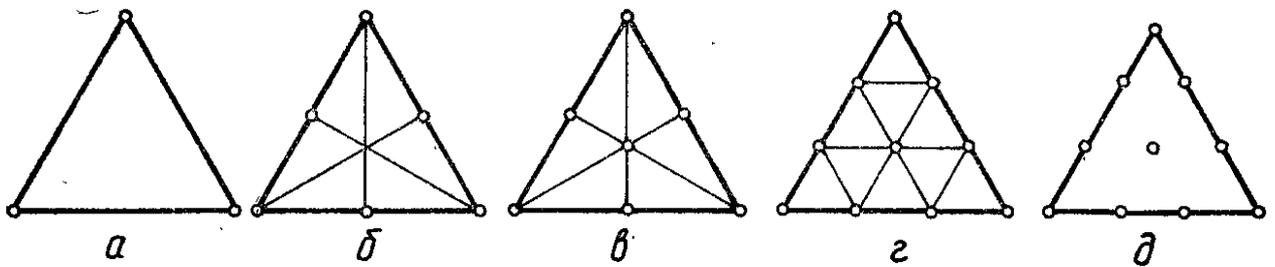


Рис. III. 29. Симплекс-решетчатые планы разных порядков (a — первого; $б$ — второго; $в$ — неполного третьего; $г$ — третьего) и D -оптимальный третьего порядка ($д$)

приведенными полиномами первой степени $Y = \sum_{1 \leq i < q} A_i v_i$. На практике используются приведенные полиномы до четвертой степени [1, 47], наиболее часто — второго порядка (III.198), (III.199). Для трехкомпонентных систем полиномы неполного третьего и третьего порядков имеют вид:

$$Y = \sum_{i=1}^3 A_i v_i + \sum_{i < j < 3} A_{ij} v_i v_j + A_{123} v_1 v_2 v_3; \quad (\text{III.201})$$

$$Y = \sum_{i=1}^3 A_i v_i + \sum_{i < j < 3} A_{ij} v_i v_j + A_{123} v_1 v_2 v_3 + \\ + A_{1-2} v_1 v_2 (v_1 - v_2) + A_{1-3} v_1 v_3 (v_1 - v_3) + A_{2-3} v_2 v_3 (v_2 - v_3). \quad (\text{III.202})$$

Коэффициенты приведенных полиномов определяются МНК (см. п. III.7) по значениям факторов и откликов в эксперименте. Наименьшее число точек в спектре плана эксперимента для оценки коэффициентов приведенного полинома степени m от q переменных равно C_{q+m-1}^m . Чаще всего используются предложенные Шеффе насыщенные симплекс-решетчатые планы — совокупность узлов сетки с шагом $1/m$ на $(q-1)$ -мерном симплексе. На рис. III.29, a — $г$ показаны симплекс-решетчатые планы первого, второго, неполного третьего и третьего порядков для $q=3$, позволяющие оценить коэффициенты приведенных полиномов соответствующей степени: в первых трех случаях планы D - и G -оптимальны (планы первого порядка оптимальны по всем критериям, связанным с матрицей D). На рис. III.29, $д$ показан D -оптимальный план Кифера третьего порядка, в котором точки на ребрах симплекса сдвинуты ближе к вершинам по сравнению с симплекс-решеткой. Матрицы этих планов приведены в прил. II.14. Там же записаны вытекающие из особенностей рассмотренных моделей и планов простые формулы для расчета оценок коэффициентов приведенных полиномов [1, 47]. Те же оценки дает и общая для линейных по параметрам моделей формула (III.114), в которой $\mathbf{b} = (A_1, \dots, A_q, A_{12}, \dots)^T$ и $\mathbf{x} = (v_1, \dots, v_q, v_1 v_2, \dots)^T$.

Пример III.42. Оценки коэффициентов в модели (III.200) рассчитаны по результатам эксперимента, представленным в табл. III.35 (индексы i при y в точках плана указывают на наличие доли i -го компонента в составе смеси);

Таблица III. 35. План и результаты эксперимента в системе «зерновой состав наполнителя — вязкость композиций»

Номера точек плана	Факторы			η , Па·с	$y = \ln \eta$
	v_1	v_2	v_3		
1	1	0	0	412,4	$y_1 = 6,02$
2	0	1	0	221,8	$y_2 = 5,40$
3	0	0	1	197,7	$y_3 = 5,29$
4	1/2	1/2	0	180,7	$y_{12} = 5,20$
5	1/2	0	1/2	134,7	$y_{13} = 4,90$
6	0	1/2	1/2	161,3	$y_{23} = 5,08$

реализован симплекс-решетчатый план второго порядка (рис. III.29, б). Для расчета оценок коэффициентов использованы формулы из прил. II.14. Так, коэффициент при $v_1 v_3$ в (III.200) $A_{13} = 4 \cdot 4,9 - 2(6,02 + 5,29) = 3,01$.

Регрессионный анализ моделей в виде приведенных полиномов выполняется по общей методике (см. п. III.8 и III.9). Однако в связи с насыщенностью планов Шеффе проверка точности описания объекта моделью проводится в $N_{\text{доп}}$ дополнительных точках [1, 47]. Полученные при этом значения отклика $\bar{y}_{s, \text{доп}}$ сравниваются с расчетными $\hat{y}_{\text{доп}}$ по t -критерию, строящемуся с учетом функции дисперсии предсказания d (III.146), которая для моделей второго порядка имеет вид [1, 47]

$$d = \sum_{i=1}^q [v_i (2v_i - 1)]^2 + 16 \sum_{i < j} v_i^2 v_j^2. \quad (\text{III.203})$$

Если число параллельных измерений $m_{\text{доп}} = m_u = \text{const}$, то нуль-гипотеза об отсутствии расхождений между \bar{y}_s и \hat{y} принимается, когда t -критерий, аналогичный по смыслу критерию (III.37), окажется меньше критического уровня:

$$t_{\text{доп}} = \frac{|\bar{y}_s - \hat{y}|_{\text{доп}}}{s_s \sqrt{1+d}} \sqrt{m_{\text{доп}}} < t_{\text{кр}}. \quad (\text{III.204})$$

Критический уровень $t_{\text{кр}}$ берется (прил. II.5) при числе степеней свободы f_s и уровне значимости $\alpha/N_{\text{доп}}$ (т. е. по мере увеличения числа дополнительных точек значение $t_{\text{кр}}$ возрастает, а значит, фактические разности $|\bar{y}_s - \hat{y}|_{\text{доп}}$ могут быть больше). Если условия однородности s_s^2 или $m_{\text{доп}} = m_u = \text{const}$ не выполняются, то t -критерий строится аналогично (III.72).

Пример III.43. В дополнение к эксперименту в пр. III.42 поставлен опыт в точке M_1 с координатами $v_1 = 0,35$, $v_2 = 0,09$ и $v_3 = 0,56$; число повторений $m_{\text{доп}} = m_u = 2$; $s_s \{\ln \eta\} = 0,05$ при $f_s = \infty$; результат $\bar{\eta}_{\text{доп}} = 119$ Па·с ($\ln \eta = 4,779$). В соответствии с (III.200) расчетное значение $\widehat{\ln \eta} \{M_1\} = 4,853$.

Функция дисперсии предсказания в точке M_1 имеет значение

$$d\{M_1\} = [0,35(2 \cdot 0,35 - 1)]^2 + [0,09(2 \cdot 0,09 - 1)]^2 + [0,56(2 \cdot 0,56 - 1)]^2 + \\ + 16(0,35^2 \cdot 0,09^2 + 0,35^2 \cdot 0,56^2 + 0,09^2 \cdot 0,56^2) = 0,6922.$$

Рассчитанный по (III.204) критерий

$$t_{\text{доп}} = \frac{|4,779 - 4,853|}{0,05 \sqrt{1 + 0,6922}} = 1,61$$

меньше чем $t_{\text{кр}} = 1,645$ ($\alpha = 0,1/1$, $f = \infty$ — см. прил. II.5), поэтому точность совпадения результатов дополнительного эксперимента и расчетов по (III.200) в точке M_1 (где вязкость минимальна) можно признать удовлетворительной.

Методы планирования и анализа эксперимента при изучении систем «смесь — свойства», включая приемы планирования в ограниченной части симплекса и отличные от (III.199), (III.201), (III.202) виды моделей, достаточно подробно рассмотрены в специальной литературе [47 и др].

Совместное воздействие на технические свойства материалов *смесевых* (взаимозависимых) факторов v и *взаимонезависимых* факторов $|x| < 1$ (параметры технологии и эксплуатации, относительное к количеству базового компонента содержание смеси и т. п.) можно оценить, анализируя изменения диаграмм «состав — свойство» под влиянием x_i . Передвигая диаграмму « v — Y » вдоль оси x_i в факторном пространстве и фиксируя ее в отдельных точках этой оси (в частности, при $x_1 = -1; 0; +1$), можно проанализировать скачкообразные изменения изолиний $Y(v) = \text{const}$ под влиянием этого фактора. При непрерывном перемещении диаграммы вдоль оси x_1 формируется треугольная *призма*, в которой изоповерхности свойства $Y(v_1, v_2, v_3, x_1) = C_{\text{тр}} = \text{const}$ образованы соответствующими изолиниями $Y(v_1, v_2, v_3) = C_{\text{тр}} = \text{const}$ [8, 41].

Пример III.44. Зависящая от зернового состава наполнителя вязкость композиций, рассматриваемых в пр. III.40 — III.43, определяется прежде всего степенью наполнения. При изменении нормализованной концентрации наполнителя x в диапазоне $1,5 \leq H/\Pi \leq 2,5$ трансформируется и диаграмма «зерновой состав — вязкость». Эти изменения отражает набор диаграмм « $v = \eta$ » (рис. III.30) для ряда фиксированных уровней $H/\Pi = 1,5; 2; 2,5$ ($x = -1; 0; 1$). Зависимость $\eta(v_1, v_2, v_3, x)$ отображается изоповерхностями вязкости внутри призмы «зерновой состав — наполнение» на рис. III.30, г. В основаниях расположены диаграммы «зерновой состав — вязкость» для граничных уровней $H/\Pi = 1,5$ и $H/\Pi = 2,5$ (совпадает с рис. III.28, б). Изолинии на гранях призмы отражают изменения вязкости композиций с тремя бинарными наполнителями — $\{v_1, v_2\}$, $\{v_1, v_3\}$ и $\{v_2, v_3\}$ при изменении H/Π .

Штриховая линия внутри призмы — геометрическое место точек минимальной вязкости на треугольных диаграммах (точек типа M_1 на рис. III.28). На объемной диаграмме видно, что изоповерхности выгнуты в сторону больших H/Π . Следовательно, существует возможность, подбирая зерновой состав наполнителя, уменьшить расход полимера в композиции при сохранении ее заданной вязкости. Так, для обеспечения $\eta_{\text{тр}} = 120$ Па · с наполнитель из смеси мелких и крупных частиц (состав на штриховой линии) требует расхода 310 кг смолы на 1 т материала, а при мелком наполнителе — 350 кг.

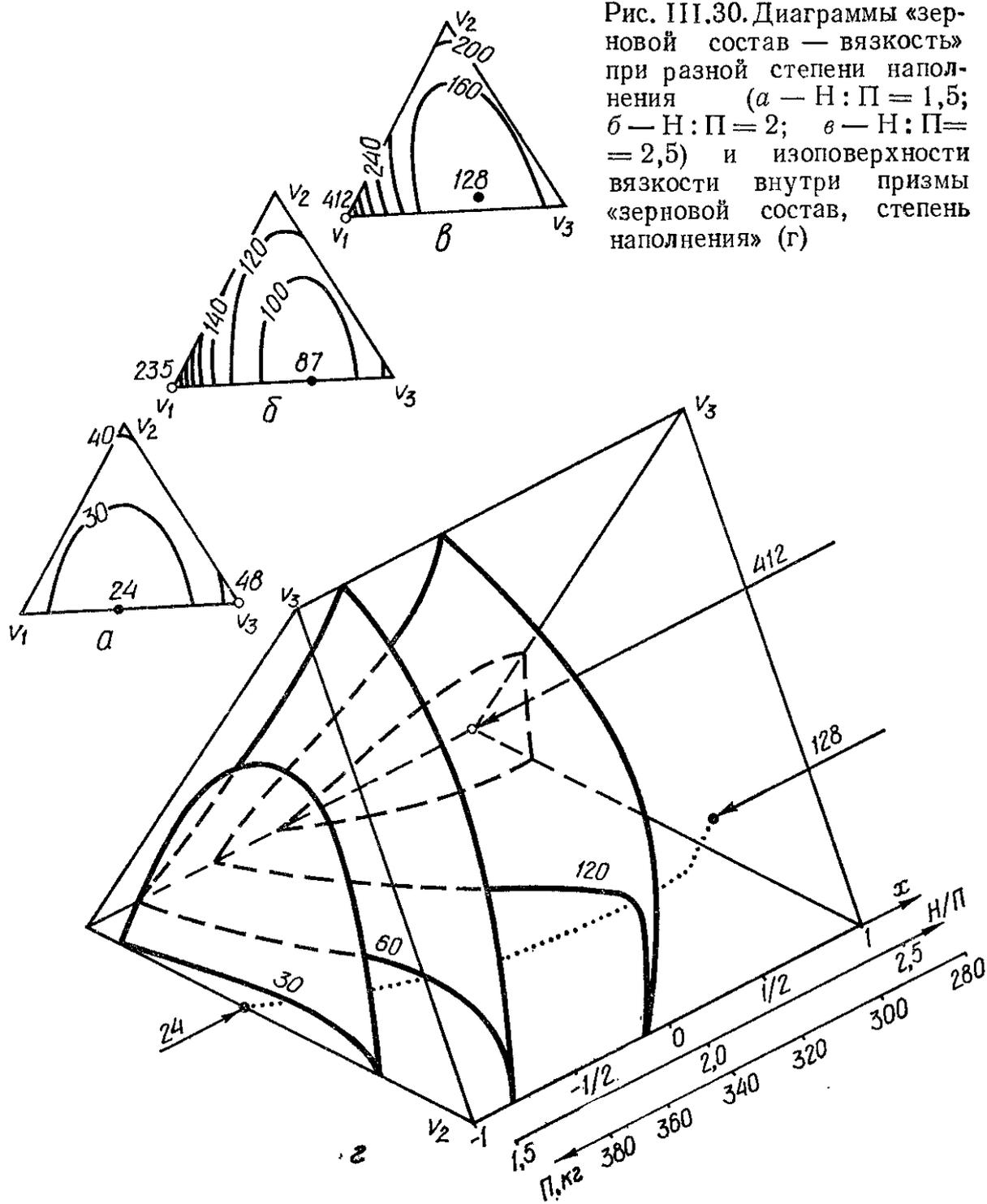


Рис. III.30. Диаграммы «зерновой состав — вязкость» при разной степени наполнения (a — $H:\Pi = 1,5$; b — $H:\Pi = 2$; e — $H:\Pi = 2,5$) и изоповерхности вязкости внутри призмы «зерновой состав, степень наполнения» (γ)

Отобразить графически изменение диаграммы «состав — свойство» под влиянием двух технологических факторов — x_1 и x_2 можно только дискретно [8, 41]: строятся и анализируются смесевые диаграммы $f(v)$ в фиксированных точках квадрата $\{x_1, x_2\}$, в частности в девяти точках-центроидах. Возможно другое отображение — в семи точках-центроидах треугольника (вершины, середины сторон и центр тяжести) строятся изменяющиеся от точки к точке квадратные диаграммы $Y(x_1, x_2)$.

Системы, свойства которых определяются группой смесевых факторов $v = (v_1, \dots, v_q)^T$ (III.192) и группой технологических факторов $x = (x_1, \dots, x_k)^T$, названы [8, 13, 41] системами «смесь, технология — свойства» и обозначены MTQ (T от англ. *technology* — тех-

нология). Полный вектор $K = (q + k)$ факторов в такой системе записывается как

$$u = (v^T, x^T)^T; \quad (\text{III.205})$$

факторная область — выпуклый многогранник, представляющий собой декартово произведение $(q - 1)$ -мерного симплекса на k -мерный куб:

$$\Omega_u = \Omega_v \times \Omega_x. \quad (\text{III.206})$$

Запись (III.206) означает, что каждой точке симплекса Ω_v соответствует гиперкуб Ω_x (и наоборот).

Для анализа и оптимизации систем MTQ полный полином m -й степени от u (III.205) не может быть использован в качестве модели в силу линейной взаимосвязи части факторов (что уже обсуждалось при рассмотрении систем MQ). Поэтому для описания систем MTQ используются: во-первых, *модели-произведения* — произведения приведенного полинома от v на полином от x ; во-вторых, *приведенные полиномы* от u . Расчет оценок коэффициентов и регрессионный анализ моделей MTQ проводятся по общим правилам, рассмотренным в п. III.9.

Если функция отклика в системе MTQ задается произведением приведенного полинома от v и полинома от x , то ее удобно представить в форме полинома от факторов одной группы, в частности v_i , коэффициенты которого — функции факторов другой группы, в частности x : $A_i = \varphi_i(x)$ и $A_{ij} = \varphi_{ij}(x)$. Таким образом, каждый коэффициент полинома (III.199) из числовой оценки превратился в функцию от независимых факторов. В частности, для случая $q = 3$ и $k = 1$ произведение полиномов второго порядка дает модель

$$\begin{aligned} Y = & (b_{0.A_1} + b_{1.A_1}x + b_{11.A_1}x^2)v_1 + \\ & + (b_{0.A_2} + b_{1.A_2}x + b_{11.A_2}x^2)v_2 + \\ & \dots \quad \dots \quad \dots \\ & + (b_{0.A_{23}} + b_{1.A_{23}}x + b_{11.A_{23}}x^2)v_2v_3. \end{aligned} \quad (\text{III.207})$$

Функции в скобках характеризуют изменение свойств чистых компонентов и их смесей в зависимости от технологического фактора.

Пример III.45. Зависимость вязкости полиэфирных композиций от зернового состава и содержания наполнителя, показанная на объемной диаграмме рис. III.30, *г* (пр. III.44), описывается моделью

$$\begin{aligned} \widehat{Y} = \ln \widehat{\eta} = & (5,46 + 1,23x - 0,66x^2)v_1 + \\ & + (5,01 + 0,83x - 0,44x^2)v_2 + \\ & + (4,83 + 0,71x - 0,24x^2)v_3 + \\ & + (-1,43 - 0,63x \pm 0)v_1v_2 + \\ & + (-2,58 - 0,42x + 0)v_1v_3 + \\ & + (-1,04 \pm 0 \pm 0)v_2v_3. \end{aligned} \quad (\text{III.208})$$

При подстановке в (III.208) $x = +1$ получается модель (III.200), описывающая диаграмму «зерновой состав — вязкость» при $H/\Pi = 2,5$ (рис. III.28, *б* и III.30, *в*). Для заданного состава наполнителя можно из (III.208)

получить модель $\eta = \psi(N/P)$; так, в частности, для смеси мелкой и крупной фракций $v_1 = v_3 = 0,5$ и $v_2 = 0$ будет получена парабола $\hat{Y} = \ln \hat{\eta} = 5,14 + 0,97x - 0,45x^2$.

Число членов в моделях-произведениях, а следовательно, и минимальное количество разных экспериментальных точек для оценки коэффициентов определяется как $L_{M-P} = N_{M-P} = C_{q+m-1}^m C_{k+m}^m$.

При $q = 3$, $k = 1$ и $m = 2$, т. е. в модели (III.208), оно равно 18. С ростом числа факторов или степени полиномов-сомножителей в моделях-произведениях растет число эффектов высокого порядка, которые во многих практических задачах несут незначительный вклад. Поэтому значительные экспериментальные затраты на N_{M-P} опытов оказываются неоправданными. Для уменьшения показателя избыточности $e_N = N/L$ в качестве моделей *MTQ* используются специальные приведенные полиномы [8].

Модели в виде приведенных полиномов *MTQ* степени m с числом факторов $K = q + k$ содержат лишь $L = C_{k+m-1}^m$ членов. Так, при $q = 3$, $k = 1$ и $m = 2$ число эффектов $L = 10$, т. е. по сравнению с (III.208) можно уменьшить объем опытов почти вдвое. Однако эти модели будут правдоподобны лишь в том случае, если априори известно о независимости эффектов синергизма и антагонизма в смесевой системе от уровня технологических факторов. Если такая информация отсутствует и нет жестких ограничений на число точек в эксперименте, то для описания системы «смесь, технология — свойства» лучше использовать модели в виде произведения приведенного полинома относительно смесевых факторов на обычный полином от взаимонезависимых (технологических) факторов, т. е. модели типа (III.207).

Для оценки коэффициентов в произведениях полиномов в качестве планов эксперимента используют *произведения соответствующих планов*. Это означает, что в каждой точке плана на симплексе реализуется план на кубе (и наоборот). При этом если оптимальны по какому-либо критерию планы-сомножители, то оптимально и их произведение.

Пример III.46. Для описания зависимости вязкости композиций от зернового состава $v = (v_1, v_2, v_3)$ и количества x наполнителя был выбран полином вида (III.207). Эксперимент для оценки его 18 коэффициентов выполнен по оптимальному насыщенному плану: в каждой из трех точек оптимального плана на отрезке ($x = -1; 0; 1$) реализован симплекс-решетчатый план второго порядка (см. рис. III.29, б). Определялась вязкость 18 разных композиций: при трех уровнях наполнения ($N/P = 1,5; 2; 2,5$) испытывались композиции с мононаполнителями и равновесными бинарными смесями. Расчет оценок коэффициентов по МНК выполнен на ЭВМ. После удаления незначимых эффектов при ПРА (см. п. III.9) получена модель (III.208) с 14 коэффициентами, изоповерхности которой показаны на рис. III.30, а.

Оптимальные планы для построения моделей *MTQ* в виде приведенных полиномов синтезируются на ЭВМ (см. п. III.14). Известен [6, 8, 41] набор синтезированных *D*-оптимальных планов с разным числом точек для приведенных полиномов до третьей сте-

пени относительно взаимозависимых факторов v и не выше второй степени относительно взаимонезависимых x .

Представление объектов в материаловедении и технологии как систем MTQ позволяет [8] глубже и полнее анализировать происходящие в них явления. Однако проблему методического и программного обеспечения такого подхода нельзя признать завершённой

III.14. Синтез оптимальных планов эксперимента

Выбор того или иного плана эксперимента (III.164) для решения конкретной задачи регрессионного анализа (т. е. задачи оценки параметров Θ ЭС-моделей системы, оценки выходов системы и других функций параметров Θ) диктуется конкретными условиями этой задачи и требованиями к результатам ее решения.

В общем виде условия типовой задачи регрессионного анализа записываются как

$$Y = F(x, \Theta) + \varepsilon = \Theta^T f(x) + \varepsilon, \quad (\text{III.209})$$

где Y — отклик; $x = (x_1, \dots, x_k)^T$ — факторы, изменяемые в области Ω_x пространства R^k (x обозначает здесь и взаимонезависимые x , и смесевые v , и любые другие нормированные управляемые входы); $F(x, \Theta)$ — функция отклика; $f(x) = (f_1(x), \dots, f_l(x), \dots, f_L(x))^T$ — базисные функции; независимые случайные ошибки ε распределены по нормальному закону с математическим ожиданием $E(\varepsilon) = 0$ и дисперсией $E(\varepsilon^2) = \sigma_\varepsilon^2(x)$.

Решение задачи определяется по результатам N_m наблюдений за величиной Y в точках плана (III.164) в виде оценок МНК. Эти оценки Θ , по аналогии с (III.114), в общем случае (III.164) и (III.209) определяются через информационную матрицу M и вектор Y_M :

$$M = \sum_{1 \leq u \leq N_m} f(x_u) f^T(x_u); \quad (\text{III.210})$$

$$Y_M = \sum_{1 \leq u \leq N_m} y_u f(x_u); \quad (\text{III.211})$$

$$\hat{\Theta} = M^{-1} Y_M = D Y_M. \quad (\text{III.212})$$

Для функции отклика МНК-оценка

$$\hat{Y} = \hat{\Theta}^T f(x). \quad (\text{III.213})$$

При постановке задачи условия (III.209) конкретизируются заданием факторной области Ω_x , вектора базисных функций и функции ошибки эксперимента. В большинстве типовых задач ошибка эксперимента однородна в области Ω_x , т. е. $\sigma_\varepsilon^2(x) = \text{const} = s_\varepsilon^2$.

Конкретизация плана эксперимента (III.164) на основании этих данных, как подчеркнуто в п. III.10, может быть осуществлена исследователем в соответствии с требованиями точности оценок $\hat{\Theta}$, $\hat{Y}(x)$ и другими, связанными со стоимостью и условиями проведения

эксперимента ограничениями и пожеланиями. Точность оценок (III.212) описывается, как и (III.131), ковариационной матрицей

$$E \{(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T\} = s_0^2 D, \quad (\text{III.214})$$

а дисперсия оценки выхода (III.213)

$$s^2 \{\hat{Y}\} = s_0^2 d(\mathbf{x}) = s_0^2 \mathbf{f}^T(\mathbf{x}) D \mathbf{f}(\mathbf{x}). \quad (\text{III.215})$$

Связанные с матрицей D (см. III.212) и функцией d (см. III.215) критерии оптимальности планов, рассмотренные в п. III.10, используются в качестве меры для выражения требований к точности оценок.

Для многих типовых, часто встречающихся в исследованиях и инженерной практике ситуаций разработано и накоплено достаточно большое число оптимальных по разным критериям планов эксперимента с разным числом точек N_m . Созданы каталоги планов [58] и др. (см. [1, 4, 6, 8]). Аналогичные и более мощные каталоги (банки планов) могут быть организованы и эксплуатироваться с помощью ЭВМ. Однако уже существующие и вновь разрабатываемые планы не могут охватить многообразия реальных задач экспериментально-статистического моделирования. Исследователю приходится либо *подгонять условия задач* (III.209) под *готовый план* (изменять факторную область, вид или форму модели, увеличивать число опытов, сознательно снижая требования к решению задачи), либо ставить и решать задачу *синтеза оптимального плана*, соответствующего *конкретной* цели моделирования.

Следует выделить три основные группы ситуаций, когда целесообразно синтезировать специальный план.

1. *Не стандартна* (отлична от куба, шара, симплекса) *факторная область*. Так, возможность варьирования факторов в ограниченной части куба (см. рис. III.1, в) выражает запись области действия

$$\Omega_x = \left\{ \begin{array}{l} |x_i| \leq 1 \quad (i = \overline{1, k}); \\ g_j(\mathbf{x}) \leq 0 \quad (j = \overline{1, l}), \end{array} \right\} \quad (\text{III.216})$$

где помимо прямых ограничений на факторы x_i указаны функциональные ограничения — функции $g(\mathbf{x})$, отсекающие от куба *запрещенные* (опасные, бессмысленные или нежелательные для экспериментов) *зоны*.

Пример III.47. Изучается возможность сокращения времени τ набора начальной прочности R полимерной композиции за счет использования в качестве катализатора твердения смеси двух добавок. Содержание добавки-модификатора структуры X_1 и добавки-ускорителя твердения X_2 варьируется в диапазонах: $0 \leq X_1 \leq 2$, $0 \leq X_2 \leq 2$ (% к массе смолы). При этом по априорным данным известно, что если количество катализатора превышает 3 %, то технологическая смесь теряет требуемую жизнеспособность, т. е. $X_1 + X_2 \leq 3$ или для нормализованных факторов $x_1 + x_2 \leq 1$. Это условие не позволяет для построения модели зависимости $\tau(x_1, x_2)$ в виде полинома второго порядка выполнить эксперимент по насыщенному плану второго порядка на квадрате (в частности, по плану № 5, описанному в п. III.12; рис.

III.24). Поэтому необходимо определить координаты точек насыщенной области для оценки шести коэффициентов модели стандартного вида (III.11), но в факторной области:

$$\Omega_x = \left\{ \begin{array}{l} -1 \leq x_1 \leq +1; \\ -1 \leq x_2 \leq +1; \\ x_1 + x_2 - 1 \leq 0. \end{array} \right\} \quad (\text{III.217})$$

2. *Не стандартен вид модели* (для оценки коэффициентов стандартной модели известны приемлемые планы эксперимента). Так, из априорных данных может быть известно, что функция отклика содержит один из эффектов высокого порядка, а необходимость экономить опыты делает нецелесообразным реализацию эксперимента по известному плану для модели с большим числом эффектов.

Пример III.48. Анализируется влияние концентрации сериной кислоты X_1 в диапазоне $1 \leq X_1 \leq 10$ % и температуры X_2 в интервале $20^\circ\text{C} \leq X_2 \leq 60^\circ\text{C}$ на вероятность разрушения полимерного материала [4] при растяжении до нормативного напряжения 7 МПа. Факторы нормализуются по соотношениям $x_1 = (\ln X_1 - 0,5) / 0,5$ и $x_2 = (X_2 - 40) / 20$. Критерием оптимизации является предложенная в [4] характеристика κ , при использовании которой расчетное значение частоты появления исследуемого события не может выйти за пределы $0 \leq \hat{p} \leq 1$:

$$\kappa = \ln[p/(1 - p)]. \quad (\text{III.218})$$

Экспериментальное значение p_u определяется при испытании в каждой точке плана $m_u = 50$ идентичных образцов.

По результатам предшествующих экспериментов известно, что эффекты нулевого, первого и второго порядков влияния концентрации кислоты на деформацию материала пропорциональны изменению температуры среды. Поэтому модель зависимости $\kappa(x_1, x_2)$ записывается в виде

$$\begin{aligned} \hat{\kappa} = & b_0 + b_1x_1 + b_{11}x_1^2 + \\ & + b_2x_2 + b_{12}x_1x_2 + b_{112}x_1^2x_2. \end{aligned} \quad (\text{III.219})$$

Из-за трудоемкости испытаний 50 образцов в каждой точке плана для оценки шести коэффициентов модели (III.219) известные планы третьего порядка на квадрате [58] не устраивают исследователя, поскольку содержат «лишние» точки. Необходимо синтезировать для модели (III.219) насыщенный оптимальный план на квадрате $|x_i| \leq 1$.

3. Модель и область действия стандартны (например, полином второго порядка на кубе), однако исследователю необходимо *включить* в план эксперимента *особые точки* (эталонные или априори наиболее экономичные решения, составы, процессы и т. п.) факторной области Ω_x , причем точки эти не содержатся ни в одном из известных в каталогах планов. В таких условиях необходимо *достроить оптимальный план* к этим особым *обязательным* точкам.

Пример III.49. Изучается возможность повышения производительности установки для подачи в стык строительных конструкций герметика за счет использования ультразвуковых насадок на сопло [6]. В частности, анализируется прирост производительности ΔQ , г/мин, в зависимости от скорости прохождения герметика в насадке ($X_1 = (1,9 \pm 1,2)\text{c}^{-1}$) и от толщины обрабатываемого ультразвуком слоя ($X_2 = (4 \pm 2)$ мм). У существующей установки $x_1 = 0,4$ и $x_2 = 0,6$; естественно, что необходимо включить точку

с такими координатами в эксперимент на квадрате $\{x_1, x_2\}$, минимизируя число опытов, связанных с переналадкой установки. Вместо шести точек известного насыщенного плана второго порядка (см. рис. III.24) в эксперимент следует ввести «особую» точку плюс пять точек, координаты которых найдены на ЭВМ.

Задача синтеза оптимального плана E^* формулируется как экстремальная задача

$$E^* = \arg \min_{E \in \Omega_E} \psi(D(E)), \quad (\text{III.220})$$

в которой целевой функцией является один из критериев оптимальности планов, связанный с ковариационной матрицей оценок коэффициентов модели (см. п. III.10), а следовательно (см. (III.212)), с информационной матрицей M плана E (III.164), входящего в Ω_E — множество возможных планов, содержащих N_m точек из области Ω_x .

Решить задачу (III.220) означает найти координаты таких N_m точек факторной области, чтобы составленный из них план E^* был лучшим по критерию ψ из всех возможных планов, обеспечивая минимум ψ для заданных условий (III.209). Оптимизационная задача (III.220) чрезвычайно сложна: целевая функция — функционалы $\psi(D(E))$ — многоэкстремальна; велика размерность области решений, в которой ведется поиск оптимального плана E^* , — она равна $(k \times N_m)$ и т. д. Поэтому приближенным решением задач синтеза оптимальных планов занимаются специалисты в области математической теории эксперимента и численных методов.

Инженер-технолог (материаловед) оценивает необходимость синтезировать план эксперимента для своей нестандартной задачи моделирования и, обращаясь к специалисту, представляет исходные данные к синтезу плана: критерий оптимальности, факторную область, вид модели, число точек плана.

Алгоритм синтеза D -оптимальных планов для разных видов моделей и факторных областей в исследованиях композиционных материалов описан в [8]. По этому алгоритму на ЭВМ решены, в частности, задачи синтеза планов, сформулированные в пр. III.47 — III.49.

Пример III.50. Точки оптимального плана на «усеченном» квадрате для решения рецептурной задачи, сформулированной в пр. III.47, показаны на рис. III.31, а. Спектр исходного плана включал три точки в углах квадрата (№ 1 — 3), что априори соответствует «хорошему» расположению опытов, и три случайные точки на поле квадрата. В ходе оптимизации именно последние точки передвинулись: точки 4 и 5 расположились на линии усечения симметрично, на максимальном удалении от центра; точка 6 имеет координаты $\{-0,2; -0,2\}$. Этот синтезированный на ЭВМ план близок к D -оптимальному насыщенному плану (№ 6 на рис. III.24), известному из каталога [58], хотя в силу усечения квадрата точки № 4 — 5 несколько сдвинуты. По плану получена модель условной вязкости τ, c , изолинии которой показаны на рис. III.31, а. Среднеквадратичная ошибка предсказания $s\{\hat{Y}\}/s_y$ для синтезированного плана (рис. III.31, б) практически не отличается от аналогичной величины для D -оптимального плана (см. рис. III.24, г) из каталога, хотя значения минимума смещены в другие зоны.

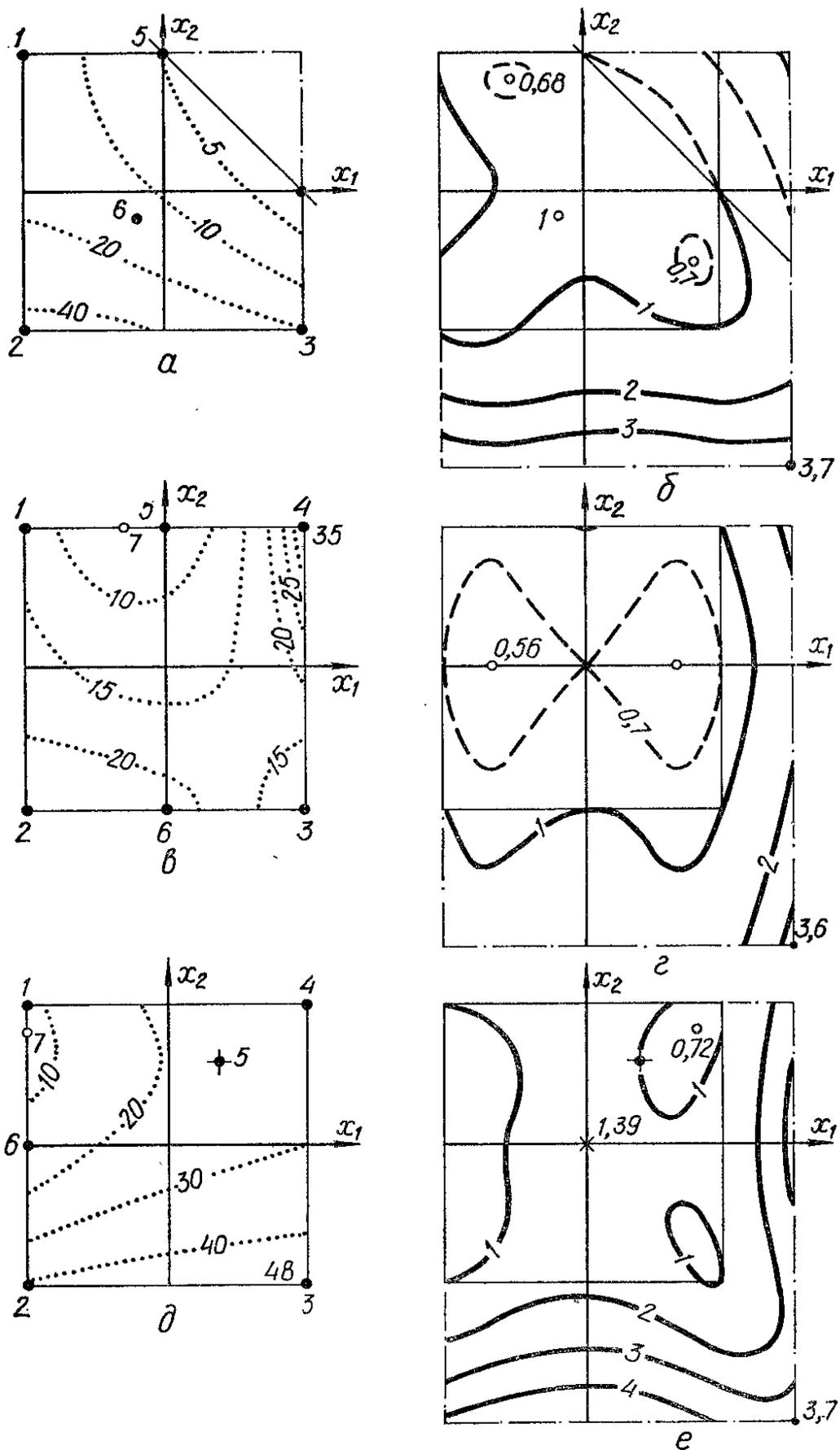


Рис. III. 31. Синтезированные на ЭВМ оптимальные планы (а, в, д) и изолинии ошибок предсказания (б, г, е) для случая усеченного квадрата (а, б), неполной кубической модели (в, г) и фиксированной точки № 5 (д, е)

Пример III.51. Для построения ипополиой кубической модели, гипотеза о которой сформулирована в пр. III.48, синтезирован D -оптимальный план, показанный на рис. III.31, в. По эксперименту в этих точках получена модель

$$\hat{y} = -1,88 + 0,12x_1 + 0,42x_1^2 - 0,62x_2 + 0,55x_1x_2 + 0,80x_1^2x_2,$$

для которой изолинии оценки вероятности p разрушения образца (при нагрузке ниже нормативной) приведены на рис. III.31, в. Как показывает анализ изолиний ошибки предсказания $s\{\hat{Y}\}/s_y$ (рис. III.31, г), синтезированный симметричный план обладает хорошей точностью (см. рис. III.24, а, г).

Пример III.52. Для технологической ситуации с фиксированной в факторном пространстве точкой (№ 5, $x_1 = 0,4$, $x_2 = 0,6$) D -оптимальный план приведен на рис. III.31, д; там же показаны изолинии прироста производительности агрегата ΔQ , %. В этой задаче особый интерес представляет сравнительный анализ изолиний ошибки предсказания $s\{\hat{Y}\}/s_y$, приведенных на рис. III.31, е, и изолиний стандартных планов на рис. III.24. В синтезированном плане за счет фиксации точки № 5 точность предсказания, конечно, хуже, чем у стандартного D -оптимального насыщенного плана (рис. III.24, г), но несколько лучше, чем у композиционного к ПФЭ 2^2 насыщенного плана (рис. III.24, в).

В перспективе роль специалистов по синтезу планов смогут выполнять в ходе диалога исследователя с ЭВМ (см. п. IV.5) экспертные системы, оснащенные программами синтеза планов.

III.15. Решение с помощью полиномиальных моделей задач оптимизации первого вида — поиска экстремумов выхода

Решение задачи оптимизации первого вида (см. «Введение») сводится к поиску в пределах некоторой области Ω экстремальных значений выхода системы \hat{Y}^* — максимума \hat{Y}_{\max} с координатами x_{\max} и (или) минимума \hat{Y}_{\min} с координатами x_{\min} . Область Ω всегда ограничена, причем она может полностью совпадать с областью Ω_x , в которой ставился эксперимент и в которой $|x_i| \leq 1$, или отличаться от области Ω_x , как правило, пересекаясь с ней в какой-то подобласти. К решению задачи могут быть привлечены любые методы оптимизации, рассмотренные в п. II.1 — II.4. Однако для поиска \hat{Y}^* в тех случаях, когда поведение системы в Ω_x адекватно описывается ЭС-моделями в виде полиномов не выше второго порядка, целесообразно применять специфический диссоциативно-шаговый метод оптимизации (ДШМ), при создании которого [4, 6, 23] учтены особенности структуры таких моделей. Этот метод, с одной стороны, наиболее узкий, поскольку не применяется к моделям иной структуры, с другой — он прост и наиболее информативен, так как по своей сути ориентирован на полное использование информации, содержащейся именно в ЭС-моделях данной структуры.

Полиномиальная k -факторная модель второго порядка в любом однофакторном сечении (это значит, что фиксированы уровни всех x_i , кроме одного, например x_1 без потери общности) переходит в

параболу $\hat{Y} = b_0 + b_1 x_1 + b_{11} x_1^2$, свойства которой проанализированы в п. III.2 (см. рис. III.2). Представление о параболе, как единственно возможной форме (прямая — вырожденная парабола при $b_{11} = 0$) связи между \hat{Y} и x_1 , развивается в понятии о *квазиоднофакторной модели* W_i , в которую входят все эффекты, связанные с данным фактором x_i , в частности $i = 1$:

$$W_i = \hat{Y} - b_{0,j} = b_i x_i + \sum_{i < j} b_{ij} x_i x_j + b_{ii} x_i^2; \quad (\text{III.221})$$

$$W_1 = \hat{Y} - b_{0,j} = (b_1 + \sum_{j=2}^k b_{1j} x_j) x_1 + b_{11} x_1^2. \quad (\text{III.222})$$

Свободный член в квазиоднофакторной модели всегда равен нулю, поскольку элемент $b_{0,j}$ отражает влияние на \hat{Y} размещения других факторов x_j на некоторых определенных уровнях.

Соотношение (III.222) описывает семейство парабол, проходящих через начало координат $\{x_i = 0, W_i = 0\}$. Поскольку квадратичный эффект b_{ii} в квазиоднофакторной модели имеет постоянное значение, то форма параболы (ширина ее раскрытия) не меняется. Изменение линейного эффекта b_i под влиянием других факторов через эффекты взаимодействия b_{ij} приводит к тому, что вершины $\{x_i^*, W_i^*\}$ множества парабол, образующих семейство (III.221), передвигаются вдоль оси x_i . Исходя из общего условия стационарности точки экстремума (II.8), можно записать

$$\frac{\partial W_i}{\partial x_i} = \left(b_i + \sum_j b_{ij} x_j \right) + 2b_{ii} x_i = 0 \quad (\text{III.223})$$

и получить функцию, по которой происходит перемещение точки x_i^* при изменении уровней других x_j :

$$x_i^* = (b_i + \sum_j b_{ij} x_j) / (-2b_{ii}). \quad (\text{III.224})$$

При этом вторая координата вершин W_i^* перемещается по функции, которая получается при подстановке (III.224) в (III.222):

$$W_i^* = (b_i + \sum_j b_{ij} x_j)^2 / (-4b_{ii}). \quad (\text{III.225})$$

Поскольку функция (III.225) одновременно указывает, насколько изменится исходная модель (или другие квазиоднофакторные модели) после перехода x_i на уровни, обеспечивающие экстремальное поведение системы, то по ней можно рассчитать *приросты* всех оставшихся в исходной модели *оценок* (не содержащих индекса «i»):

$$\left. \begin{aligned} \Delta b_0 &= -0,25 b_i^2 / b_{ii}; \\ \Delta b_j &= -0,5 b_i b_{ij} / b_{ii}; \\ \Delta b_{jj} &= -0,25 b_{ij}^2 / b_{ii}; \\ \Delta b_{es} &= -0,5 b_{ie} b_{is} / b_{ii} \quad (e \neq s \neq i). \end{aligned} \right\} \quad (\text{III.226})$$

Поскольку каждый фактор x_j в области Ω_x изменяется только в диапазоне $-1 \leq x_j \leq +1$, причем влияет на координату x_i^* (III.224) вершины параболы *линейно*, то можно найти *граничные положения* как этой вершины x_i^* , так и самой квазиоднофакторной параболы W_i , подставив соответственно в (III.224) и (III.221) значения $x_j = \pm 1$ с тем знаком, который максимизирует $\sum_j b_{ij}$ в зависимости от знака при эффектах взаимодействия:

$$x_i^* = (b_i \pm \sum_j b_{ij}) / (-2 b_{ii}); \quad (\text{III.227})$$

$$W_i = (b_i \pm \sum_j b_{ij}) x_i + b_{ii} x_i^2. \quad (\text{III.228})$$

Из (III.227) следует, что x_i^* *всегда находится* в области эксперимента Ω_x , если выполняется условие

$$|b_i| + \sum_j |b_{ij}| \leq 2 |b_{ii}|, \quad (\text{III.229})$$

и никогда *не входит* в эту область, если при нарушении (III.229) дополнительно существует соотношение

$$|b_i| - \sum_j |b_{ij}| \geq 2 |b_{ii}|. \quad (\text{III.230})$$

Пример III.53. В пр. III.39 при исследовании влияния на прочность бетона R , МПа, соотношения П/Ц (x_1), концентрации суперпластификатора $C_{\text{сп}}$ (x_2) и температуры T (x_3) ТВО получена модель (III.191), из которой по (III.222) следуют три квазиоднофакторные модели:

$$W_1 = (5,5 + 2x_2 + 1,5x_3)x_1 + 1,9x_1^2; \quad (\text{III.231})$$

$$W_2 = (0,8 + 2x_1 + 0,4x_3)x_2 - 1,8x_2^2; \quad (\text{III.232})$$

$$W_3 = (6,8 + 1,5x_1 + 0,4x_2)x_3 + 0,6x_3^2; \quad (\text{III.233})$$

описывающие семейства парабол (рис. III.32) с граничными положениями:

$$\left. \begin{aligned} W_1 &= (5,5 \pm 3,5)x_1 + 1,9x_1^2; & 2 \leq b_1 \leq 9; \\ W_2 &= (0,8 \pm 2,4)x_2 - 1,8x_2^2; & -1,6 \leq b_2 \leq 3,2; \\ W_3 &= (6,8 \pm 1,9)x_3 + 0,6x_3^2; & 4,9 \leq b_3 \leq 8,7. \end{aligned} \right\} \quad (\text{III.234})$$

Координаты экстремума (минимума для W_1 и W_3 , но максимума для W_2 — рис. III.32) передвигаются (III.224) вдоль соответствующей оси x_i по линейным функциям:

$$\begin{aligned} x_{1(\min)} &= (5,5 + 2x_2 + 1,5x_3) / (-2 \cdot 1,9) = \\ &= -1,45 - 0,53x_2 - 0,39x_3; \quad -2,37 \leq x_1^* \leq -0,53; \end{aligned} \quad (\text{III.235})$$

$$x_{2(\max)} = +0,22 + 0,56x_1 + 0,11x_3; \quad -0,44 \leq x_2^* \leq +0,89; \quad (\text{III.236})$$

$$x_{3(\min)} = -5,66 + 1,25x_1 - 0,33x_2; \quad -7,25 \leq x_3^* \leq -4,08. \quad (\text{III.237})$$

Поскольку для квазиоднофакторной модели W_3 (III.234) не выполняется условие (III.229), а именно: $|6,8| + |1,9| = |8,7| > |2 \cdot 0,6| = |1,2|$, и, кроме того, действительно соотношение (III.230), а именно $|6,8| - |1,9| = |4,9| > |1,2|$, то вершины парабол всегда находятся вне области эксперимента, что и отражено в (III.237). Противоположен результат анализа для

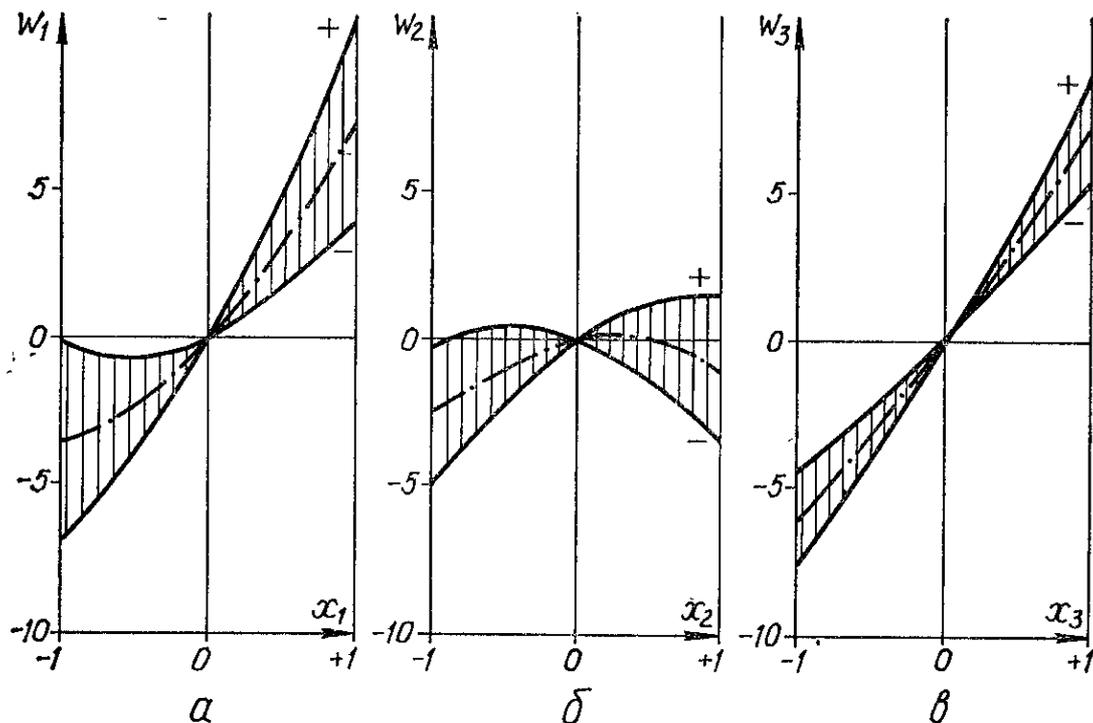


Рис. III. 32. Три квазиоднофакторных модели влияния на прочность бетона R («+» — верхняя граница, «-» — нижняя):
 a — П/Ц; b — суперпластификатора; c — температуры ТВО

парабол W_2 , которые всегда находятся в области Ω_x , поскольку для них выполняется условие (III.229). В отношении парабол W_1 однозначный вывод невозможен, поскольку, как видно из (III.235) и по рис. III.32, a , их вершины (минимумы) под влиянием $C_{сп}$ и T могут пересекать левую (нижнюю) границу области эксперимента.

Если координата x_i^* (вершина квазиоднофакторной параболы, т. е. максимум или минимум \hat{Y}) определяется непрерывной функцией (III.224), то вторая экстремальная точка $x_{i,opt}^*$, наиболее далеко отстоящая от \hat{Y}^* и соответствующая другому крайнему состоянию системы — минимуму (максимуму) \hat{Y} , всегда лежит на границе области Ω_x , т. е. при $x_i = +1 \vee -1$. При этом она располагается на «дальней» границе, т. е. если $x_i^* > 0$, то $x_{i,opt}^* = -1$ и наоборот. При передвижении x_i^* под влиянием эффектов взаимодействия (III.224) значение $x_{i,opt}^* = +1 \vee -1$ остается постоянным до тех пор, пока x_i^* не перейдет через начало координат, после чего сразу примет другое дискретное граничное значение $x_{i,opt}^* = -1 \vee +1$. В точке $x_i^* = 0$ оба значения $x_{i,opt}^* = \pm 1$ равноценны. Вершина параболы не переходит через начало координат и, следовательно, граничное значение $x_{i,opt}^* = +1 \vee -1$ не меняет, знак, если выполняется условие

$$|b_i| > \sum_j |b_{ij}|. \quad (III.238)$$

Пример III.54. Анализ предела прочности бетона R по данным пр. III.53 показывает, что:

первый фактор обеспечивает $W_{1,max}^*$, когда он находится на верхнем уровне (рис. III.32, a), причем это значение не меняется в зависимости от

уровней других факторов, так как выполняется условие (III.239) $|5,5| > |2| + |1,5| = |3,5|$;

значению $W_{2. \min}^*$ могут соответствовать (рис. 32, б) как $x_2 = -1$, так и $x_2 = +1$, поскольку условие (III.238) нарушено: $|0,8| < |1,5| + |0,4| = |1,9|$;

в отношении действия на прочность бетона фактора x_3 , судя по $W_{3. \max}^*$ (рис. III.32, в), выводы аналогичны тем, которые сделаны для $W_{1. \max}^*$.

Вышеизложенные представления о квазиоднофакторных моделях систематизированы и дополнены в табл. III.36, лежащей в основе *диссоциативно-шагового метода* оптимизации (ДШМ), основные этапы которого показаны на рис. III.33. Первые этапы достаточно подробно детализированы. Эта блок-схема фрагментарно рассмотрена и в п. IV.5, посвященном диалогу «ЭВМ — технолог». При реализации ДШМ порядок некоторых этапов может быть изменен, если кроме задач поиска Y^* исследователь ставит и другие инженерные задачи (см. п. III.16, III.17).

Пример III.55. По модели (III.191) из пр. III.39 и III.54 нужно найти максимум и минимум прочности с учетом стандартной ошибки предсказания $s\{\hat{R}\}$. Последняя определяется в зависимости от координат точки x_i^* по соотношению, полученному из (III.144) для симметричных планов эксперимента V_i , с учетом (III.185), (III.186):

$$s^2\{\hat{Y}\} = s_0^2 [T_1 + (T_3 + 2T_2) \sum_i x_i^2 + (T_6 + 2T_5) \sum_{i < j} x_i^2 x_j^2 + T_4 \sum_i x_i^4]. \quad (\text{III.239})$$

Для плана V_3 ($n_0 = 1$ — см. прил. II.12) при $s_0 = 0,63$ МПа соотношение (III.239) конкретизируется [6] как

$$s^2\{R\} = 0,1147 - 0,0485(x_1^2 + x_2^2 + x_3^2) - 0,0386(x_1^2 x_2^2 + x_1^2 x_3^2 + x_2^2 x_3^2) + 0,1544(x_1^4 + x_2^4 + x_3^4). \quad (\text{III.240})$$

В соответствии с квазиоднофакторными моделями (III.231)—(III.233) и рис. III.32 поиск R_{\max} выполняется за три шага, а поиск минимума — за пять шагов (рис. III.34).

При *поиске максимума*, во-первых, в связи с выполнением условия (E) в модель (1) подставлено $x_3 = +1$ и получена модель (2), во-вторых, аналогично установлено, что $x_1 = +1$ и получена модель (3), в-третьих, из (G. I) следует, что координата $x_2^* = -0,5b_2/b_{22} = 0,89$, в которой $R_{\max} = 37,22$ МПа. По (III.240) ошибка в точке максимума ($x_1 = x_3 = +1$, $x_2 = 0,89$) оценивается как $s\{\hat{R}_{\max}\} = 0,52$ МПа.

При *поиске минимума* на первом шаге по условию (G. III) определено $x_3 = -1$ и получена двухфакторная модель (7), на втором — условие (B) указало на необходимость перехода к двум конкурирующим моделям — (8) и (9). На третьем шаге в связи с (G. I) определено по (III.14), что $x_{i. \text{opt}}^* = -0,53$ (10) и $\hat{R}_x = 10,57$ МПа (11) при $s\{\hat{R}_x\} = 0,52$ МПа (12). На четвертом — по конкурирующей модели (6), для которой выполняется условие (G. III), определено $x_{i. \text{opt}}^* = -1$ (13), $\hat{R}_{\min} = 7,8$ МПа и $s\{\hat{R}_{\min}\} = 0,56$ МПа.

В связи с использованием в ходе минимизации \hat{R} *принципа конкурирующих моделей* очень важен пятый шаг, на котором не только определяется, что \hat{R}_{\min} лежат на ветви (9) ... (15) при $x_1 = x_3 = -1$ и $x_2 = +1$, но и сравни-

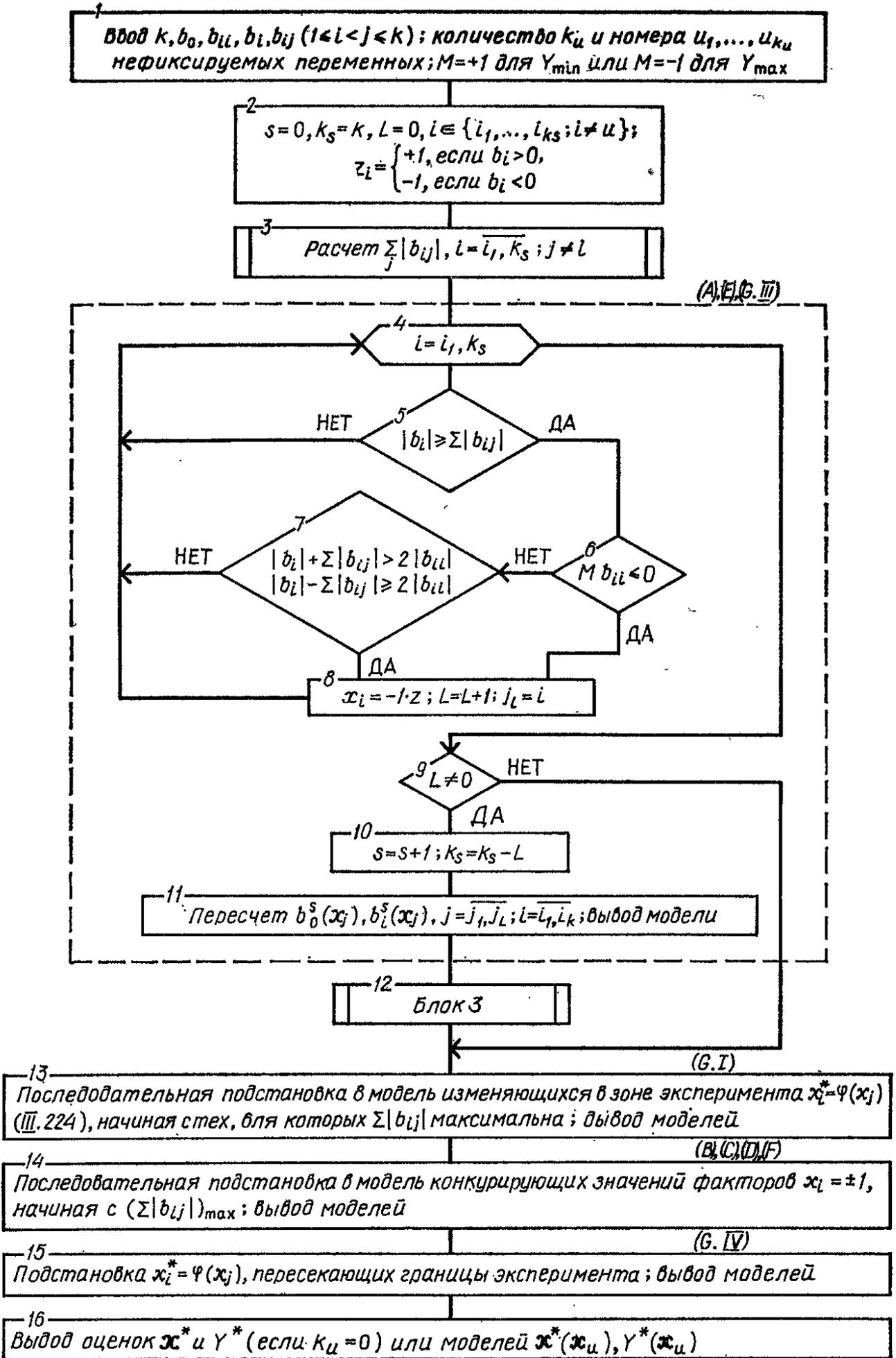


Рис. III. 33. Блок-схема алгоритма диссоциативно-шагового метода

Таблица III.36. Поиск оптимальных x_i^* по k -факторной модели с помощью ДШМ

Поиск решения для				Положение $x_{e. ext}$			
min Y		max Y		min Y max Y			
b_{ii}	b_i	b_{ii}	b_i	$\sum b_{ij} $	$ b_i + \sum b_{ij} > 2 b_{ii} $ (I)		
					$ b_i + \sum b_{ij} \leq 2 b_{ii} $ (I)	$ b_i - \sum b_{ij} \geq 2 b_{ii} $ (III)	$ b_i - \sum b_{ij} < 2 b_{ii} $ (IV)
	> 0		< 0	$ b_i \geq \sum b_{ij} $ (A)	-1 (A)		
				$ b_i < \sum b_{ij} $ (B)	-1 или +1 (B)		
≤ 0	0	≥ 0	0	$b_{ij} = 0$ (C)	при $b_{ii} \neq 0$: -1 или +1 (C') при $b_{ii} = 0$: любое (C'')		
				$ b_i < \sum b_{ij} $ (D)	-1 или +1 (D)		
	< 0		> 0	$ b_i \geq \sum b_{ij} $ (E)	+1 (E)		
				$ b_i < \sum b_{ij} $ (F)	-1 или +1 (F)		
> 0	> 0	< 0	< 0		$x_{i. ext}$ находится за границами эксперимента -1 (G'.III)	$x_{i. ext}$ переходит через границы эксперимента, за которыми равен 1 (G.IV)	
	0	0	0	$x_{i. ext}$ находится в зоне эксперимента (G.I)	-		
	< 0	> 0	> 0		+1 (G''.III)		

ваются полученные результаты (12) и (15) с помощью t -критерия, построенного аналогично (III.72), но учитывающего число опытов в плане N (4):

$$t = (\hat{R}_x - \hat{R}_{\min}) \sqrt{N} / \sqrt{r^2 \{\hat{R}_x\} + s^2 \{\hat{R}_{\min}\}} \quad (III.241)$$

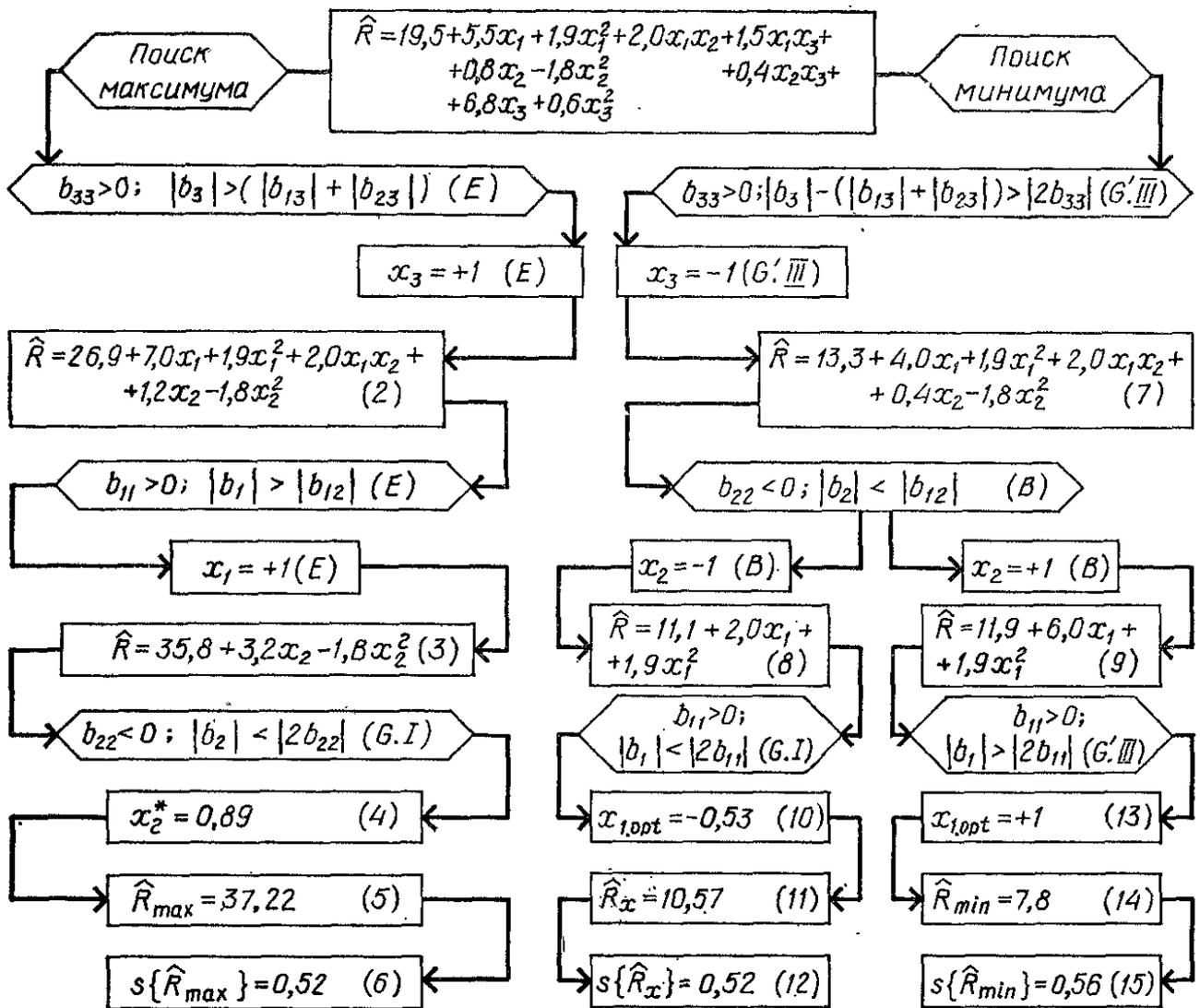


Рис. III. 34. Схема поиска R_{min} и R_{max} диссоциативно-шаговым методом

Проверка по (III.241) нуль-гипотезы $H_0: \eta\{R_x\} = \eta\{R_{min}\}$ для № 15 показала, что по t -критерию

$$t = (10,57 - 7,80) \sqrt{15} / \sqrt{0,52^2 + 0,56^2} = 14,$$

ее следует отвергнуть, так как $t > t_{\alpha} = \{f_3 = 30\} = 3,65$ даже при $\alpha = 0,001$ (см. прил. II.5). Следовательно, координаты R_{min} определены однозначно.

Одно из достоинств оптимизации ДШМ состоит в том, что оператор (исследователь) может построить алгоритм оптимизации (рис. III.33) так, чтобы после каждого шага (или серии шагов) *обсуждать* в технологических терминах результаты последовательной оптимизации. При этом целесообразно начинать оптимизацию с тех факторов, для которых сравнительно *невелики затраты* на регулировку и которые достаточно *просто поддерживать* на любом необходимом уровне (в частности, концентрации химических добавок, время некоторых технологических циклов и т. п.). Последовательность оптимизируемых факторов строится так, чтобы решение в отношении *основных технико-экономических факторов* (в частности, расходы цемента или полимера, технологической энергии и т. п.) принималось на *последних этапах*, что позволяет

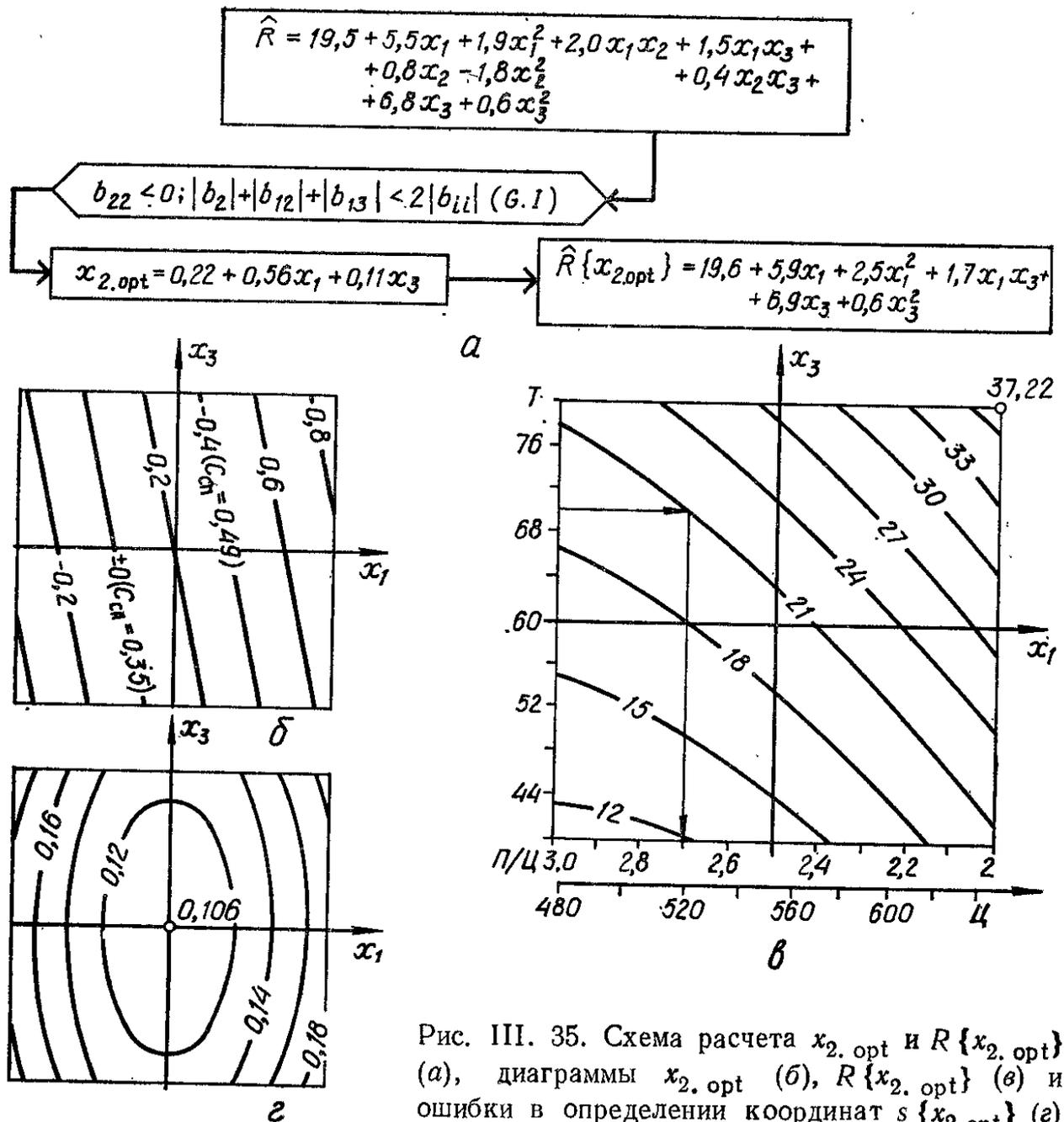


Рис. III. 35. Схема расчета $x_{2,opt}$ и $R\{x_{2,opt}\}$ (а), диаграммы $x_{2,opt}$ (б), $R\{x_{2,opt}\}$ (в) и ошибки в определении координат $s\{x_{2,opt}\}$ (г)

в дальнейшем оперативно решать ряд задач оптимизации второго вида, в том числе об экономии ресурсов (см. п. III.16, III.17).

Пример III.56. Начинать оптимизацию прочности бетона R в условиях пр. III.53 — III.55 целесообразно с поиска оптимальной концентрации суперпластификатора $C_{сп} = X_2 = 0,35 \pm 0,35 \%$, поскольку дозировка этого компонента не представляет большой сложности на любом уровне.

Так как условие (G.I) (табл. III.36) выполняется (рис. III.35) и оптимум x_2 всегда находится в области эксперимента, то следует рассчитать функцию оптимального управления дозировкой суперпластификатора

$$x_{2,opt} = \frac{0,8 + 2x_1 + 0,4x_3}{-2(-1,8)} = 0,22 + 0,56x_1 + 11x_3, \quad \text{(III.242)}$$

из которой следует, что каждому расходу цемента (соотношение П/Ц = X_1) и каждой температуре ТВО (X_3) соответствует своя оптимальная концентрация добавки; изолинии линейной модели (III.242) показаны на рис. III.35, б.

На втором этапе необходимо рассмотреть, как будет изменяться R в зависимости от Ц и Т, если концентрация суперпластификатора будет оптимальной для любого сочетания этих рецептурно-технологических факто-

ров. Ответ на этот вопрос даст подстановка $x_{2,opt}$ (III.242) в исходную модель (III.191). Для расчетов новой модели целесообразно использовать формулы прироста оценок (III.226):

$$\begin{aligned} \Delta b_0 &= -0,25 \cdot 0,8^2 / (-1,8) = 0,09; \\ \Delta b_1 &= -0,5 \cdot 0,8 \cdot 1,5 / (-1,8) = 0,33; \quad \Delta b_2 = 0,09; \quad \Delta b_{12} = 0,17; \\ \Delta b_{11} &= -0,25 \cdot 1,5^2 / (-1,8) = 0,31; \quad \Delta b_{22} = 0,02. \end{aligned}$$

Учет приростов в модели (III.191) приводит к модели прочности мелкозернистого бетона с оптимальным содержанием суперпластификатора

$$\hat{R}\{x_{2,opt}\} = 19,6 + 5,9x_1 + 2,5x_1^2 + 6,9x_3 + 0,6x_3^2 + 1,7x_1x_3, \quad (III.243)$$

изолинии которой показаны на рис. III.35, в. Комплекс диаграмм на рис. III.35, б, в не только указывает единственное значение $\hat{R}_{max} = 37,22$ МПа в точке $x_1 = x_3 = +1$ (правая диаграмма) при $x_{2,opt} = 0,89$ или $C_{сп} = 0,66\%$ (левая диаграмма), но и позволяет решать так называемые регулировочные задачи. Например, задана прочность $R = 21$ МПа при температуре ТВО 70° ; определяются при $x_3 = (70 - 60)/20 = 0,5$, во-первых, необходимый расход цемента $x_1 = -0,4$ или П/Ц = $-(-0,4) \times 0,5 + 2,5 = 2,7$, во-вторых, оптимальная концентрация суперпластификатора $x_{2,opt} = -0,05$ или $C_{сп} = (-0,05) \cdot 0,35 + 0,35 = 0,33\%$ в точке $\{-0,4; +0,5\}$.

Следует отметить, что решение об оптимальной дозировке добавки принимается на основе ЭС-модели (III.191), имеющей ошибку предсказания (III.240). Естественно, что и оптимальная дозировка $x_{2,opt}$ (III.242) оценивается [6, 8, 23] с дисперсией

$$s^2\{x_{i,opt}\} = \left\{ \left(s^2\{b_i\} + \sum_j (x_j^2 s^2\{b_{ij}\}) + \left[b_i^2 + \sum_j (x_j^2 b_{ij}) \right] \frac{s^2\{b_{ii}\}}{b_{ii}^2} \right) \right\} / \frac{b_{ii}^2}{4}, \quad (III.244)$$

функция которой выводится из (III.224) аналогично функции случайных величин (III.144).

С учетом информации о плане B_3 (№ 3 в табл. III.33) модель (III.244) конкретизируется как

$$s^2\{x_{2,opt}\} = 0,0308 (0,3657 + 0,6051x_1^2 + 0,1442x_3^2).$$

Изолинии $s\{x_{2,opt}\}$ в виде семейства эллипсов с центром $x_1 = x_3 = 0$ показаны на рис. III.35, г. Наиболее точно значение $x_{2,opt}$ определяется в центре эксперимента, где $s\{x_{2,opt}\} = 0,106$; по мере удаления от него ошибка увеличивается. Результат $x_{2,opt}\{x_1 = x_3 = +1\} = 0,89$ следует рассматривать только совместно с $s\{x_{2,opt}\} = 0,184$.

Полиномиальные модели могут быть использованы для поиска оптимума и вне области эксперимента Ω_x . Если границы допустимой области поиска оптимума Ω определены, то кроме общих методов оптимизации (см. гл. II) в ряде случаев удается применить и ДШМ, предварительно нормализовав факторы в этой области за счет ведения новых — $X_{0,i}$ и ΔX_i (В.2), (В.3).

В том случае, когда ЭС-модель линейна по факторам (III.87), поиск оптимума вне области эксперимента Ω_x можно вести по одной из модификаций градиентного метода (гл. II) — по методу *крутого восхождения Бокса — Уилсона* [1, 9], аналогичному методу наискорейшего спуска, рассмотренному в п. III.3. Особенность

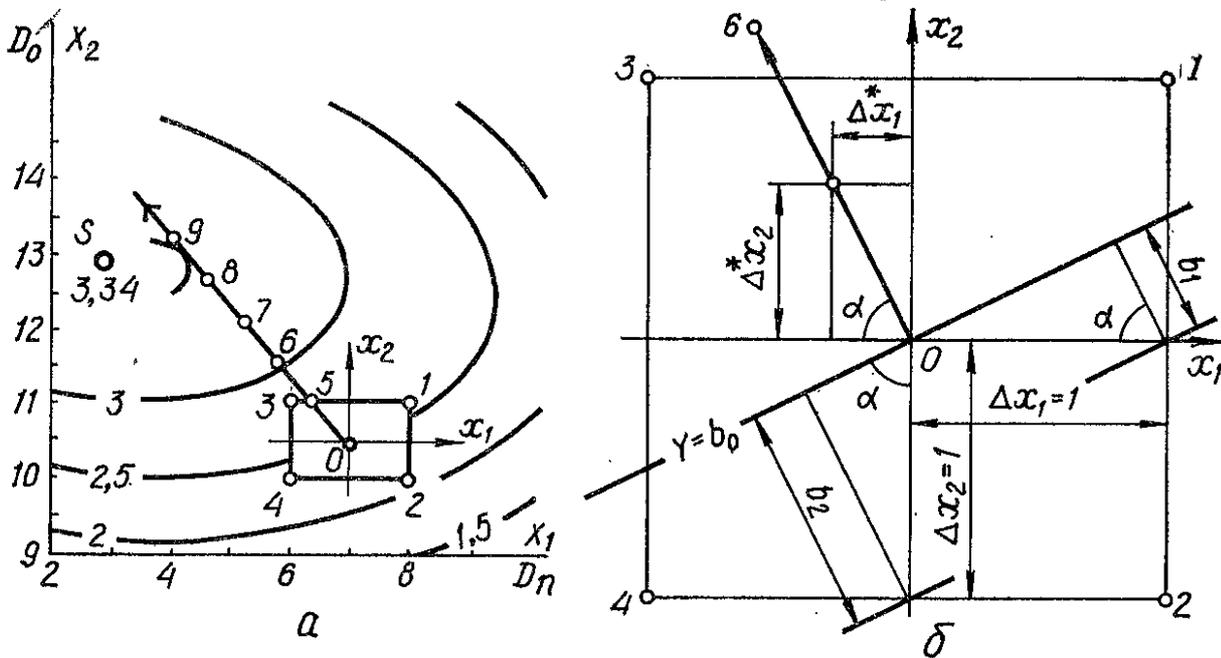


Рис. III. 36. Экспериментальный поиск оптимума по методу Бокса-Уилсона: a — общий принцип выбора точек эксперимента; b — геометрические зависимости при формировании градиента

его заключается в том, что оценки градиента (II.9) определяются по результатам полного (рис. III.36, a — точки 1 — 4) или дробного факторного эксперимента как оценки коэффициентов линейной ЭС-модели.

Чтобы от центра эксперимента (точка с координатами $0, \dots, 0$ области Ω_x) перейти в окрестность точки оптимума показателя качества Y_{opt} (точка S на рис. III.36, a), необходимо сделать ряд экспериментальных рабочих шагов (точки 5 — 9) по кратчайшему пути, т. е. по градиенту ΔY (II.9). Для реализации такого движения следует изменять факторы x_1, \dots, x_n пропорционально соответствующим коэффициентам линейной модели и в ту сторону, куда указывает знак коэффициента.

Так, в ситуации с двумя нормализованными факторами (рис. III.36, b) движение от центра $x_1 = x_2 = 0$ по градиенту (перпендикуляр к изолинии $\hat{Y} = b_0$), проходящий под углом $180^\circ - \alpha$ к оси x_1 , осуществится всегда, когда будет постоянным соотношение шагов по координатным осям $\Delta x_2^* / \Delta x_1^*$, равное $\operatorname{tg} \alpha$. При изменении x_1 на $\Delta x_1 = 1$ изолиния сдвигается на величину b_1 и $\cos \alpha = b_1 / 1 = b_1$. Аналогично при изменении x_2 на $\Delta x_2 = 1$ изолиния сдвинется на величину b_2 и $\sin \alpha = b_2 / 1 = b_2$. Следовательно, $\operatorname{tg} \alpha = \sin \alpha / \cos \alpha = b_2 / b_1 = \Delta x_1^* / \Delta x_2^*$. Для движения по ∇Y ставится пошаговый эксперимент последовательно в точках 5—9 (рис. III.36, a) до тех пор, пока в некоторой точке (9) не начнется снижение экспериментальных значений выхода по сравнению с лучшим значением \tilde{y}_u на линии градиента. В эксперименте натуральные факторы изменяются на величину ΔX_i^* , пропорциональную произведению $b_i \Delta X_i$; коэффициент пропорциональности λ выбирается одинаковым для факторов так, чтобы один из шагов $\Delta X_i^* = \lambda b_i \Delta X_i$ был удобен для реализации экспериментатору.

Пример III.57. На первом поисковом этапе разработки полимерного клея для монтажа железобетонных конструкций [8] необходимо максимизировать прочность R , МПа, и изгиб клеевого шва, регулируя концентрацию двух компонентов — $D_{\text{п}}$ и $D_{\text{о}}$.

Результаты ПФЭ (табл. III.37) в четырех вершинах квадрата (рис. III.36,а) описываются линейной моделью

$$R = 2,48 - 0,135x_1 + 0,225x_2.$$

Шаг крутого восхождения (табл. III.37) выбирается так, чтобы было удобно изменять $D_{\text{о}}$ по 0,5 %, что дает общий множитель $\lambda = 0,5/0,1125 = 0,444$. Движение от центра $\{D_{\text{п}} = 7, D_{\text{о}} = 10,5\}$ осуществляется одновременным уменьшением содержания $D_{\text{п}}$ на 0,6 % и увеличением $D_{\text{о}}$ на 0,5 %.

Таблица III.37. Матрица планирования ПФЭ 2^2 , результаты эксперимента и расчет крутого восхождения

Параметры эксперимента	$D_{\text{п}}$	$D_{\text{о}}$	Предел прочности R , МПа
Основной уровень $X_{0,i}$	7	10,5	
Интервал варьирования ΔX_i	1	0,5	
Уровни нормализованных факторов x_i в точках:			
1	+1	+1	2,57
2	+1	-1	2,12
3	-1	+1	2,84
4	-1	-1	2,39
Оценки коэффициентов b_i	-0,135	+0,225	
Произведение $b_i \Delta X_i$	-0,135	+0,1125	
Шаг $\Delta X_i^* = \lambda b_i \Delta X_i$ ($\lambda = 0,444$)	-0,6	0,5	
Уровни натуральных факторов X_i в точках линий ∇Y :			
5	6,4	11	Мысленный опыт
6	5,8	11,5	3,03
7	5,2	12	3,11
8	4,6	12,5	3,18
9	4	13	3,15

Пятый опыт пропущен (хотя был рассчитан как «мысленный»), поскольку точка 5 относится к Ω_x . В точке 8 обнаружена $R_{\text{max}} = 3,18$, а после эксперимента в точке 9 (где $R_9 < R_8$) движение следует прекратить и приступить к моделированию в области оптимума.

Основной недостаток метода крутого восхождения в том, что результат неинвариантен к изменению интервалов варьирования ΔX_i (В.3) в области эксперимента Ω_x : при изменении в m раз интервала варьирования ΔX_i абсолютное значение оценки коэффициента b_i также возрастает в m раз, следовательно, шаг движения

по градиенту ΔX_i^* возрастет в m^2 раз; в этих условиях инвариантны только знаки составляющих градиента [9].

Метод крутого восхождения дает *полезные результаты* в том случае, когда *время ожидания отклика мало* (например, от начала смешивания компонентов для изготовления образцов до получения окончательной числовой оценки показателя качества готового материала). В этом случае можно быстро построить первую многофакторную линейную модель, используя близкий к насыщенномуДФЭ (см. п. III.11), потом сделать ряд шагов по первому градиенту; выйдя в новую область, построить вторую линейную и, вновь определив откорректированное направление градиента, сделать шаги к оптимуму и т. д. Однако время ожидания отклика при исследовании свойств вяжущих бетонов и других твердеющих композитов, как правило, исчисляется сутками и месяцами, что резко сужает круг задач, решаемых в материаловедении и технологии с использованием этого метода.

В целом задачу поиска по ЭС-моделям экстремальных значений каждого из выходов системы можно считать решенной.

III.16. Решение с помощью полиномиальных моделей задач оптимизации второго вида — минимизации расхода ресурса

В практике анализа, проектирования и управления в области строительного материаловедения и технологии *наиболее распространены* оптимизационные задачи второго вида (см. рис. В. 5) — необходимо достигнуть *заданного уровня** функционирования системы $Y_{j,тр} = \text{const}$, израсходовав при этом *минимальный ресурс* x_i .

При решении таких задач возникает ряд трудностей как формально-математических, так и логико-профессиональных. К числу первых, в частности, относится тот факт, что нелинейное уравнение для определения уровня любого x_i , соответствующего $Y_{j,тр} = \text{const}$, получается из ЭС-модели в неявном виде:

$$(b_0 - \hat{Y}_{j,норм}) + \sum_{i=1}^k b_i x_i + \sum_{i < j} b_{ij} x_i x_j + \sum_{i=1}^k b_{ii} x_i^2 = 0. \quad (\text{III.245})$$

При этом одному значению $Y_{j,тр} = \text{const}$ соответствует не одна пара корней (за исключением параболы при $k = 1$), а *множество точек на изоповерхности* второго порядка в k -мерном пространстве, на которой и нужно найти значение x , соответствующее минимуму расхода ресурса по одному из факторов. Эти трудности могут быть преодолены с помощью методов нелинейного программирования (см. гл. II), а также при достаточно гибком применении принципов ДШМ оптимизации (см. п. III.15), направленных на пошаговое

* «Требуемый уровень» — предельное состояние системы (см. «Введение»), поэтому естественно, что в общем случае его можно превысить.

уменьшение размерности факторного пространства за счет стабилизации части факторов на оптимальных уровнях $x_{j,\text{opt}}$. Процедура численного решения задач, связанных с уравнением (III.245), включая построение на ЭВМ соответствующих изоповерхностей, сводится к определению для однофакторного уравнения от x_1 (без потери общности) корней $x_1^{(0)}$ после фиксации всех $(k-1)$ -х факторов на некоторых уровнях $x_j = \text{const}$, что отражается на значениях свободного члена $b_0\{x_j\}$ и линейного эффекта $b_1\{x_j\}$:

$$(b_0\{x_j\} - Y_{\text{тр}}) - b_1\{x_j\}x_1 + b_{11}x_1^2 = 0; \quad (\text{III.246})$$

$$x_1^{(0)} = -\frac{b_1\{x_j\}}{2b_{11}} \pm \sqrt{\frac{1}{b_{11}} \left\{ Y_{\text{тр}} - \left[b_0\{x_j\} - \frac{(b_1\{x_j\})^2}{4b_{11}} \right] \right\}}. \quad (\text{III.247})$$

Изолиния $Y_{\text{тр}}$ на плоскости $\{x_1, x_2\}$ возникает при пошаговом изменении в (III.247) второго фактора $j = 2$ (без потери общности). Шаг ΔX_2 выбирается в зависимости от точности расчетов и размера диаграммы на дисплее, графопостроителе и т. п. Объемная изоповерхность $Y_{\text{тр}}$ внутри куба (см. рис. III.27, б) возникает при дополнительном пошаговом изменении $j = 3$ и т. п.

К числу логико-профессиональных трудностей в оптимизационных задачах второго вида относится, в частности, выбор решения в том случае, когда желательно получить экономию ресурса не по одному x_1 , а по нескольким факторам. При этом нередко приходится использовать неформальные критерии и работать в режиме диалога «ЭВМ — исследователь» (см. п. IV.5). Аналогичные трудности возникают и тогда, когда задана группа нормативов свойств $Y_{1,\text{тр}}, Y_{2,\text{тр}}, \dots$, для которых нужно найти компромиссное решение (см. рис. В. 5, г); отдельные рекомендации для поиска таких решений даны в п. III.17.

Для некоторых типовых инженерных задач, связанных с минимизацией расхода ресурсов на основе ЭС-моделей, разработаны [4, 6, 8, 22, 23 и др.] достаточно подробные рекомендации по решению, однако их применение всегда требует неформального подхода к существованию материаловедческих и технологических задач.

Пример III.58. Изменение прочности бетона R , МПа, при управлении Ц/П, T и $C_{\text{сп}}$ (см. пр. III.33), описываемое моделью (III.191), графически отображается семейством изоповерхностей-гиперболоидов (рис. III.27, б.) Для изоповерхности $R_{\text{тр}} = 30$ МПа (рис. III.37, а) наиболее экономичный состав по расходу тепловой (технологической) энергии, характеризуемой при прочих равных условиях температурой изотермической выдержки T , соответствует точке А. Она лежит на плоскости $x_1 = +1$ (наибольший расход Ц) и в то же время принадлежит плоскости оптимальной переменной концентрации суперпластификатора $x_{2,\text{opt}}$ (III.242, III.243). При одновременной подстановке в (III.243) условия $x_1 = +1$ и требования $R_{\text{opt}} = 30$ получается уравнение

$$-2 + 8,6x_3 + 0,6x_3^2 = 0,$$

один из корней которого $x_3^{(0)}\{A\} = 0,234$ дает координату точки А. Наименьшая температура ТВО, обеспечивающая $R_{\text{тр}} = 30$ МПа, равна $T\{A\} = 20 \times$

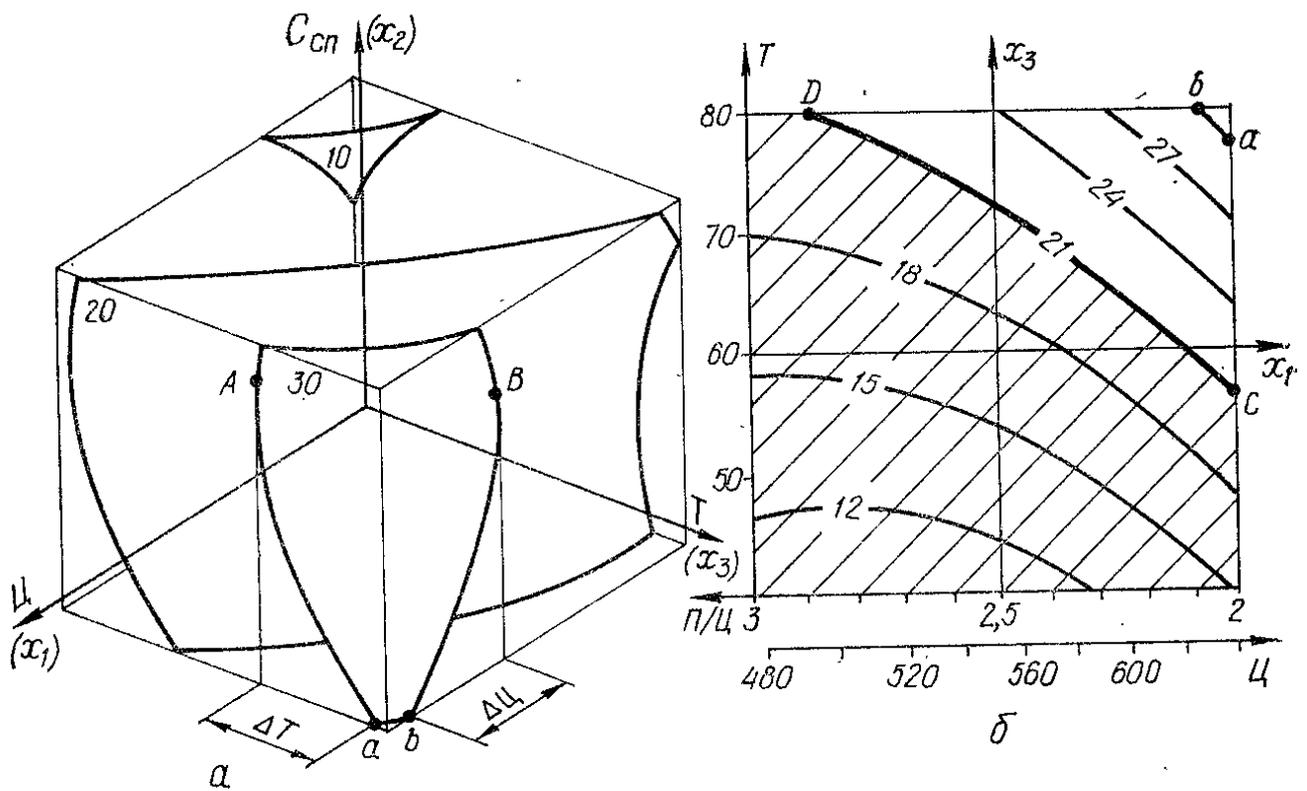


Рис. III. 37. Изоповерхности прочности бетона R (а) и изолинии R (б) для бездобавочного бетона («дно» куба)

$\times 0,234 + 60 = 64,7$ °С. Следует отметить, что этот же результат можно получить для точки A непосредственно по регулировочной диаграмме на рис. III.35, в.

Экономичность решения оценивается при сравнении с базовым вариантом технологии, в качестве которого логично принять бездобавочный бетон. Прочность такого «эталонного» материала описывается моделью

$$R_g = 16,9 + 3,5x_1 + 1,9x_1^2 + 6,4x_3 + 0,6x_3^2 + 1,5x_1x_3, \quad (\text{III.248})$$

полученной при подстановке $x_2 = -1$ в общую модель (III.191). Изолинии (III.248) показаны на диаграмме III.37, б («дно» куба на рис. III.37, а). Температура в базовом варианте (точка a) по уравнению

$$-7,7 + 7,9x_3 + 0,6x_3^2 = 0,$$

полученному при подстановке в (III.248) $x_1 = +1$, должна быть $T = 78,2$ °С ($x_3^{(0)}\{a\} = 0,912$). Следовательно, за счет введения в бетоны из равноподвижных смесей [6] оптимальной добавки $C_{сп}$ можно при сохранении требуемой прочности $R_{тр} = 30$ МПа снизить температуру ТВО на

$$\Delta T = T\{a\} - T\{A\} = 78,2 - 64,7 = 13,5$$
 °С.

Для $R_{тр} = 30$ МПа наиболее экономичный состав по расходу цемента — точка B , где $x_1\{B\} = 0,343$ (при $x_3 = +1$ расчет аналогичен вышеизложенному), а $\Pi/\text{Ц} = 2,5 - 0,343 \cdot 0,5 = 2,33$. Расход цемента на 1 т бетона составит $\text{Ц} = 1000(1 + 2,33) = 300$ кг/т, что при средней плотности свежеуложенной мелкозернистой бетонной смеси $\gamma = 1925$ кг/м³ [6] дает расход цемента $\text{Ц}\{B\} = 578$ кг/м³.

Эталонный бетон имеет $R_{тр} = 30$ МПа при минимальном расходе $\text{Ц}\{b\} = 631$ кг/м³ (точка b : $x_1 = 0,907$; $\Pi/\text{Ц} = 2,05$; $\text{Ц} = 328$ кг/т — рис. III.37, б). Следовательно, при введении оптимальной добавки, уровень которой $x_{2,opt} = 0,84$ или $C_{сп} = 0,65$ % легко рассчитывается по (III.242), можно сэкономить цемента $\Delta\text{Ц} = \text{Ц}\{b\} - \text{Ц}\{B\} = 631 - 578 = 53$ кг/м³, т.е. уменьшить его расход на 8,4 %.

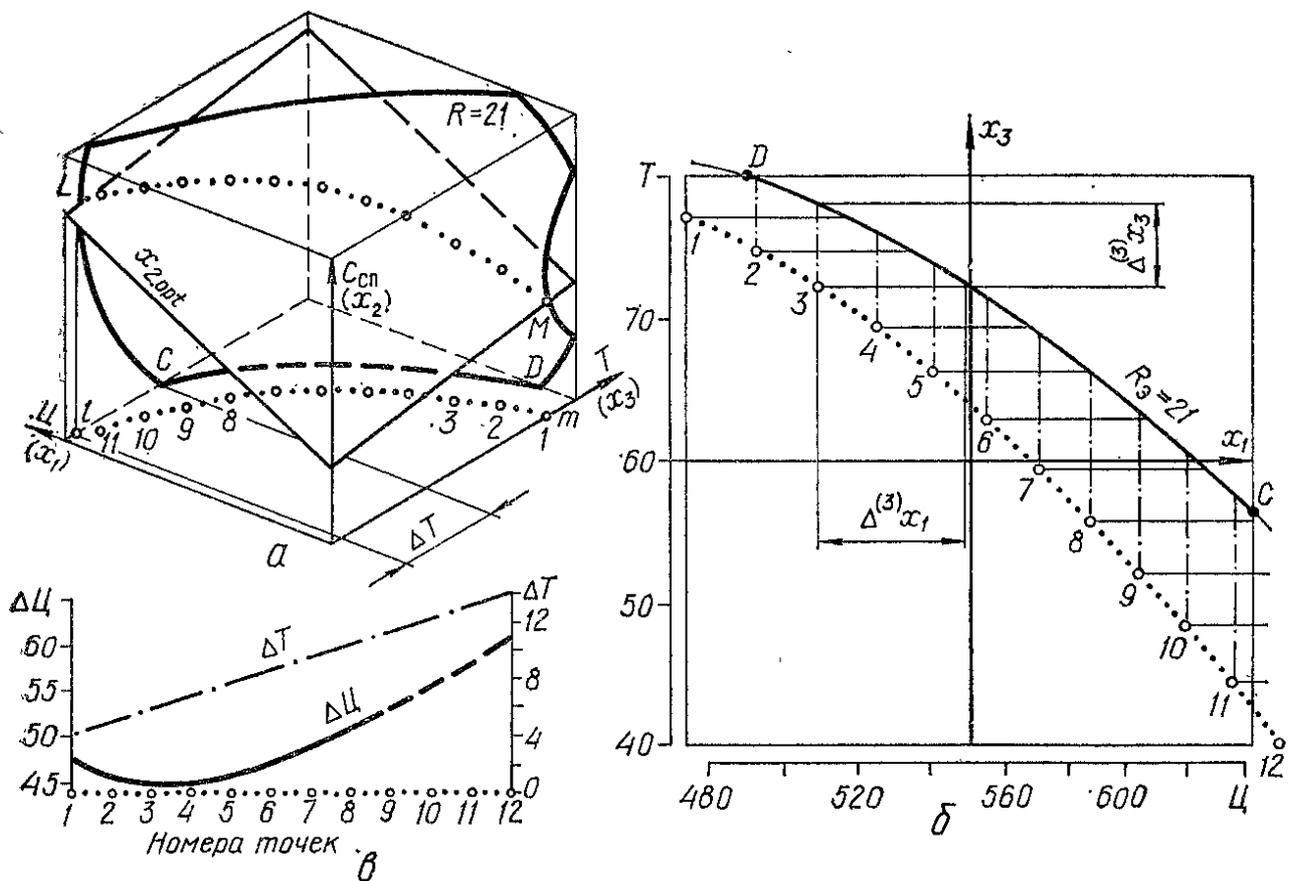


Рис. III. 38. Прочность бетона R при $x_{2, opt} = \varphi(x_1, x_3)$ в трехфакторном пространстве:

a — при $R = 21$ МПа линии LM соответствуют минимальные ζ и T ; b — графоаналитический метод расчета $\Delta \zeta$ и ΔT ; c — результаты расчета $\Delta \zeta$ и ΔT вдоль линии LM

При ориентации одновременно на экономию обоих видов ресурса (тепловой энергии и цемента) вопрос не имеет однозначного ответа. Так, для $R_{тр} = 21$ МПа (это задание соответствует 70 %-й прочности бетона марки М300 после ТВО) все возможные решения для эталонного бетона (рис. III.37, б) отражаются множеством точек на соответствующей изолинии от точки D (минимальный расход цемента $\zeta\{D\} = 497$ кг/м³ при $T\{D\} = 80$ °С) до точки C (минимальная температура ТВО $T\{C\} = 58$ °С при $\zeta\{C\} = 642$ кг/м³).

Решения, соответствующие $R_{тр} = 21$ МПа, при введении добавки пластификатора с переменной оптимальной концентрацией $x_{2, opt}$ лежат на изолинии ML (рис. III.38, а), образующейся при пересечении изоповерхности $R = 21$ МПа и плоскости $x_{2, opt} = \varphi(x_1, x_3)$. Если эту линию спроектировать на плоскость эталонного бетона, то проекция lm будет находиться левее линии CD для $R_э = 21$ МПа, т. е. в области меньшего расхода цемента и пониженной температуры ТВО. Область между линиями lm и CD — это *пространство экономии ресурсов* за счет оптимальной переменной дозировки добавки. Для решения задачи об экономии ресурсов рекомендуется [6] численный метод — изолиния lm (рис. III.38, б) квантуется на N_L равных частей и в каждой точке определяются расстояния по горизонтали (экономия цемента) и по вертикали (снижение температуры ТВО) до эталонной кривой CD .

Результаты расчетов $\Delta \zeta$ и ΔT могут быть представлены в виде графиков (рис. III.38, в), соответствующих изменению экономии ресурсов при движении вдоль изолинии $R_{тр} = 21$ МПа от высокотемпературных режимов ТВО бетонов с пониженным расходом цемента (точка 1) к низкотемпературным ТВО для бетонов с высоким ζ (точка 11). Окончательный выбор того или иного пути от линии CD к линии lm полностью формализован быть не может, так как приоритет критериев ζ и T (или связанной с ними себестоимости изделий) зависит от условий конкретного производства.

Для сравнения вариантов инженерных решений необходимы их те или иные числовые характеристики. Выше сравнение эталонного бетона и бетона с химдобавкой-регулятором проводилось по R_{\max} , по C_{\min} или T_{\min} при $R_{\text{тр}}$ и пр. Можно сформулировать и другие *обобщающие показатели* для регулировочных диаграмм на рис. III.36, б, рис. III.37, б и т. п. Так, в [8] введена оценка вероятности P_F появления в исследованной области факторного пространства Ω_x инженерных решений, удовлетворяющих заданному нормативу. На рис. III.37, б заштрихована «запрещенная» (см. гл. II) область $\Omega_{\text{зпр}}$, в которой $R < R_{\text{тр}} = 21$ МПа. Незаштрихованная область $\Omega_{\text{тр}}$ содержит множество *допустимых решений*, удовлетворяющих $R \geq R_{\text{тр}}$. По сути, показатель P_F — геометрическая вероятность [32, 53], оцениваемая для двухфакторного случая как отношение (процентное) площадей

$$P_F = 100S\{\Omega_{\text{тр}}\}/S\{\Omega_x\}. \quad (\text{III.249})$$

Площадь области допустимых решений $S\{\Omega_{\text{тр}}\}$ можно рассчитать как двойной интеграл на отрезках $[-1, +1]$ численными методами (см. гл. I), а $S\{\Omega_x\}$ для квадрата при $|x_i| \leq 1$ всегда равна 4. Так, при $R = 21$ МПа обобщающий показатель для эталонного бетона (рис. III.37, б) лишь $P_F = (0,76/4) \cdot 100 = 19\%$, в то время как для бетона с оптимальной добавкой рис. III.35, в) он возрастает до 44 %.

Необходимо подчеркнуть, что все решения по минимизации ресурсов носят вероятностный характер, так как принимаются на основе ЭС-моделей, предсказывающих оценку отклика с ошибкой $s\{\hat{Y}\} = \varphi(X)$ и имеющих ошибки оценок коэффициентов $s\{b\}$. При заданном уровне $Y_{\text{тр}} = \text{const}$ выход системы — детерминированная величина, но соответствующие ему уровни факторов $x_i^{(0)}\{Y = \text{const}\}$ превращаются в случайные величины. Их среднеквадратичная ошибка $s\{x_i^{(0)}\}$ зависит как от ошибки предсказания выхода по модели $s\{\hat{Y}\}$ (а следовательно, от ошибки эксперимента s_y и координат точки $x_i^{(0)}$ — см. (III.145) и (III.240)), так и от геометрической формы той гиперповерхности $\hat{Y} = f(x)$, по которой решается задача (см., например, выражение (III.244) для $s^2\{x_{i.\text{opt}}\}$ в пр. III.55). В результате множество точек $x_i^{(0)}\{Y = \text{const}\}$ располагается, вообще говоря, не только на изолинии $\hat{Y} = \text{const}$, но и возле нее в некотором коридоре, ширина которого пропорциональна $s\{x_i^{(0)}\}$. Методы оценки доверительных интервалов для случайных величин $x_i^{(0)}\{Y = \text{const}\}$, получаемых при решении задачи оптимизации второго вида, достаточно сложны и пока получены [8, 23] для ограниченного круга ситуаций, причем ряд методов ориентирован не на аналитическое, а на численное построение таких оценок. В целом проблематика решения задач минимизации расхода ресурсов по ЭС-моделям требует интенсивного развития работ по созданию методического и программного обеспечения.

III.17. Принятие инженерных решений по комплексу экспериментально-статистических моделей

В материаловедческих и технологических задачах поведение исследуемой (проектируемой, управляемой) системы характеризуется обычно *группой критериев* качества $Y_1, \dots, Y_j, \dots, Y_Q$, *координаты оптимума* которых $x^*\{Y_1\}, \dots, x^*\{Y_j\}, \dots, x^*\{Y_Q\}$, как правило, *не совпадают* (см. «Введение», рис. В. 5). В силу этой объективной закономерности возникает новый круг проблемных вопросов, связанный с принятием *компромиссных решений* в многокритериальных инженерных задачах. Выбор компромиссного решения не может быть формализован без введения дополнительных правил [6, 8, 9, 23, 43, 48 и др]. Среди них при начальном ознакомлении с существом вопроса целесообразно отметить следующие.

Размер группы критериев качества должен быть уменьшен за счет установления *корреляционных связей* между критериями (с риском α , см. п. III.4) и исключения тех из них, изменение которых допустимо оценивать по другим остающимся критериям.

Можно построить (см. гл. II) на основе единичных критериев некоторый *интегральный критерий* как функцию $V = \psi(Y_1, \dots, Y_j, \dots, Y_Q)$ и для него искать экстремальные значения. Примером такого интегрального критерия является себестоимость единицы продукции.

Можно с помощью квалифицированных экспертов установить *приоритет критериев*, проранжировав их как $Y^{(i)}$ по степени важности для состояния данной системы в целом. Компромиссная задача решается последовательно путем поиска области $\Omega\{Y^{(1)}\}$, удовлетворяющей требованиям по наиболее важному критерию, затем в этой области определяется подобласть $\Omega\{Y^{(2)}\}$, удовлетворяющая требованиям по второму критерию, и т. д. Принцип приоритета весьма плодотворен — он используется при построении интегральных критериев, в частности критерия желательности [9].

Для поиска компромиссных решений весьма удобно использовать оптимизацию ДШМ (см. п. III.15), особенно в условиях диалоговой системы «ЭВМ — человек» (см. п. IV.5). Сначала для каждого из Q критериев Y_j находятся *индивидуальные координаты оптимума* $x^*\{Y_j\}$ в виде k -мерных векторов. Далее формируется матрица размером $k \times Q$, в которой Q критериев размещаются ранжированно по приоритетам. В матрице выбирается и *стабилизируется* тот фактор x_i , уровни которого $x_i^*\{Y_j\}$ для всех или наиболее важных критериев *совпадают*. Таким образом, размерность факторного пространства *сокращается* на единицу, а результаты стабилизации фактора на общем уровне $x_{i, \text{общ}}$ оцениваются по снижению уровня тех критериев, для которых $x_{i, \text{общ}} \neq x_i^*\{Y_j\}$. Если результаты приемлемы для технолога, то процедура повторяется, начиная с поиска $(k - 1)$ -й индивидуальной координаты оптимума.

Этот способ позволяет принимать компромиссное решение очень гибко, оценивая в инженерных терминах результаты каждого ша-

га, и при необходимости возвращаться назад для «раскачивания» решения по другим вариантам $x_{i.общ}$.

После того как за R шагов размерность факторного пространства сокращена до $k - R \ll 3$, принятие компромиссных решений резко упрощается, поскольку появляется возможность графически отобразить любое количество изоповерхностей для Q показателей качества. На последнем этапе, как правило, используется [4, 6, 8] комплекс двухфакторных диаграмм, каждая из которых может быть представлена континуально (семейством изолиний) или дискретно (таблица с шагом квантования Δx_i), в зависимости от возможностей периферийных устройств вычислительной техники.

При классификации компромиссных задач по инженерным формулировкам наиболее распространенной оказывается задача о *выборе рецептуры и технологических режимов* получения материалов с *комплексом свойств*, отвечающих в общем случае двусторонним («коридорным») нормативам

$$Y_{j.тр.мин} \leq Y_j \leq Y_{j.тр.макс} \quad (j = \overline{1, Q}). \quad (\text{III.250})$$

В области Ω_x сначала должна быть определена разрешенная (III.250) подобласть $\Omega_{тр}$, а затем в ней точка, удовлетворяющая или требованиям экономии ресурсов (см. п. III.16), или наибольшей устойчивости решения к колебаниям процесса (см. п. IV. 3) и др. [4, 6, 8, 23].

Пример III.59. При разработке композиционного материала [8] для ремонта и восстановления железобетонных конструкций, испытывающих существенные динамические нагрузки, необходимо обеспечить его повышенные демпфирующие свойства по сравнению с чистой эпоксидной смолой. Исходя из условий работы конструкций, нормативные уровни назначены для модуля упругости $E_{тр} \leq 3,5$ ГПа и для предела прочности на изгиб $R_{тр} \geq 30$ МПа. Кроме того, для получения жизнеспособной технологической смеси пластическая прочность композиции $P_m = 0,5$ МПа должна достигаться не раньше чем за $\tau = 1$ ч и не позже чем за $\tau = 4,5$ ч.

Для управления структурой и прочностью композита использованы наполнители разной природы (цемент, поливинилхлоридный порошок — ПВХ, их смеси) и модифицирующие эпоксидную матрицу добавки (кремний-органическая жидкость — КОЖ и латекс). Если в многофакторных моделях стабилизировать на лучших уровнях количество отвердителя и наполнителей, то влияние изменения концентрации (% от массы смолы) модификаторов $X_1 = C_k = 0,5 \pm 0,5$ и $X_2 = C_l = 15 \pm 15$ на механические свойства и технологический показатель $Y = 10^3 (\ln \tau)$ описывается двухфакторными моделями:

$$E = 2,85 - 0,57x_1 + 0,90x_1^2 - 0,64x_2 + 0,29x_2^2; \quad (\text{III.251})$$

$$R = 20,53 - 6,80x_1 + 1,66x_1^2 - 16,41x_2 + 4,81x_2^2 + 3,85x_1x_2; \quad (\text{III.252})$$

$$\hat{Y} = 711 - 345x_1 - 398x_1^2 + 1274x_2 + 547x_2^2 + 461x_1x_2. \quad (\text{III.253})$$

Изолинии модуля упругости E (III.251) показаны на рис. III.39, а. Для допустимой подобласти $P_P\{E\} = 68\%$. Заданный уровень E не может быть обеспечен за счет одной модифицирующей добавки. Минимальный расход КОЖ обеспечивается при максимальном расходе латекса $x_2\{A\} = +1$ ($C_l = 30\%$) и составляет $x_{1.мин}\{A\} = -0,783$ ($C_k = 0,11\%$). Минимальный

расход латекса $x_{2, \min} \{B\} = -0,838$ ($C_{л} = 2,4 \%$) достигается при оптимальной концентрации КОЖ $x_{1, \text{opt}} \{B\} = 0,317$ ($C_{к} = 0,66 \%$).

Для прочности R (III.252) подобласть допустимых решений (рис. III.39, б) $P_F \{R\} = 26 \%$, т. е. меньше, чем для первого показателя, и расположена в иной части исследованной области Ω_x . В нее включается и немодифицированная эпоксидная матрица ($x_1 = x_2 = -1$; $C_{к} = C_{л} = 0$), для которой $R = R_{\max} = 54,6$ МПа. Немодифицированная матрица — самое экономичное решение по критерию R .

Изолинии времени τ , ч, формирования заданной структуры (рассчитываются по модели (III.253) с пересчетом $\tau = e^{0,001Y}$) показаны на рис. III.39, в. Допустимая подобласть представлена коридором ($P_F \{\tau\} = 64 \%$). Выше верхней границы структурообразование замедляется большим содержанием латекса; нижняя граница, отделяющая смеси с малой жизнеспособностью, вызывает сужение коридора при повышении $C_{к} \geq 0,5 \%$. Наиболее экономичны смеси при немодифицированной матрице ($x_1 = x_2 = -1$; $1 < \tau = 1,48 \leq 4,5$ ч).

Подобласть $\Omega_{\text{тр}}$, соответствующая одновременно трем требованиям к композиционным материалам, оценивается при «совмещении» всех диаграмм (рис. III.39, г). Справа эта подобласть в виде треугольного сектора ограничена изолинией $\tau = 1$ ч, сверху — изолинией прочности $R = 30$ МПа, а слева — изолинией модуля упругости $E = 3,5$ ГПа, причем именно на ней между точками L и M лежат наиболее экономичные составы. Точка L соответствует минимальному содержанию КОЖ $C_{к} = 0,29$, что, естественно, больше, чем его концентрация в точке A ($C_{к} = 0,11 \%$, рис. II.37, а) за счет введения сверху ограничения по прочности композита. Точка M , в которой $C_{л} = 2,9 \%$, незначительно отличается от минимально возможного расхода латекса $C_{л} \{B\} = 2,4 \%$ (рис. III.37, а). Координаты точек на ЭВМ рассчитываются численным итерационным методом, так как затруднительно аналитическое совместное решение пары уравнений, следующих из (III.25) — (III.253), при неявном виде приравниваемых координат.

Подобласть допустимых решений относительно невелика ($P_F \{E, R, \tau\} = 5 \%$), поэтому целесообразность реализации рецептуры, например $C_{к} = 0,5 \%$ и $C_{л} = 5 \%$, существенно зависит от технических возможностей точного дозирования модифицирующих добавок ($\Delta C_{к} < 0,1 \%$ и $\Delta C_{л} < 1 \%$).

С помощью группы двухфакторных диаграмм можно решить еще одну важную инженерную задачу, характерную для практики материаловедческих исследований.

В ряде случаев целесообразно анализировать поведение системы по некоторому j -му показателю качества при условии, что *уровень* другого ее отклика постоянен — $Y_1 = \text{const}$ (без потери общности). Такой анализ назван [4] *изопараметрическим*. Так, в технологии бетона уровни свойств затвердевшего материала нередко (см., частности, пр. III.53) сравниваются между собой при условии постоянства такого косвенного реологического параметра, как подвижность бетонной смеси. Эти смеси называются равноподвижными, а в более общей терминологии — изореологическими. В технологии ячеистых бетонов их механические показатели сравниваются при одинаковых плотности или пористости материала (изопоризованные композиты) и т. п.

Экспериментальное определение составов и технологии изготовления изопараметрических образцов весьма трудоемко, так как требуются предварительные опыты для достаточно точного попада-

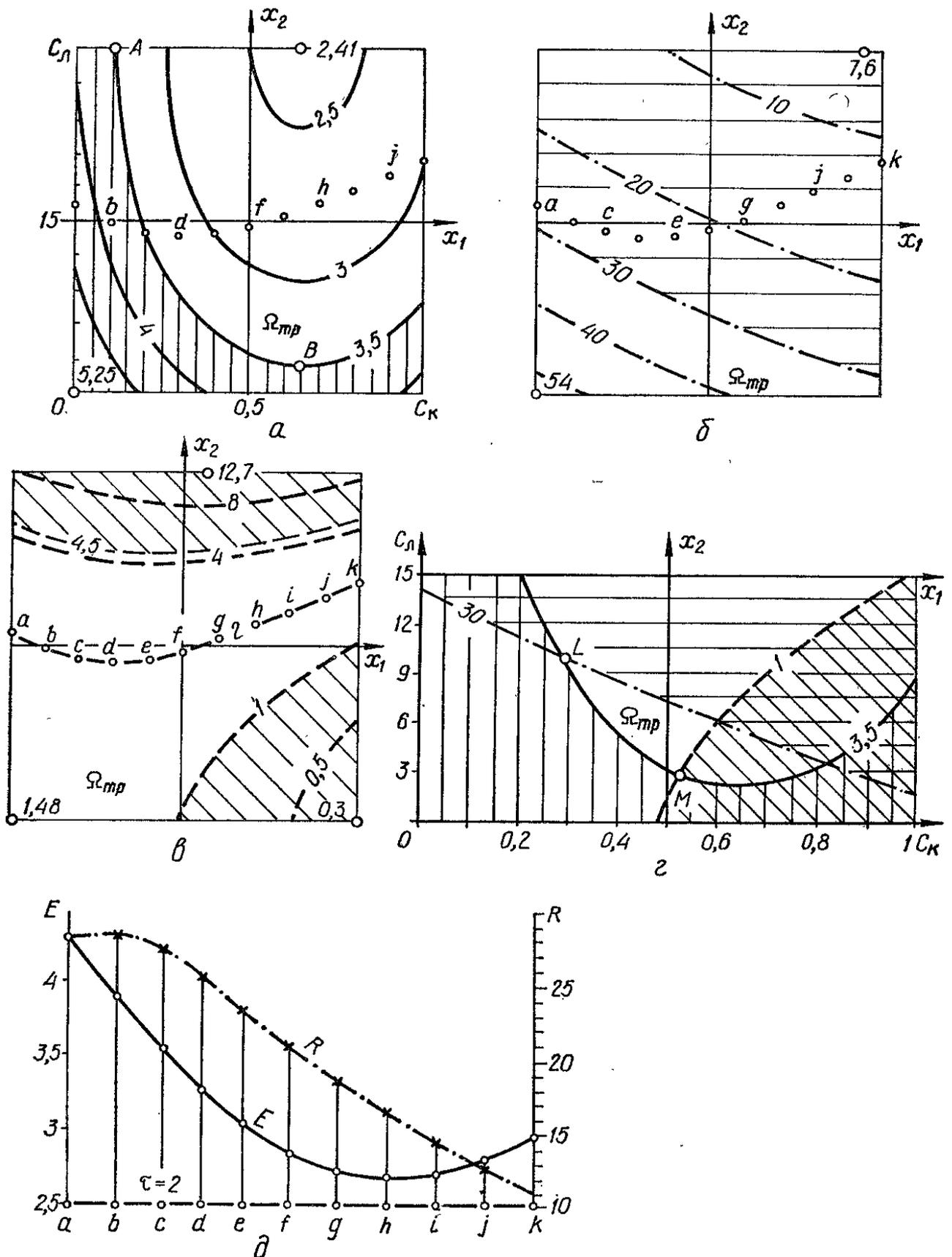


Рис. III. 39. Комплексный анализ свойств (E — модуль упругости, R — предел прочности, τ — время отвердевания композита по двухфакторным моделям, x_1 — концентрация кремнийорганической жидкости, x_2 — концентрация латекса); a — изолинии E и область допустимых решений $E_{\text{Тр}} < 3,5$ МПа; b — изолинии R и область $R_{\text{Тр}} > 30$ МПа; v — изолинии τ и область $1 < \tau < 4,5$ ч; e — допустимая область $\Omega_{\text{Тр}}$ для всех трех свойств; d — изопараметрический анализ E и R при $\tau = 2 = \text{const}$

ния на заданный уровень свойства, например на заданную вязкость смеси. В то же время использование комплекса ЭС-моделей позволяет решать такие задачи достаточно просто и точно.

По определению (см. «Введение»), на любой изолинии двухфакторной диаграммы уровень отклика системы постоянен — $Y_1 = f(x_1, x_2) = \text{const}$, а следовательно, движение вдоль изолинии возможно только тогда, когда каждому изменению фактора x_2 соответствует функционально заданное изменение фактора $x_1 = \psi(x_2)$. Поскольку уровни факторов x_1 и x_2 вдоль изолинии $Y_1 = \text{const}$ изменяются, то это влечет за собой изменения j -го ($j = \overline{2, Q}$) свойства материала. Анализируя эти изменения, можно получить *новую* материаловедческую и технологическую *информацию* [6, 8, 13 и др.].

Пример III.60. В условиях задачи о разработке полимерсодержащего композита (пр. III.59) представляет интерес анализ изменения механических свойств затвердевшего материала при условии постоянства его начальных структурных характеристик. Одной из таких характеристик является пластическая прочность смеси P_m . Одинаковая пластическая прочность $P_m = 0,5$ МПа достигается смесями разного состава за разное время τ , но на изолиниях $\tau = \text{const}$ (рис. III.39, в) допустимо считать, что структурообразование твердеющих композиций находится в одинаковой стадии.

На рис. III.39, в вдоль изолинии $\tau = 2$ ч сделано слева направо (т. е. при увеличении содержания C_k) 10 шагов, границы которых обозначены точками a, \dots, k . Длина шага произвольна, но фиксирована. Между соседними точками i и j длина дуги изолинии l_{i-j} определяется на ЭВМ численным интегрированием [8]. Движение вдоль этой изолинии дает прямую $\tau = \Phi_\tau(\sum l_{i-j}) = 2 = \text{const}$, которая и принимается за ось абсцисс в изопараметрическом анализе. Если же двигаться по точкам a, \dots, k , но на диаграммах $E = f_E(x_1, x_2)$ и $R = f_R(x_1, x_2)$ (рис. III.39, а, б), то такое движение будет пересекать изолинии этих свойств. Полученные таким образом графики $E = \Phi_E(\sum l_{i-j})$ и $R = \Phi_R(\sum l_{i-j})$ в общем случае могут быть параболами четвертой степени [4, 8]. Переход от a к k , обеспечиваемый одновременным изменением концентраций добавок C_k и C_d , вызывает сначала снижение модуля упругости, а затем его некоторый рост. В тех же условиях предел прочности на изгиб композита сначала практически не реагирует на малые добавки C_k , а затем начинает снижаться. Между точками c и g кривые обоих процессов идут параллельно, и в этой области показатели E и R будут сильно коррелировать (что и ожидается, исходя из общей теории прочности материалов). Однако наличие на кривой $R = \Phi_R(\sum l_{i-j})$ начального горизонтального участка указывает на то, что существуют рецептуры композита, у которых при равном начальном структурообразовании упругость снижена относительно сильнее, чем их когезионная прочность. Это и позволяет получать материалы с повышенными демпфирующими свойствами.

Вышеизложенные методы решения компромиссных задач на двухфакторных диаграммах могут быть развиты на трехфакторные ситуации, рассматриваемые на объемных диаграммах. Это нередко оказывается необходимым в тех случаях, когда область допустимых решений P_F слишком мала и целесообразно искать возможности ее расширения до заданного значения $P_{F, \text{тр}}$ за счет изменения уровней третьего фактора.

Пример III.61. При разработке композита для ремонта конструкций материал с заданным комплексом свойств (см. пр. III.59) был получен в весьма

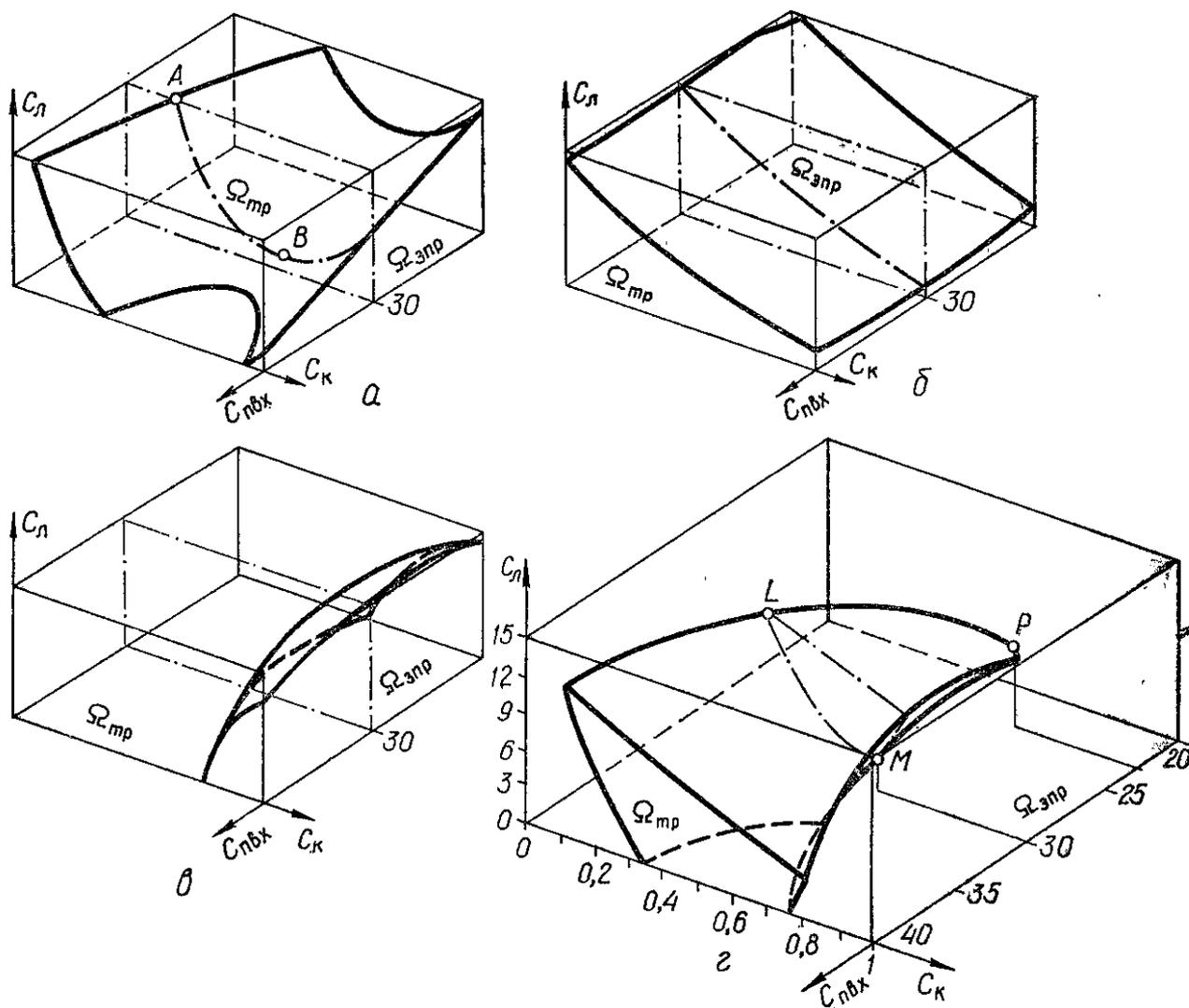


Рис. III.40. Комплексный анализ свойств композита по трехфакторным моделям (x_3 — массовое наполнение частицами ПВХ):

a — изоповерхность, ограничивающая область $E_{\text{тр}} < 3,5$ ГПа; b — то же $R_{\text{тр}} > 30$ МПа; $в$ — то же, $\tau > 1$ ч; $г$ — допустимая область для всех трех свойств; точка P соответствует минимальному расходу ПВХ; точки A, B, M комментируются на рис. III.39

малой подобласти $P_F\{R, E, \tau\} = 5\%$, что вызвало повышение требований к точности дозировки компонентов. Этот материал содержит 30% наполнителя поливинилхлорида (ПВХ), поэтому в дополнение к модификации эпоксидной матрицы добавками целесообразно проанализировать возможности изменения свойств за счет управления количеством ПВХ. Влияние варьирования содержания ПВХ при одновременной модификации матрицы описывается трехфакторными моделями, для которых граничные изоповерхности ($E = 3,5$ ГПа, $R = 30$ МПа и $\tau = 1$ ч) показаны на рис. III.40; для $C_{\text{л}}$ и $C_{\text{пвх}}$ показана только половина факторного пространства; по оси $C_{\text{пвх}}$ масштаб нормализованного фактора для наглядности увеличен вдвое.

Для модуля упругости (рис. III.40, a) разрешенная область расположена выше изоповерхности $E = 3,5$ ГПа и представляет собой вырезку из эллиптического цилиндра. По мере увеличения количества наполнителя минимально допустимая дозировка латекса уменьшается, а при $C_{\text{пвх}}$ более 35% могут быть получены отвечающие нормативу безлатексные композиты; вследствие этого область $\Omega_{\text{тр}}$ с увеличением $C_{\text{пвх}}$ расширяется. На когезионную прочность (рис. III.40, b) уровень дозировки $20 \leq C_{\text{пвх}} \leq 40\%$ влияет незначительно. В этих же условиях нижняя граница жизнеспособности смеси $\tau = 1$ ч отодвигается в сторону больших концентраций КОЖ. Это вызывает увеличение $\Omega_{\text{тр}}\{\tau\}$ с ростом $C_{\text{пвх}}$ (рис. III.40, $в$).

Одновременное выполнение требований по всему комплексу свойств обеспечивается в подобласти $\Omega_{\text{тр}}$, имеющей сложную объемную форму (рис. III.40, а), образованную при пересечении граничных изоповерхностей. С увеличением количества ПВХ (по сравнению с $C_{\text{ПВХ}} = 30\%$ — см. рис. III.39, а) точка минимального расхода КОЖ модификатора сдвигается от L влево, достигая $C_K = 1\%$, а точка минимального расхода латекса сдвигается от M вниз и, начиная с $C_{\text{ПВХ}} = 34\%$, возможно получение требуемых композиций без латекса.

При $C_{\text{ПВХ}} = 40\%$ доля допустимых решений достигает $P_V = 13\%$, а в точке P , соответствующей минимально возможному содержанию наполнителя $C_{\text{ПВХ}} = 23\%$, эта доля стремится к нулю. В то же время точка P является весьма важной для технолога, так как характеризует предельное состояние системы при заданных нормативах функционирования. Правее точки P , например, при $C_{\text{ПВХ}} = 20\%$ получить композит с заданным комплексом свойств нельзя. Если же расход ресурса $C_{\text{ПВХ}} > 20\%$ невозможен, то следует или искать добавочные рецептурно-технологические факторы, управление которыми смогло бы повлечь образование подобласти $\Omega_{\text{тр}}$ с $P_F > 0$, или снижать некоторые нормативные требования $Y_{j,\text{тр}}$ (например, перейти от $R_{\text{тр}} = 30$ МПа к $R_{\text{тр}} = 20$ МПа — см. рис. III.39, б).

Поиск точки P и выбор технологических решений при разработке композиционных материалов предложено [8] осуществлять на ЭВМ численным методом (предпочтительно в режиме диалога «ЭВМ — исследователь» — см. п. IV.5). На внешние устройства ЭВМ подается информация в графической или табличной форме об изменении всех нормируемых показателей свойств в заданном сечении факторного пространства, в частности при $C_{\text{ПВХ}} = 20\%$. Сечение последовательно «передвигается» с шагом $\Delta C_{\text{ПВХ}}$ в направлении повышения расхода ресурса до тех пор, пока не появится область $\Omega_{\text{тр}}$ нужного размера $P_F\{E, R, \tau, \dots\}$.

Следует еще раз подчеркнуть, что изменение нормативных требований $Y_{j,\text{тр}}$ к качеству функционирования системы приводит к трансформации области $\Omega_{\text{тр}}$, изменению доли разрешенных решений P_F и уровней минимального расхода ресурсов $x_{i,\text{min}}\{\Omega_{\text{тр}}\}$.

Вышеизложенные методы поиска рецептурно-технологических решений по комплексу полиномиальных моделей ориентируются на использование на последних этапах ДШМ двухфакторных диаграмм, наиболее удобных для профессионально-логического обсуждения. Сведение всех многофакторных ситуаций к двухфакторным не может служить существенным ограничением применимости методов, поскольку, во-первых, можно построить основанные на «плоских» диаграммах более сложные системы отображения (куб — пр. III.39, призма — пр. III.44, «квадрат на квадрате» и «треугольник на треугольнике» — [8, 41 и др.]), во-вторых, можно изменить по тем или иным соображениям уровни фиксации не входящих в отображение факторов. Возможности этих методов расширяются в связи с развитием оперативного применения в работе технологов вычислительной техники, в том числе персональных компьютеров с программным обеспечением графических построений на дисплее и режима диалога «ЭВМ — технолог».

Глава IV. ПРИНЦИПЫ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

IV.1. Использование моделей в режиме имитации

Любая модель в принципе имитационная, поскольку она имитирует (лат. *imitatio*— подражание) реальный объект. Однако сложились термины «имитационное моделирование»* и «имитационная модель», которые характерны для вполне определенного класса действий с моделями объектов при решении инженерных, технико-экономических и других задач. *Имитация* в этом смысле (43) — это *анализ системы с помощью вариантных расчетов*.

Имитационная модель воспроизводит процесс функционирования объекта во времени, причем имитируются *элементарные явления*, составляющие процесс, с сохранением их *логической структуры* и *последовательности* протекания во времени, что позволяет по исходным данным получить сведения о состоянии процесса в определенные моменты времени [55].

В технических объектах проводится предварительная *декомпозиция* (лат. *de...* — приставка, означающая отмену и т. п.) процесса функционирования объекта на составляющие подпроцессы или подсистемы и для описания тех из них, для которых это возможно и целесообразно, используются функциональные зависимости — аналитические модели (уравнения математической физики, экспериментально-статистические и др.— см. «Введение»). Для остальных подпроцессов строятся алгоритмы, описывающие правила функционирования подсистем с той или иной степенью формализации. При реализации этих правил может быть предусмотрено как машинное обращение к неформальной информации — *базе данных* о поведении объекта, специально собранных в памяти ЭВМ для решения определенного круга задач (это могут быть и простейшие таблицы, фиксирующие требования стандартов на промышленную продукцию, и описания всех изобретений в данной отрасли техники и т. п.), так и *взаимодействие* человека-оператора (исследователя, проектировщика) с ЭВМ в режиме *диалога* (см. п. IV. 5).

Имитационная система кроме комплекса имитационных моделей данного объекта включает две группы математического обеспечения: специально разработанный для данной системы набор алгоритмов и программ, обеспечивающий эффективный диалог человека и ЭВМ, носит название внутреннего математического обеспечения, а алгоритмы и программы стандартного типа для вспомогательных (сервисных) задач типа ввода и вывода информации, накопления базы данных называются внешним математическим обеспечением. Техническое обеспечение имитационной системы включает в себя прежде всего средства вычислительной техники, ввода и вывода информации, а также средства связи между ЭВМ и человеком.

* В английском языке ему соответствует термин *simulation*, который в русском языке ближе к понятию «ложное поведение».

Таким образом, имитационная система может рассматриваться [55] как *машинный аналог* сложного реального процесса. Эксперимент с реальным процессом функционирования системы заменяется *машинным экспериментом* с математической моделью в режиме имитации этого процесса в ЭВМ.

Имитационное моделирование на ЭВМ является эффективным (а нередко единственным) инструментом исследования и проектирования сложных систем, когда реальный эксперимент невозможен или нецелесообразен по технико-экономическим причинам. К достоинствам имитационного моделирования, в частности, относятся: возможность исследовать поведение объекта в любых условиях (естественно, в области действия модели!); значительное сокращение продолжительности испытаний по сравнению с натурным экспериментом; гибкость структуры модели и возможность замены ее элементов (моделей подсистем, алгоритмов, базы данных и пр.) в ходе экспериментов с целью поиска оптимального варианта как имитационной модели, так и самой системы при ее проектировании.

К недостаткам имитационного моделирования относится [55] то, что в результате машинной реализации всей цепи анализа системы будет получено *частное численное решение*, отвечающее фиксированным параметрам системы, уровням факторов, алгоритмам поведения и другим элементам структуры модели. Поэтому для анализа результатов процесса функционирования системы необходимо проводить имитационный эксперимент *множественно, варьируя* исходные условия каждой реализации. Это требует значительных затрат машинного времени на проведение имитационных экспериментов.

Одна из наиболее развитых областей применения имитационных моделей при решении инженерных задач — это исследование процессов функционирования систем, подверженных случайным воздействиям, методом *статистического моделирования*. При этом *исходные условия* каждой реализации *задаются* как *случайные величины*, закон распределения которых *известен*, а *результаты* множества таких реализаций, называемых *статистическими испытаниями*, рассматриваются как *выборочная совокупность* числовых значений выхода системы. Для этой совокупности нужно найти статистические оценки неизвестного распределения (по тем же алгоритмам, которые применяются для обработки результатов натурального эксперимента) и сделать инженерные выводы о поведении системы. В основе статистического моделирования лежит универсальный численный метод решения математических задач при моделировании случайных величин — метод Монте-Карло* [54].

Пример IV.1. Рассматривается многофакторная технологическая система [4], для которой требуется оценить изменение во времени промежуточного показателя качества технологии — формовочной прочности бетона R_{ϕ} , МПа,

* Монте-Карло — столица княжества Монако, знаменитая старейшими игорными домами, в которых используется рулетка — простейший механический прибор для генерирования случайных чисел от 0 до 36, одно из которых является выигрышным.

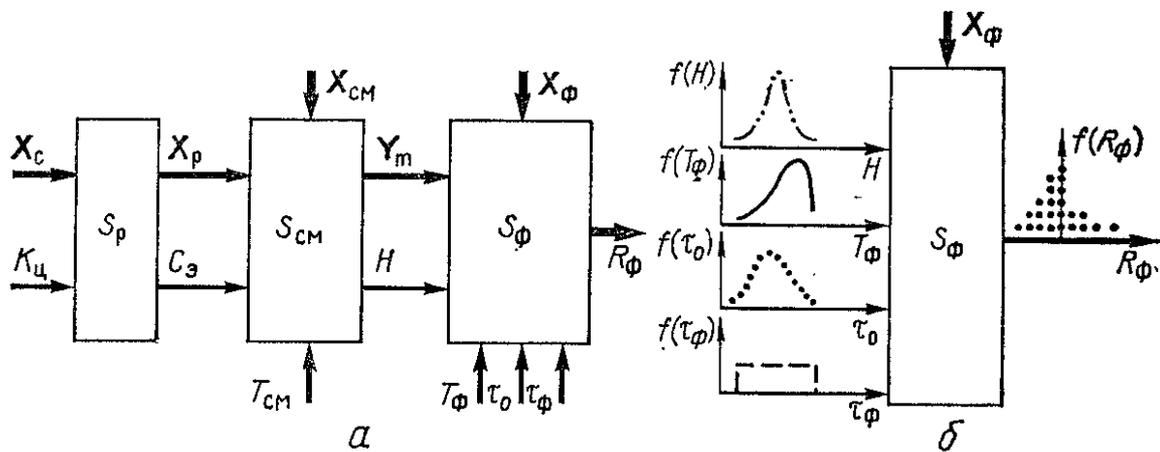


Рис. IV. 1. Декомпозиция технологического процесса (а) и схема анализа устойчивости одного из блоков по методу Монте-Карло (б)

необходимой для принятия решений о снятии бортовой оснастки. Систему при декомпозиции можно представить в виде трех последовательно соединенных подсистем S_i (рис. IV.1, а):

формирования рецептуры смеси S_p , имеющей на входе группу сырьевых факторов X_p , из которых выделен особо класс цементов K_c на трех дискретных уровнях (— 1 — шлакопортландцемент, 0 — портландцемент, + 1 — быстро-твердеющий портландцемент);

смешивания компонентов S_{cm} , имеющей на входе группу рецептурных факторов X_p (выделена концентрация добавки-электролита $C_э = (3 \pm 1) \%$ от массы цемента) и группу факторов собственно процесса смешивания X_{cm} (выделена температура $T_{cm} = (50 \pm 30)^\circ \text{C}$ специально подогреваемой смеси);

формования изделий S_ϕ , имеющей на входе, во-первых, структурно-механические показатели качества смеси $Y_{т.см}$ после подсистемы S_{cm} , среди которых выделен инженерный реологический показатель $H = (50 \pm 40) \text{ с}$ (такие факторы, являющиеся одновременно выходами предшествующей системы, называются *агрегирующими*); во-вторых, группу факторов собственно процесса формования X_ϕ , среди которых выделены три: температура смеси при формовке $T_\phi = (50 \pm 30)^\circ \text{C}$, время термостатической выдержки смеси до формовки $\tau_0 = (15 \pm 5) \text{ мин}$ и время дополнительного пребывания (твердения) в форме до снятия бортоснастки $\tau_\phi = (30 \pm 15) \text{ мин}$.

После декомпозиции необходимо описать происходящие в подсистемах S_p , S_{cm} и S_ϕ процессы, используя предпочтительно уравнения математической физики M_1 и концептуальные модели M_2 . Однако пока не удалось получить полные комплексы таких моделей, обеспечивающие описание происходящих в системе (рис. IV.1, а) явлений и предсказывающие численное значение R_ϕ с удовлетворительной для инженеров точностью. Поэтому, продолжая перспективные исследования для получения M_1 и M_2 , в большинстве практических задач приходится строить решения на основе моделей M_3 — экспериментально-статистических.

Вне зависимости от типа используемых моделей M_i исследование системы методом Монте-Карло позволяет оценить ее устойчивость к случайным колебаниям входных факторов X_i во времени. При статистическом испытании системы каждый из факторов (рис. IV.1, б — подсистема S_ϕ) принимает численное значение в соответствии со своим законом распределения случайной величины, а в результате таких случайных сочетаний X_i будет получена гистограмма распределения показателя качества системы R_ϕ .

Метод Монте-Карло уже дал интересные и полезные инженерные результаты при использовании в режиме имитации моделей

заполнения объема материала зернами разных размеров и форм [25], при моделировании деформации и прочности бетона методами механики разрушения [31], при исследовании проницаемости бетона [50], при анализе однородности смешивания технологических смесей [8] и в ряде других материаловедческих и технологических задач.

IV.2. Генерирование случайных чисел на ЭВМ

Для статистического моделирования необходимо иметь последовательность *случайных чисел*. Если специально не оговаривается, то подразумевается, что эти случайные числа представляют собой независимую выборку из генеральной совокупности, имеющей равномерное распределение на отрезке $[0,1]$. Такая выборка может быть получена на физическом датчике, примером которого является падающая монета, генерирующая два случайных числа — $x = 0$ («орел») и $x = +1$ («решка»):

$$\begin{matrix} x_i \\ p_i \end{matrix} \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}. \quad (\text{IV.1})$$

С помощью n таких независимых генераторов, вообще говоря, получается любое n -разрядное число в двоичной системе счисления (например, 10001, выпавшее на пяти одновременно брошенных монетах — число 17). При работе с ЭВМ используются обеспечивающие (IV.1) электронные датчики, в частности анализаторы шума в лампах. При простых расчетах на микрокалькуляторах можно пользоваться готовыми таблицами случайных чисел. В прил. II.10 приведена таблица, содержащая 400 случайных цифр, объединенных по пять. Из нее выбирается любое число, начиная с которого формируется последовательность случайных чисел с нужным числом разрядов. Так, начиная с третьей цифры «5», можно генерировать случайные равномерно распределенные числа с двумя знаками после запятой: 0,51; 0,56; 0,91;...

При работе с универсальными ЭВМ физические датчики или хранящиеся в памяти таблицы случайных чисел используются редко. Как правило, случайные числа заменяются *псевдослучайными последовательностями*, имеющими очень большой период повторяемости (например, $2^{29} - 2^{38}$ [54]):

Существует ряд способов получения на ЭВМ псевдослучайных чисел ξ , соответствующих равномерному распределению в интервале $[0,1]$. Основной принцип рекуррентный — по предыдущему числу ξ_n вычисляется по специально подобранной функции $\Phi(\xi_n)$ следующее число ξ_{n+1} :

$$\xi_{n+1} = \Phi(\xi_n). \quad (\text{IV.2})$$

Самый простой способ состоит в том, что сначала берется некоторое число ξ_n с $2m$ знаками, которое возводится в квадрат, и получается число ξ_n^2 с $4m$ знаками, из которого выбирают в ка-

честве числа ξ_{n+1} средние $2m$ цифр. Способ называется *методом середины квадратов*. Его недостаток состоит в том, что получается больше, чем нужно, малых значений.

Пример IV.2. Из исходного числа 0,2306 необходимо получить три псевдослучайных числа по методу середины квадратов:

$$\left. \begin{array}{l} \xi_0 = 0,2306; \xi_0^2 = 0,05317636; \\ \quad \quad \quad \downarrow \quad \downarrow \\ \xi_1 = 0,3176; \xi_1^2 = 0,10086976; \\ \quad \quad \quad \quad \quad \downarrow \quad \downarrow \\ \xi_2 = 0,0869; \xi_2^2 = 0,00755161; \\ \quad \quad \quad \quad \quad \quad \quad \downarrow \quad \downarrow \\ \xi_3 = 0,7551\dots \end{array} \right\} \text{(IV.3)}$$

Наибольшее распространение в типовом программном обеспечении ЭВМ получил *метод сравнения* (или вычетов) [54]. Суть метода состоит в том, что выбираются числа g, a_n и M , не имеющие общих делителей, причем число M — большое. На первом этапе вычисляется набор чисел a_1, a_2, a_3, \dots , которые представляют собой остатки от деления числа ga_n (с учетом $a_i = 1$) на число $M = 2^s$, где s — длина разрядной сетки ЭВМ. Так, для 32-разрядных машин удовлетворительные (при $n \geq 500$) псевдослучайные равномерно распределенные числа получены, в частности, при $g = 5^{13}$. На втором этапе остатки набора a_1, a_2, a_3, \dots умножаются на 2^{-32} , образуя ряд случайных чисел на отрезке $[0, 1]$.

Для получения значений x_1, x_2, x_3, \dots , случайной величины, имеющей *заданное распределение* $f(x)$, можно воспользоваться общей теоремой (53), согласно которой значение x_i является решением уравнения, содержащего случайную величину ξ_i , равномерно распределенную на отрезке $[0, 1]$:

$$\xi_i = \int_{-\infty}^x f(x) dx. \quad \text{(IV.4)}$$

Так, известно (см. прил. I.4), что для величин x_1, x_2, x_3, \dots , равномерно распределенных на отрезке $[a, b]$, $f(x) = (b - a)^{-1}$, следовательно,

$$\xi_i = \int_a^{x_i} \frac{1}{b-a} dx = \frac{x_i - a}{b-a}, \quad \text{(IV.5)}$$

откуда получается формула для перехода от ξ_i $[0, 1]$ к x_i $[a, b]$:

$$x_i = \xi_i (b - a) + a. \quad \text{(IV.6)}$$

Пример IV.3. Из четырех чисел ξ_i (IV.3), образованных в IV.2, для отрезка $[-3, +5]$ будет получено:

$$\begin{aligned} x_1 &= 0,2306(5 + 3) - 3 = -1,1552; & x_2 &= 0,3176 \cdot 8 - 3 = - \\ &- 0,4592; & & \text{(IV.7)} \\ x_3 &= 0,0869 \cdot 8 - 3 = -2,3048; & x_4 &= 0,7551 \cdot 8 - 3 = 3,0408. \end{aligned}$$

Для получения из ξ_i $[0,1]$ нормально распределенных случайных чисел используются иные правила, поскольку уравнение (IV.4) будет в явном виде неразрешимо относительно x_i . В частности, можно рассчитать величину ε , распределенную приближенно нормально со средним $\eta = 0$ и дисперсией $\sigma^2 = 1$, как $\varepsilon = (2,308 + 0,271 \xi)/(1 + 0,992 \xi + 0,045 \xi^2)$, (IV.8) причем если $\xi \leq 0,5$, то следует подставлять $\xi' = 1 - \xi$, а перед ε ставить знак минус.

Пример IV.4. Из четырех равномерно распределенных чисел (VI.3) рассчитываются четыре нормально распределенных числа ε :

$$\begin{aligned} \varepsilon_1 &= -[2,308 + 0,271(1 - 0,2306)]/[1 + 0,992(1 - 0,2306) + 0,045(1 - \\ &\dots 0,2306)^2] = 1,4060; \\ \varepsilon_4 &= +(2,308 + 0,271 \cdot 0,7551)/(+1 + 0,992 \cdot 0,7551 + 0,045 \cdot 0,7551^2) = \\ &= 1,7747. \end{aligned} \quad (\text{IV.9})$$

Эффективные алгоритмы генерирования случайных величин, имеющих заданное распределение, могут быть построены по методу исключения, на суммировании простых распределений для описания более сложного и др. [54]. Как правило, один из таких алгоритмов входит в программное обеспечение ЭВМ. Для оценки пригодности совокупности генерированных случайных величин к дальнейшему использованию в методе статистических испытаний может проводиться анализ такой совокупности с точки зрения ее адекватности генерируемому закону распределения. Одним из простейших тестов для нормальной совокупности является расчет оценок среднего, дисперсии, коэффициента асимметрии и эксцесса и их сравнение с заданными доверительными границами при n реализациях (см. п. III.5).

IV.3. Анализ методом Монте-Карло устойчивости технологических решений, принимаемых по экспериментально-статистическим моделям

При анализе m результатов натурального эксперимента в одной точке k -факторного пространства (см. п. III.3 — III.5) предполагается, что все эти k исследуемых рецептурно-технологических факторов имеют неизменные уровни и именно эта стабильность определяет центр их группирования — среднее значение выхода системы \bar{y} . В то же время рассеяние результатов в некотором интервале $[y_{v(\min)}, y_{v(\max)}]$ обуславливается неконтролируемым действием случайных факторов ξ ; оно оценивается среднеквадратичным отклонением $s\{\bar{y}\}$, которое приравнивается к ошибке эксперимента s_{ε} .

Новое среднее значение выхода системы \hat{y} при переходе ее в другую точку факторного пространства, как известно (см. п. III.7 — III.17), достаточно просто предсказывается по экспериментально-статистической модели. Можно оценить ошибку предсказания $s\{\hat{y}\}$, связанную (см. п. III.8) с постоянной в исследованной области ошибкой эксперимента s_{ε} и с мерой точности модели d , зависящей от свойств плана эксперимента и от координат точки предсказания.

Однако ни ошибка эксперимента s_y , ни ошибка предсказания $s\{\hat{y}\}$ не могут охарактеризовать *устойчивость* выхода системы Y к неизбежным колебаниям во времени тех k факторов x_i , которые и в натурном эксперименте, и при прогнозе \hat{Y} по модели принимались фиксированными ($x_i = \text{const}$ ($i = \overline{1, k}$)). Эти колебания факторов во времени $x_i(\tau_e)$ приведут к дополнительному рассеянию значений Y ; при этом рецептурно-технологическое решение ($x_i = \text{const}$) следует считать неустойчивым, если малым колебаниям $x_i(\tau_e)$ будет соответствовать большое рассеяние Y_e , оцениваемое среднеквадратичным отклонением $s\{Y_e\}$, коэффициентом вариации $v\{Y_e\}$ и т. п. При сравнении двух рецептурно-технологических решений А и В, обеспечивающих равные средние значения выхода $Y^{(A)} = Y^{(B)}$, предпочтительнее то, которое имеет большую устойчивость (меньшее $s\{Y_e\}$ и т. п.).

Решение проблемы устойчивости при натурном эксперименте с системой весьма трудоемко и требует специального планирования эксперимента, основанного на многомерном дисперсионном анализе. Поэтому разработана [4, 8] ориентированная на машинный эксперимент методика, согласно которой:

на первом этапе строится экспериментально-статистическая k -факторная модель поведения системы;

на втором этапе независимо анализируются закономерности распределения случайных значений уровней факторов $x_{i.e}$, характерных для данного производства;

на третьем этапе генерируются в ЭВМ случайные числа для имитации в соответствии с результатами второго этапа случайных колебаний факторов $x_{i.e}$ вокруг выбранного рецептурно-технологического решения ($x_i = \text{const}$);

на четвертом этапе методом Монте-Карло по модели «проигрываются» n реализаций, соответствующих случайным уровням k факторов, т. е. n раз производится расчет \hat{Y} ;

на пятом этапе по алгоритмам, применяемым при статистическом анализе результатов натурального эксперимента в одной точке, оцениваются статистические характеристики генерированной в ходе реализации метода Монте-Карло выборочной совокупности \hat{Y} и рассчитывается критерий устойчивости рецептурно-технологического решения.

Пример IV.5. Прочность свежееотформованного бетона R_ϕ , получаемого при функционировании системы, показанной на рис. IV.1 (см. пр. IV.1), описывается [4] шестифакторным полиномом второй степени. Если изделие формируется через $\tau_0 = 15$ мин после начала процесса из смеси с реологическим показателем $H = 50$ с, приготовленной на шлакопортландцементе ($x_3\{K_{II}\} = -1$) с добавкой электролита $C_9 = 3\%$, то заключительный этап процесса в подсистеме S_ϕ в зависимости от двух факторов: $T_\phi = X_1 = (45 \pm 25)^\circ \text{C}$ — температуры смеси в форме, $\tau_\phi = X_5 = (30 \pm 15)$ мин — времени пребывания

$$\hat{Y} = 10^3 \ln R_\phi = 613 + 482x_1 - 241x_1^2 + 140x_1x_5 + 397x_5 - 398x_5^2. \quad (\text{IV.10})$$

Анализ производственных условий показал, что температура смеси может изменяться вокруг проектного уровня по нормальному закону $f_N(\Delta T_{\phi. \varepsilon})$ со средним $(\Delta T_{\phi. \varepsilon}) = 0$ и среднеквадратичным отклонением $s\{\Delta T_{\phi. \varepsilon}\} = 2,5^{\circ} \text{C}$, а время пребывания в форме до снятия бортоснастки отклоняется от проектного уровня на ± 3 мин, причем в этом интервале распределение случайных значений $\Delta t_{\phi. \varepsilon}$ можно считать равномерным. Следовательно, далее факторы x_i в (IV.10) должны рассматриваться как случайные величины, состоящие из двух слагаемых;

$$x_i = \Omega_i + (\omega_i p_i). \quad (\text{IV.11})$$

Первое слагаемое Ω_i — детерминированная величина; это проектный уровень фактора, вокруг которого происходят колебания. Второе слагаемое $\omega_i p_i$ — случайная величина; собственно случайной является аргумент p_i (англ. *play* — игра), который будет изменяться по заданному закону, а масштабный коэффициент ω_i сокращает нормализованное факторное пространство до размеров «игровой зоны» вокруг Ω_i . Он вычисляется как отношение характеристики случайных производственных колебаний $\Delta X_{i. \varepsilon}$ к диапазону варьирования факторов при построении модели ΔX_i . Следовательно, для T_{ϕ} коэффициент $\omega_1 = 2,5/25 = 0,1$ и относится к среднеквадратичному отклонению $s\{p_1\} = 1$, а для τ_{ϕ} коэффициент $\omega_2 = 3/15 = 0,2$ и относится к полудиапазону равномерных колебаний $[-1, +1]$.

Распределение генерированных 200 случайных чисел p_1 и p_2 в виде распечатанных на ЭВМ гистограмм показано на рис. IV.2, а, б. Случайные колебания p_1 , связанные с температурой, должны были иметь нормальное распределение со средним значением $\bar{p}_1 = 0$, среднеквадратичным отклонением $s_1 = 1$, коэффициентом асимметрии $\hat{A}_1 = 0$ и эксцессом $\hat{E}_1 = 0$, а у генерированного распределения $\hat{p} = 0,04$, $\hat{s}_1 = 1,01$, $\hat{A}_1 = -0,13$ и $\hat{E}_1 = 0,04$, т. е. его можно признать (см. п. III.6) нормальным. Случайные колебания p_2 , связанные с временем, должны были иметь равномерное распределение на отрезке $[-1, +1]$ с $\bar{p}_2 = 0$, $s_2 = [1 - (-1)] / \sqrt{12} = 0,577$; $A_2 = 0$; $E_2 = -1,2$, а у генерированного распределения $\bar{p}_2 = -0,03$, $s_2 = 0,592$, $\hat{A}_2 = 0,09$, $\hat{E}_2 = -1,22$, т. е. его также можно признать соответствующим требованиям.

Целесообразно проверить устойчивость технологических решений, отвечающих требованию «обеспечить прочность бетона при распалубке изделий $R_{\phi} \leq 2$ МПа», т. е. $Y = 693$. Из множества решений, соответствующих этому требованию и лежащих на изолинии 693 (рис. IV.2, в); наиболее интересны два: в точке А ($x_1 = \Omega_1^A = +1$, $x_2 = \Omega_2^A = -0,25$ или $T_{\phi} = 70^{\circ} \text{C}$, $\tau_{\phi} = 26,2$ мин), где результат достигается за самое короткое время, и в точке В ($x_1 = \Omega_1^B = -0,04$, $x_2 = \Omega_2^B = 0,49$ или $T_{\phi} = 44^{\circ} \text{C}$, $\tau_{\phi} = 37,4$ мин), где результат достигается при самой низкой температуре смеси.

Подстановка в (IV.10) сформированных по (IV.11) случайных величин

$$x_1^{(A)} = 1 + 0,1p_1; \quad x_2^{(A)} = -0,25 + 0,2p_2; \quad (\text{IV.12})$$

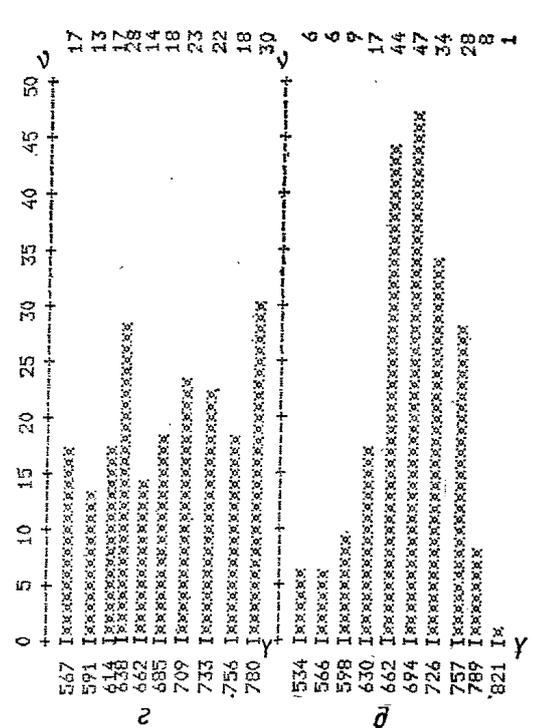
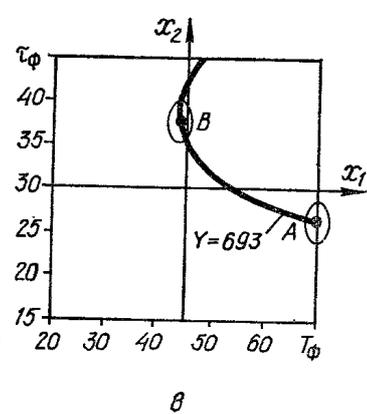
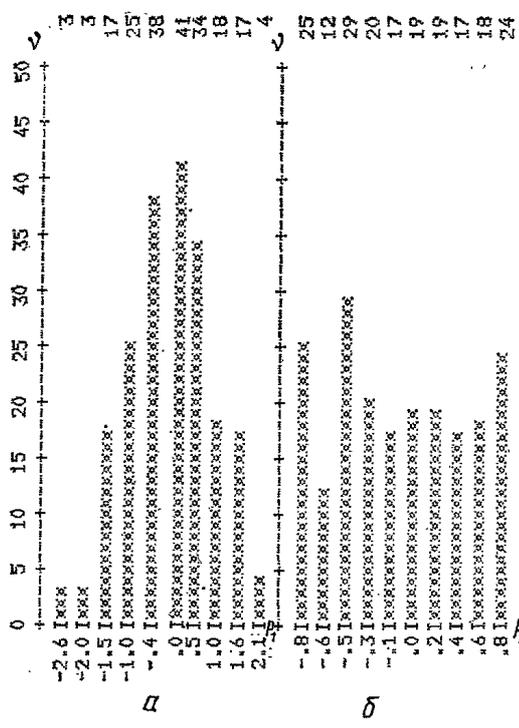
$$x_1^{(B)} = -0,04 + 0,1p_1; \quad x_2^{(B)} = 0,49 + 0,2p_2 \quad (\text{IV.13})$$

приводит к двум «игровым» моделям:

$$Y^{(A)} = 693 - 4p_1 - 2p_1^2 + 116p_2 - 16p_2^2 + 3p_1p_2, \quad (\text{IV.14})$$

$$Y^{(B)} = 693 + 57p_1 - 2p_1^2 + 0p_2 - 16p_2^2 + 3p_1p_2. \quad (\text{IV.15})$$

Результаты 200 проигрышей с помощью генерированных случайных чисел p_1 (рис. IV.2, а) и p_2 (рис. IV.2, б) по моделям (IV.14) и (IV.15) дали две



совокупности, гистограммы которых показаны на рис. IV.2, *г*, *д* (v — «эмпирические» частоты, %).

В точке *A* ($\bar{Y} = 685$, $s = 68$, $A = -0,12$, $E = -1,22$) получено почти равномерное распределение выхода Y с коэффициентом вариации $v^{(A)} = 10\%$, а в точке *B* ($\bar{Y} = 687$, $s = 58$, $A = -0,45$, $E = 0,30$) — левоасимметричное с коэффициентом вариации $v^{(B)} = 8\%$. Поскольку оба распределения отличны от нормального, то лучше сравнивать квантили распределений (III.49), (III.89), соответствующие заданной степени риска, например $\alpha = 0,1$. Для такого риска минимальные значения Y_α соответствуют 20-му результату в генерированных методом Монте-Карло совокупностях, отраженных гистограммами на рис. IV.2, *г*, и *д*. Для точки *A* минимально допустимое значение $Y_{0,1}^{(A)} = 582$, что соответствует $R_\Phi = e^{0,001Y} = 1,79$ МПа; для точки *B* $Y_{0,1}^{(B)} = 607$ или $R_\Phi = 1,83$ МПа, следовательно, технологическое решение *B*, обеспечивающее пониженную температуру формовочной смеси, более устойчиво к заданным возмущениям.

При имитационном моделировании вообще, а при реализации метода Монте-Карло на экспериментально-статистических моделях в частности целесообразно использовать все идеи и методы математической теории эксперимента, которые были рассмотрены в гл. III, что позволяет резко уменьшить число вариантов, а в ряде задач получить простые полиномиальные описания сложных вероятностных ситуаций (см., например, в [4] задачу о резерве транспорта при производстве товарного бетона на БСЗ).

При планировании имитационного эксперимента в качестве переменных могут выступать не только рецептурно-технологические факторы, изменяющие координаты точек Ω_i , но и характеристики диапазонов ω_i варьирования случайных величин p_i , а также сами законы распределения $f(P_i)$ этих величин.

Опыт решения задач анализа устойчивости рецептурно-технологических решений [8 и др.] методом Монте-Карло показывает эффективность и полезность такого подхода, однако его методическое и программное обеспечение пока далеки от совершенства.

IV.4. Метод Монте-Карло при моделировании структуры и свойств композиционных материалов

Реальные искусственные строительные материалы неоднородны по структуре. Так, бетоны на уровне *макроструктуры* состоят из зерен крупного заполнителя (щебня) разного размера и формы и цементно-песчаной матрицы. Характерный дефект структуры — трещины на границе раздела матрицы и заполнителя, в частности, из-за седиментационных явлений при твердении бетона. На следующем уровне (называемом *мезоструктурой*) легко обнаруживается, что цементно-песчаная матрица также неоднородна — она состоит из зерен крупнодисперсного наполнителя (песка) и матрицы из цементного камня, в которой распределены некоторые дефекты в виде трещин и крупных пор. На уровне *микроструктуры* сама по себе цементная матрица также неоднородна — она состоит из затвердевшего цементного геля с твердыми включениями в виде зерен це-

ментного клинкера и тонкомолотого наполнителя. Характерные элементы структуры на этом уровне — поры различной конфигурации и технологические трещины. В зависимости от целей исследования в структуре реальных материалов могут быть выделены как иные уровни, так и иные структурные элементы [6, 15, 25, 31 и др].

Проблема моделирования структуры и ее влияния на свойства материала является одной из самых актуальных в материаловедении, однако она пока находится в начальной стадии разрешения. Целесообразно выделить несколько базовых задач в этой проблеме. Во-первых, в структуре должны быть выявлены характерные элементы, которым придается идеализированная геометрическая форма (сфероиды, цилиндры и т. п.) с заданным законом распределения размеров и которые наделяются комплексом свойств (плотностью, модулем упругости, проводимостью и т. п.) с уровнями, распределенными по известным законам. Во-вторых, структурные элементы должны быть «расставлены» в пространстве (ограниченная плоскость, куб, сфера, изделие сложной конфигурации и т. п.) по некоторым правилам, в число которых часто входит случайное расположение i -го элемента в незанятой другими элементами области. В-третьих, должны быть установлены закономерности взаимодействия элементов структуры (как между собой, так и со средой, внешней по отношению к материалу) при изменении напряжений и деформаций, при тепломассообмене и пр. В-четвертых, необходимо составить алгоритмы пошагового изменения структурных элементов и характера взаимосвязей между ними при пошаговом изменении внешней среды. Все эти задачи, особенно третьей и четвертой групп, очень сложны и пока, вообще говоря, еще не сформулированы достаточно полно и корректно.

Некоторые представления [25, 31], связанные с моделированием структуры, поясняются на рис. IV.3.

В плоский контейнер размером $A \times A$ (без потери общности $A = 1$), у которого левая стенка и дно совпадают с осями координат, в случайном порядке «упаковываются» элементы структуры.

Для круглого зерна 1 (пора, наполнитель и т. п.) задаются три случайных числа, которые определяют случайный выбор одного из возможных радиусов $R^{(1)}$, расстояние центра зерна от левой стенки $x_1^{(1)}$ и от дна $x_2^{(1)}$. Поскольку контейнер пока пустой, то все эти числа запоминаются ЭВМ. Следующее зерно 2 считается неупакованным и не запоминается, поскольку оно пересеклось со стенкой (нарушено условие $R^{(2)} \leq x_i^{(2)} \leq 1 - R^{(2)}$). Зерно 3 укладывается в контейнер, так как оно попало на пустое место (выполнено условие $|x_i^{(3)} - x_i^{(1)}| \geq R^{(1)} + R^{(3)}$), а вот информация о зерне 4 не запоминается в ЭВМ, так как оно пересекает уже уложенное зерно 1. Повторяя процедуру n раз, можно «заполнить» контейнер, промоделировав случайную укладку зерен, радиусы которых подчиняются заданному закону распределения, и тем самым оценить, в частности, пустотность (пористость, проницаемость и т. п.) материала. Следует отметить, что существуют алгоритмы,

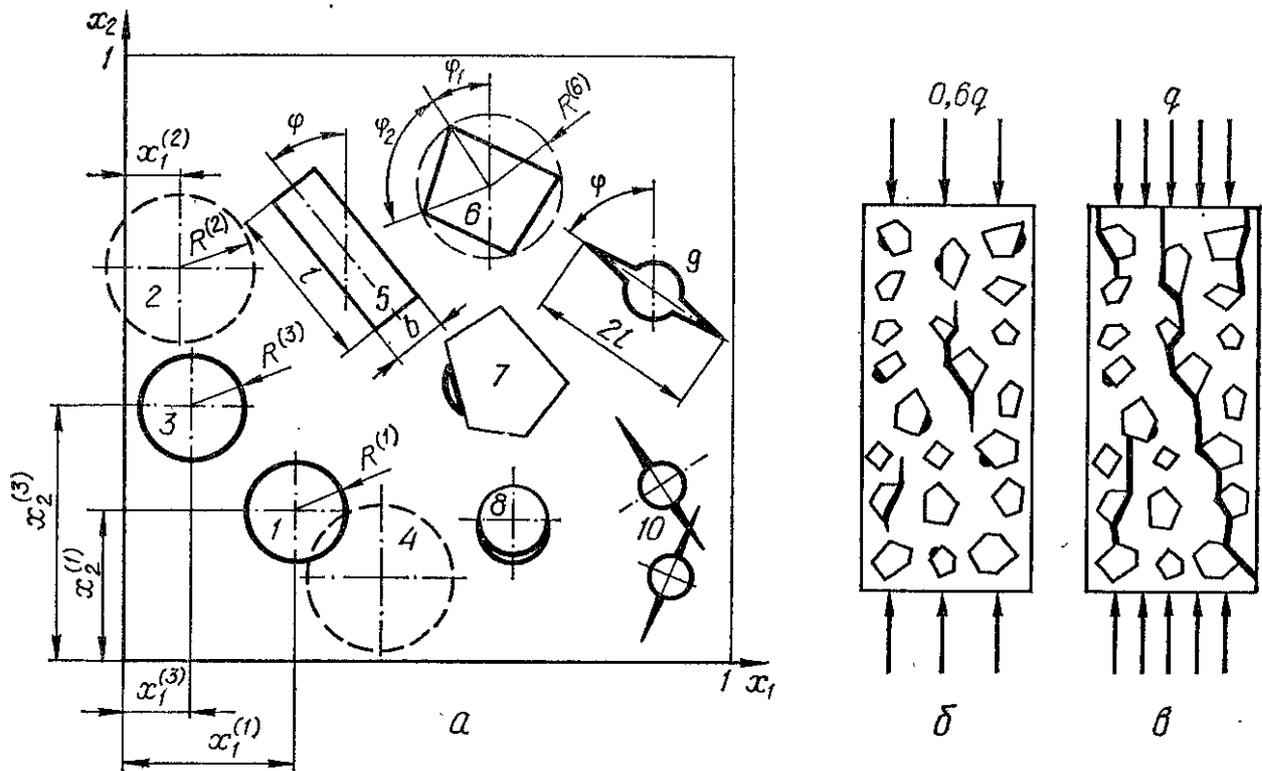


Рис. IV.3. Имитация структуры материала:

a — упаковка контейнера элементами структуры; *б, в* — изменения в структуре при изменении внешней нагрузки

«сдвигающие» зерно, попавшее на занятое место, до тех пор, пока оно не упакуется. Кроме того, у каждого зерна *1, 2, 3, ...* может быть свой случайно распределенный или связанный с размером зерна уровень физического свойства (модуль упругости, электропроводность, плотность...), что расширяет круг решаемых с помощью такого «контейнера» инженерных задач.

Структурные элементы могут иметь форму, отличную от сферической. Для фиксации прямоугольника *5* (кристалл, волокно и т. п.) кроме координат центра x_i нужно задать еще три случайные величины: размер одной из сторон l , соотношение между сторонами $l:b$, угол наклона φ к оси координат. Для фиксации многоугольника *6* (крупный наполнитель и т. п.) дополнительно задаются радиус описанной окружности $R^{(6)}$ и n координат (четыре угла φ_i) его вершин.

Важнейшими структурными элементами, особенно в задачах моделирования деформации и прочности материалов методами механики разрушения, являются дефекты — трещины в матрицах разного структурного уровня [31]. Трещины на контактах матрицы и включения (зерно *7* на рис. IV.3, *a*) распределяются в случайном порядке сначала по зернам, а потом на одну из граней выбранного зерна. Если имитируются трещины, возникающие от седиментационного расслоения смеси, то они располагаются всегда на нижней части зерна (*8* на рис. IV.3, *a*). Трещины около отверстий (пор) моделируются структурным элементом *9*, для которого обычно задается случайный угол наклона φ , а также, при необходимости, случайные радиусы поры R и длина трещины l . Допустимо такое

расположение двух пор, при котором трещины пересекаются (элемент 10 на рис. IV. 3, а), образуя непрерывную слившуюся трещину.

В зависимости от постановки задач моделирования структуры и свойств композиционных материалов могут быть распределены и другие элементы: ленты разной длины и конфигурации, каналы с узлами запираания, сети и т. п., а сам контейнер может быть объемным.

Алгоритм, описывающий закономерности поведения структуры при изменении условий внешней среды, строится, как правило, прежде всего как пошаговый процесс поведения каждого типа элементов структуры (с учетом возможного взаимодействия с контактирующими соседними элементами). Поскольку единичный элемент достаточно сильно упрощен, то для описания происходящих с ним изменений построить математическую модель не так сложно, как для структуры в целом. При этом можно использовать любые модели, но наиболее удачными, как правило, являются модели M_2 —концептуальные (см. «Введение»), в которых четко прослеживаются физические особенности процесса. Например [31], пора I в нагруженном материале описывается формулами технической механики как круговой концентратор напряжений. После того как пиковые напряжения превысят некоторый уровень и появится радиальная трещина, произойдет переход к другим формулам технической механики, описывающим деформации элемента 9. Когда длина трещины значительно превысит радиус отверстия или две соседние трещины сольются в элемент 10, то для описания процесса деформации потребуется третий комплекс формул (таблиц, графиков и т. п.).

После того как произошедшие со всеми индивидуальными элементами изменения зафиксированы в машине, они могут быть тем или иным образом интегрированы (или усреднены, или представлены, мажорантами и т. п.) и переданы для обсуждения оператору-исследователю. Если после данного шага никаких существенных перемен (границы которых заранее определены исследователем) не произошло, то будет автоматически совершен следующий шаг в изменении условий внешней среды и так последовательно (рис. IV. 3, б, в) до наступления критического состояния в структуре (разрушение, электрический пробой, сквозной проход жидкости и др.). Уровень влияния внешней среды (нагрузка, электрическое напряжение, давление жидкости и др.), при котором достигается критическое состояние, а также общий характер процесса поведения материала при воздействии внешней среды сравниваются с результатами натурального эксперимента. На основе такого сравнения и принимается решение об адекватности модели.

Преимущества вышеизложенного принципиально нового подхода к анализу структуры и свойств композиционных материалов очевидны. Необходимо подчеркнуть две особенности такого моделирования: во-первых, на любом этапе исследования в связи с появлением более сильных физических гипотез или точных формул можно некоторую часть имитационной модели заменить; во-вторых,

в ходе реализаций статистических испытаний можно выявить те сочетания структурных элементов и влияния внешней среды, при которых моделируемое явление идет по неизвестному направлению, что открывает возможности целенаправленного создания новых материалов и технологий.

IV.5. Решение рецептурно-технологических задач на ЭВМ в режиме диалога

При решении задач численными методами инженер-технолог взаимодействует с ЭВМ, оснащенной программным обеспечением. Программное обеспечение составляют: 1) программы, реализующие разные численные методы; 2) управляющие программы (операционные системы и специальные проблемно-ориентированные программные средства, разрабатываемые для решения определенного класса задач), которые организуют вычислительный процесс и обеспечивают взаимодействие с ЭВМ специалиста, решающего задачу, — пользователя ЭВМ. Это взаимодействие осуществляется в одном из двух режимов.

В *пакетном режиме* пользователь заранее определяет последовательность применяемых методов, задает их параметры, все исходные данные задачи, вводит эту информацию в ЭВМ (в пакете задания) и по завершению всех заданных расчетов получает конечные результаты. После их анализа при необходимости уточнить результат или изменить отдельные исходные данные снова выполняется полный расчет.

В *режиме диалога* (в *интерактивном режиме*) пользователь и ЭВМ задают друг другу вопросы и получают ответы. При этом алгоритм решения задачи может быть построен по сложной многоветвевой схеме, в которой переходы на ту или иную ветвь не заданы жесткой логикой, а выбираются в ходе диалога человеком. Это делает процесс решения инженерных задач более гибким и полным, позволяя учесть неформализованную информацию. Алгоритм может уточняться и пересматриваться в ходе выполнения в зависимости от текущих результатов. Если при пакетном режиме работы с ЭВМ управление вычислительным процессом осуществляется статически, то в режиме диалога этот процесс становится динамическим на основе оперативного обмена информацией между ЭВМ и пользователем через дисплей.

Пример IV.6. Фрагмент возможного диалога технолога с ЭВМ показан в табл. IV.1. Проводится анализ диссоциативно-шаговым методом (см. п. III.15, рис. III.33) технологических решений, которые обеспечивают максимум прочности бетона (см. пр. III.39) описываемой моделью (III.191).

На экране дисплея высвечиваются: 1) задаваемые технологу программой инструкции и вопросы со знаком «=» в конце фразы; 2) справа от знака равенства вводимые через клавиатуру ответы технолога; 3) получаемые в ходе решения рецептурно-технологической задачи промежуточные и окончательные результаты.

Как видно, в ходе данного диалога технолога не удовлетворила информация, полученная в строке 17, и он решил снять введенные в строках 8 и 9

Таблица IV.1. Диссоциативно-шаговая оптимизация в режиме диалога технолога с ЭВМ

Номер строки	Текст на экране дисплея
01	Вы работает с программой диссоциативно-шагового метода
02	Введите следующие данные
03	— количество факторов = 3
04	— значения коэффициентов
05	$B_0 = 19.5$ $B_1 = 5.5$ $B_{11} = 1.9$ $B_{12} = 2.0$ $B_{13} = 1.5$
06	$B_2 = 0.8$ $B_{22} = -1.8$ $B_{23} = 0.4$
07	$B_3 = 6.8$ $B_{33} = 0.6$
08	— количество нефиксируемых факторов = 1
09	— какие факторы не фиксируются = X1
10	— введите 1, если оценивается минимум
11	или -1, если оценивается максимум = -1
12	Результаты оптимизации
13	Шаг 1 — при $X_3 = +1$
14	$Y = 26.9 + 7.0X_1 + 1.9X_1X_1 + 2.0X_1X_2$
15	$+ 1.2X_2 - 1.8X_2X_2$
16	Шаг 2 — при $X_2 = 0.333 - 0.556X_1$
17	$Y = 27.1 + 7.7X_1 + 2.5X_1X_1$
18	— введите 1, если хотите изменить исходные данные
19	или 0, если расчет прекращается = 1*
20	— количество нефиксируемых факторов = 0
21	Результаты оптимизации
22	Шаг 1 — при $X_1 = +1$ $X_3 = +1$
23	$Y = 35.8 + 3.2X_2 - 1.8X_2X_2$
24	Шаг 2 — оптимальное решение
25	при $X_1 = 1.000$ $X_2 = 0.889$ $X_3 = 1.000$
26	$Y = 37.22$

* После такого ответа на экране высвечиваются для корректировки строки с исходными данными: 03, 04, 05, 06, 07, 08, 09, 10 и 11

ограничения. Поэтому, отвечая на вопрос в строках 18 и 19, он ввел «1» и изменил значение параметра в строке 8 (показано строкой 20) на «0»; в этом случае вопрос в следующей строке не ставится (она стирается с экрана). После пуска программы по скорректированным данным получено точное решение задачи в строках 25 и 26.

Диалоговая система программ для решения рецептурно-технологических задач должна быть оснащена средствами, позволяющими в интерактивном режиме осуществлять управление алгоритмами, данным, специальными подсистемами анализа решений. *Управление алгоритмами* предполагает: возможность изменения алгоритма поиска оптимума (например, смена покоординатного спуска на градиентный в тупиковой точке, см. п. П. 3); изменение процедур расчета отдельных характеристик (в частности, замену расчета среднего на оценку медианы); замену условий останова при поиске оптимального решения или корня уравнения и т. п. *Управление данными* предполагает, в частности, изменение отдельных входных экспериментальных данных уточненными, добавление новых данных, замену точек плана имитационного эксперимента и пара-

метров генерируемых распределений случайных величин, увеличение или снижение точности при поиске оптимума, изменение нормативных значений параметров и т. п. *Специальные подсистемы* в режиме диалога должны обеспечить возможность: анализировать окрестности оптимума с помощью представления на экране дисплея изолиний критерия оптимальности с учетом ошибок модели; выделять и варьировать ограничения для определения путей улучшения рецептурно-технологических решений; использовать результаты в виде удобных таблиц и графиков и т. д.

Наиболее рациональной для решения рецептурно-технологических задач на ЭВМ представляется система программ, которая включает: богатую и пополняемую библиотеку функциональных программных модулей, реализующих разные численные методы; обеспечивающие программные средства, позволяющие работать на ЭВМ и в диалоговом режиме на удобном для пользователя языке и с использованием автоматизированных пакетов программ в тех случаях, когда знания и участие технолога не могут быть использованы в вычислительном процессе.

Следует еще раз подчеркнуть, что выбор того или иного численного метода для поиска приближенного решения зависит, в частности, от двух взаимосвязанных (но не взаимозаменяемых!) причин — от класса математических моделей, описывающих инженерную ситуацию с известной полнотой и заданной точностью и от возможностей, используемых для реализации вычислительной процедуры ЭВМ (включая программное обеспечение). Диапазон методов при этом весьма велик — от простейших аналитических расчетов на микрокалькуляторах [49] до вычислительного эксперимента [60] на мощных вычислительных комплексах последнего поколения.

1.1. Элементы линейной алгебры

1. Определённый набор (совокупность, комплект, система) объектов, рассматриваемый как одно целое, называется *множеством* S , а его составляющие — *элементами* множества. Множество, заданное перечислением объектов a, b, c, \dots , обозначается в виде $S = \{a, b, c, \dots\}$; если же имеется правило, по которому определяется принадлежность объекта a множеству, то это множество записывается как $S = \{a: \dots\}$ или $S = \{a | \dots\}$, где в правой части формулируются условия, которым должен удовлетворять объект, чтобы войти в данное множество S , т. е. принадлежать ему: $a \in S$ (\in — знак принадлежности). В множестве S могут быть выделены *подмножества* S_1, S_2, \dots , каждое из которых формируется по тем или иным признакам объектов, образующих S . Все элементы подмножества S_1 входят (знак \subset или \subseteq) в множество S , т. е. $S_1 \subset S$.

○ **Пример к 1.** Результаты N контрольных испытаний бетона на прочность R_i , зафиксированные в лабораторном журнале за месяц, могут рассматриваться как единая совокупность из N чисел, характеризующая качество работы предприятия, т. е. как множество $S = \{R_1, \dots, R_N\}$. Среди результатов испытаний есть данные, по которым бетон бракуется как не набравший заданной прочности $R_{\text{тр}} = 28$ МПа. Результаты, относящиеся к бракованным бетонам, образуют подмножество $S_1 = \{R : R < 28\} = \{R | R < 28\}$; частный результат $R_a = 30$ МПа входит в множество S (т. е. $R_a = 30 \in S$), но не входит в подмножество S_1 (т. е. $R_a \notin S_1$). ●

2. Положение какого-либо элемента (например, точки) в множестве S (например, на плоскости) определяется с помощью специальных величин, называемых *координатами*. Для выявления отношений между ними совокупность координат организуется в *систему координат*, причем координаты взаимно однозначно соответствуют элементам множества S . В обширном и практически важном классе задач, связанных с множеством чисел, положение объекта оценивается относительно таких подмножеств, как прямые линии; такая система из k координат организует линейное k -мерное *координатное пространство* R^k . Прямоугольная система координат в евклидовом пространстве называется *декартовой*.

○ **Пример к 2.** А. Если начало декартовых координат поместить в левый нижний угол прямоугольной панели, то координаты ее центра тяжести в пространстве R^3 могут быть записаны: по длине панели $L = X_1 = 150$ см, по ширине $B = X_2 = 30$ см, по высоте $H = X_3 = 120$ см.

Б. Если технологический регламент на изготовление панели состоит из пяти параметров: расход цемента Π , кг; расход воды B , дм^3 , расход заполнителя Z , кг; время вибрирования смеси τ , мин, и температура ТВО T °С, то реализация технологии для данной панели ($\Pi = 150$ кг, $B = 60$ дм^3 , $Z = 500$ кг, $\tau = 3$ мин, $T = 85$ °С) дает координаты точки в пятимерном пространстве R^5 . ●

* Приложение составлено на основе ряда учебных пособий для студентов высших технических учебных заведений и справочников, в том числе [3, 12, 16, 21, 32, 35, 38, 49, 56, 60].

3. Упорядоченное множество действительных чисел a_1, a_2, \dots, a_k называется k -мерным вектором \mathbf{a} . Он указывает положение точки a по отношению к началу координат. Величина a_i задает i -ю компоненту вектора \mathbf{a} . Если все $a_i = 0$, то вектор называется нулевым. Сложение в пространстве R^k векторов \mathbf{a} и \mathbf{b} дает новый вектор \mathbf{c} с компонентами

$$* \quad c_i = a_i + b_i, \quad (\text{П.1})$$

причем справедливы свойства коммутативности ($\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$) и ассоциативности $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$. Умножение вектора \mathbf{a} на скаляр (число) p соответствует увеличению всех компонент a_i в p раз: $\mathbf{b} = p\mathbf{a} = (pa_1, \dots, pa_i, \dots, pa_k)$. Скалярное произведение двух k -мерных векторов есть

$$\text{скаляр } q = (\mathbf{a}\mathbf{b}) = a_1b_1 + \dots + a_ib_i + \dots + a_kb_k = \sum_{i=1}^k a_ib_i. \quad (\text{П.2})$$

Длиной вектора \mathbf{a} в k -мерном пространстве является его квадратичная норма, определяемая как число

$$\|\mathbf{a}\| = \sqrt{\mathbf{a}\mathbf{a}}. \quad (\text{П.3})$$

○ Пример к 3. А. Координаты центра тяжести панели образуют трехмерный вектор $\mathbf{a} = (150, 30, 120)$. Для уменьшенной в десять раз геометрической модели этой панели вектор центра тяжести $\mathbf{b} = p\mathbf{a} = 0,1\mathbf{a} = (15, 3, 12)$.

Норма вектора \mathbf{a} , или расстояние ρ , см, от угла панели до ее центра тяжести согласно (П.2) и (П.3),

$$\|\mathbf{a}\| = \rho = \sqrt{150^2 + 30^2 + 120^2} = 194,4.$$

Б. В пятимерном технологическом пространстве Ц, В, З, т, Т одно технологическое решение дает вектор $\mathbf{a} = (150, 60, 500, 3, 85)$, а другое (возникшее, например, в результате применения химических добавок) — $\mathbf{b} = (140, 55, 515, 3, 80)$. Разность их $\mathbf{b} - \mathbf{a}$ даст новый пятимерный вектор $\mathbf{c} = (-10, -5, +15, 0, -5)$, характеризующий изменение (экономия) ресурсов на изготовление панели за счет принятия нового технологического решения.

В Один вектор $\mathbf{a} = (600, 300, 50, 50)$ характеризует расход, кг, четырех компонентов на изготовление 1 т полимерного герметизирующего материала, а другой $\mathbf{b} = (0, 3; 1; 2; 0,1)$ — цены исходных материалов, руб/кг. Произведение векторов $\mathbf{c} = \mathbf{a}\mathbf{b}$

$$c = 600 \cdot 0,3 + 300 \cdot 1 + 50 \cdot 2 + 50 \cdot 0,1 = 585 \text{ руб.}$$

дает стоимость материалов на 1 т готовой продукции. ●

4. Система векторов $\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n$ линейно зависима, если существуют такие отличные от нуля скаляры p_i , при которых возникает равенство

$$\sum_{i=1}^n p_i \mathbf{a}_i = 0. \text{ В этом случае один из векторов можно представить как}$$

линейную комбинацию остальных.

Множество всех k -мерных векторов называется k -мерным евклидовым, или векторным, пространством V_k . Векторное пространство V_k натянуто на некоторую систему векторов, если каждый вектор из V_k можно представить в виде линейной комбинации векторов этой системы. Базис векторного пространства есть система линейно независимых векторов, на которую натянуто пространство V_k .

○ Пример к 4. Два вектора $\mathbf{a} = (150, 30, 120)$ и $\mathbf{b} = (15, 3, 12)$ линейно зависимы, так как выполняется равенство $1\mathbf{a} - 10\mathbf{b} = 1(150, 30, 120) - 10 \times (15, 3, 12) = (150 - 150), (30 - 30), (120 - 120) = (0, 0, 0)$. Действительно, в п. 1.3 было определено, что \mathbf{b} получено как $0,1\mathbf{a}$. ●

5. Совокупность $N = mn$ чисел a_{ij} , расположенных в прямоугольной таблице, имеющей m строк и n столбцов, называется *матрицей* A размером $m \times n$:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}, \quad (\text{П.4})$$

Элементы матрицы a_{ij} ($i = \overline{1, m}$; $j = \overline{1, n}$), находящиеся в j -м столбце, образуют вектор-столбец A_j размером $m \times 1$

$$A_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \cdots \\ a_{ij} \\ \cdots \\ a_{mj} \end{pmatrix}, \quad (\text{П.5})$$

а элементы a_{ij} , находящиеся в i -й строке, — вектор-строку A_i размером $1 \times n$

$$A_i = (a_{i1}, a_{i2}, \dots, a_{ij}, \dots, a_{in}). \quad (\text{П.6})$$

Элементы a_{ij} при $i = j$ называются *диагональными элементами* (a_{ii}), остальные элементы a_{ij} ($i \neq j$) — *внедиагональными*. Совокупность некоторых элементов a_{ij} , для которой $m' \leq m$ и $n' \leq n$, образует *подматрицу* A' размером $m' \times n'$ матрицы A .

Если $m = n$, то матрица A называется *квадратной порядка n* (размером $n \times n$), в противном случае — *прямоугольной*. Квадратная матрица, у которой $a_{ij} = 0$, а все диагональные элементы $a_{11}, \dots, a_{ii}, \dots, a_{nn}$ отличны от нуля, называется *диагональной*. Если все $a_{ii} = 1$, то такая диагональная матрица называется *единичной* и обозначается E .

○ **Пример к 5.** В комбинат входят три завода, из которых в 1987 г. выпущены железобетонные изделия пяти видов (табл. П.1). Ассортимент комбината можно представить матрицей A размером 3×5 , в которой есть вектор-столбец A_2 , характеризующий выпуск стеновых панелей, и вектор-строка A_3 , характеризующая выпуск изделий на заводе № 3; из матрицы A выделена квадратная подматрица A' порядка 3, характеризующая номенклатуру изделий, предназначенных для кассетной технологии:

$$A = \begin{pmatrix} 20 & 10 & 0 & 5 & 10 \\ 10 & 10 & 4 & 0 & 6 \\ 60 & 20 & 0 & 10 & 10 \end{pmatrix}; \quad (\text{П.7}) \quad A_2 = \begin{pmatrix} 10 \\ 10 \\ 20 \end{pmatrix}; \quad (\text{П.8})$$

$$A_3 = (60, 20, 0, 10, 10); \quad (\text{П.9}) \quad A' = \begin{pmatrix} 10 & 10 & 5 \\ 10 & 10 & 0 \\ 60 & 20 & 10 \end{pmatrix}. \quad \bullet \quad (\text{П.10})$$

6. Матрица A^T , размером $n \times m$, образуемая из исходной матрицы A размером $m \times n$ заменой строк на столбцы называется *транспонированной*. При транспонировании вектор-столбца A_j (П.5) получается вектор-строка A_i и наоборот. Матрица A называется *симметрической*, если $A^T = A$. Две матрицы A и B равны, если они имеют одинаковый порядок и равные элементы $a_{ij} = b_{ij}$ для всех i и j .

Таблица П.1. Выпуск продукции на комбинате, тыс. м³

Завод-изготовитель ($i = \overline{1, m}$)	Вид изделий ($j = \overline{1, n}$)				
	№1 Панели перекрытий	№2 Панели стеновые	№3 Колонны и балки	№4 Лестничные марши и площадки	№5 Комплектующие изделия
№1	20	10	0	5	10
№2	10	10	4	0	6
№3	60	20	0	10	10

○ Пример к 6. Матрица (П. 7) транспонируется в матрицу

$$A^T = \begin{pmatrix} 20 & 10 & 60 \\ 10 & 10 & 20 \\ 0 & 4 & 0 \\ 5 & 0 & 0 \\ 10 & 6 & 10 \end{pmatrix}, \quad (\text{П.11})$$

а вектор-столбец (П.8) — в вектор-строку

$$A_2^T = (10, 10, 20). \quad \bullet \quad (\text{П.12})$$

7. Для матриц одинакового размера **A** и **B** может быть определена их сумма **C**, причем аналогично (П.1) $c_{ij} = a_{ij} + b_{ij}$.

Произведение матриц **AB** определено только тогда, когда число столбцов первой матрицы равно числу строк второй матрицы, т. е. матрица **A** имеет размер $m \times n$, а матрица **B** — размер $n \times r$. При этом элемент c_{ij} , стоящий в матрице **C** размером $m \times r$ на пересечении i -й строки и j -го столбца,

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}. \quad (\text{П.13})$$

В общем случае произведение матриц некоммукативно, т. е. **AB** \neq **BA**, даже если **BA** существует.

Операции над матрицами обладают следующими свойствами:

- 1) $(A + B) + C = A + (B + C)$;
- 2) $A + B = B + A$;
- 3) $(A + B)C = AC + BC$;
- 4) $(AB)C = A(BC)$;
- 5) $EA = AE = A$;
- 6) $(A + B)^T = A^T + B^T$;
- 7) $(AB)^T = B^T A^T$.

○ Пример к 7. По аналогии с табл. П.1 выпуск продукции за 1988 г. характеризуется матрицей

$$A^{88} = \begin{pmatrix} 25 & 8 & 0 & 5 & 7 \\ 12 & 12 & 6 & 0 & 10 \\ 62 & 21 & 0 & 10 & 11 \end{pmatrix}. \quad (\text{П.14})$$

После ее суммирования с матрицей (П.7), обозначенной A^{87} , будет получена матрица A^Σ , характеризующая выпуск продукции за два года:

$$A^\Sigma = A^{87} + A^{88} = \begin{pmatrix} 45 & 18 & 0 & 10 & 17 \\ 22 & 22 & 10 & 0 & 16 \\ 122 & 41 & 0 & 20 & 21 \end{pmatrix}. \quad (\text{П.15})$$

Цены на 1 м³ железобетонных изделий (табл. П.1) разные: панели перекрытий — 50 руб/м³, панели стеновые — 55 руб/м³, колонны и балки — 90 руб/м³, лестничные марши и площадки — 60 руб/м³ и комплектующие дета-

ли — 75 руб/м³. Эта информация образует матрицу В размером $n \times r = 5 \times 1$ (вектор-столбец)

$$B = \begin{pmatrix} 50 \\ 55 \\ 90 \\ 60 \\ 75 \end{pmatrix}. \quad (\text{П.16})$$

Каждый элемент c_{ij} матрицы $C = A^{88} B$, рассчитанный по (П.13), имеет, в данном случае, смысл общего объема годовой продукции в денежном выражении, например для второго завода за 1988 г.

$$c_{21} = 12 \cdot 50 + 12 \cdot 55 + 6 \cdot 90 + 0 \cdot 60 + 10 \cdot 75 = 2550 \text{ тыс. руб.} \quad (\text{П.17})$$

Результирующая матрица С имеет размер $m \times r = 3 \times 1$ (вектор-столбец)

$$C = \begin{pmatrix} 2515 \\ 2550 \\ 2890 \end{pmatrix}. \quad (\text{П.18})$$

8. *Определитель* n -го порядка квадратной матрицы А размером $n \times n$ есть число (скаляр), определяемое выражением $\sum_{i=1}^n (-1)^{i+1} a_{i1} |M_{i1}|$ и обозначаемое символом

$$|A| = \det A = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}, \quad (\text{П.19})$$

где $|M_{i1}|$ есть *минор* элемента a_{i1} , или *определитель подматрицы* порядка $n - 1$, которая получается из матрицы (П.4) в результате вычеркивания (исключения) i -й строки и 1-го столбца.

Любой элемент a_{ij} квадратной матрицы А имеет свой минор $|M_{ij}|$. *Главные миноры* матрицы А — это миноры диагональных элементов a_{ii} . Величина $A_{ij} = (-1)^{i+j} |M_{ij}|$ называется *алгебраическим дополнением* элемента a_{ij} матрицы А.

Если А — матрица размером 2×2 , то

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Для вычисления определителя матрицы размером 3×3

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

можно воспользоваться следующим простым способом. Пусть матрица В получена из А путем приписывания ее 1-го и 2-го столбцов в качестве 4-го и 5-го столбцов В:

$$B = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{pmatrix}. \quad (\text{П.20})$$

Определитель матрицы А равен сумме всех произведений элементов, стоящих на диагоналях, причем слагаемое берется со знаком «+», если

оно находится на диагонали, идущей слева вниз направо, и со знаком «—», если на диагонали, идущей слева вверх направо:

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - \\ - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12}. \quad (\text{П.21})$$

Основные свойства определителей:

- 1) если все элементы столбца или строки равны нулю, то определитель также равен нулю;
- 2) значение определителя не изменяется, если строки и столбцы матрицы поменять местами;
- 3) если матрица B получена из A перестановкой двух строк или столбцов, то $|B| = -|A|$;
- 4) определитель равен нулю, если две строки (два столбца) матрицы одинаковы или если одну из них можно получить из другой поэлементным умножением на скаляр;
- 5) если столбец (строку) умножить на скаляр и затем прибавить к другому столбцу (строке) матрицы, то значение определителя не изменяется;
- 6) значение определителя изменится в α раз, если все элементы столбца (строки) матрицы умножить на скаляр α .

Следом матрицы A размером $m \times n$ называется сумма ее диагональных элементов (обозначается $\text{tr}A$). Таким образом, $\text{tr}A = a_{11} + a_{22} + \dots + a_{kk}$, где k — меньшее из чисел m и n .

○ Пример к 8. Квадратная матрица A (П.10) порядка $n = 3$ была получена из (П.7) как подматрица, описывающая номенклатуру изделий для кассетной технологии. Ее определитель (П.19) можно рассчитать по правилу П.20 — (П.21):

$$|A| = \det A = \begin{vmatrix} 20 & 10 & 5 \\ 10 & 10 & 0 \\ 60 & 20 & 10 \end{vmatrix} = 20 \cdot 10 \cdot 10 + 0 \cdot 10 \cdot 60 + 10 \cdot 20 \cdot 5 - \\ - 5 \cdot 10 \cdot 60 - 0 \cdot 20 \cdot 20 - 10 \cdot 10 \cdot 10 = -1000. \quad (\text{П.22})$$

Миноры и алгебраические дополнения к ним для матрицы A :

$$\left. \begin{array}{ll} |M_{33}| = 20 \cdot 10 - 10 \cdot 10 = 100; & A_{33} = (-1)^{3+3} \cdot 100 = 100; \\ |M_{32}| = 20 \cdot 0 - 5 \cdot 0 = -50; & A_{32} = (-1)^{3+2} (-50) = 50; \\ |M_{31}| = 10 \cdot 0 - 5 \cdot 10 = -50; & A_{31} = (-1)^{3+1} (-50) = -50; \\ |M_{23}| = 20 \cdot 20 - 6 \cdot 10 = -200; & A_{23} = (-1)^{2+3} (-200) = 200; \\ |M_{22}| = 20 \cdot 10 - 5 \cdot 60 = -100; & A_{22} = (-1)^{2+2} (-100) = -100; \\ |M_{21}| = 10 \cdot 10 - 5 \cdot 20 = 0; & A_{21} = (-1)^{2+1} (0) = 0; \\ |M_{13}| = 10 \cdot 20 - 10 \cdot 60 = -400; & A_{13} = (-1)^{1+3} (-400) = -400; \\ |M_{12}| = 10 \cdot 10 - 0 \cdot 60 = 100; & A_{12} = (-1)^{1+2} \cdot 100 = -100; \\ |M_{11}| = 10 \cdot 10 - 0 \cdot 20 = 100; & A_{11} = (-1)^{1+1} \cdot 100 = 100. \end{array} \right\} \quad (\text{П.23})$$

Главные миноры: $|M_{11}| = 100$; $|M_{22}| = -100$; $|M_{33}| = 100$. След матрицы A : $\text{tr}A = 20 + 10 + 10 = 40$. ●

9. Матрица A называется *вырожденной*, если ее определитель равен нулю.

Матрица имеет *ранг r* , если наибольший порядок ее квадратной подматрицы, определитель которой не обращается в нуль, равен r .

Квадратная матрица с ненулевым определителем называется *матрицей полного ранга*, или *невырожденной* матрицей.

Если A — невырожденная матрица n -порядка, то существует такая матрица A^{-1} , называемая *обратной* к A , для которой произведение AA^{-1} есть единичная матрица, т. е. $AA^{-1} = E$. Матрица A^{-1} является также невырожденной.

Для невырожденных квадратных матриц A, B, C размером $n \times n$ справедливы следующие положения:

$$(AB)^{-1} = B^{-1}A^{-1},$$

если $AB = AC$, то $B = C$.

Обратная матрица A^{-1} для данной невырожденной матрицы A находится по формуле

$$A^{-1} = \frac{1}{|A|} \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1i} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2i} & \dots & A_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{ni} & \dots & A_{nn} \end{pmatrix}^T = \frac{1}{|A|} \begin{pmatrix} A_{11} & A_{21} & \dots & A_{i1} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{i2} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & \dots & A_{in} & \dots & A_{nn} \end{pmatrix}. \quad (\text{П.24})$$

Процедура обращения матриц применяется при решении систем линейно независимых уравнений $AX = b$. Так как уравнения линейно независимы, то A — невырожденная матрица и $A^{-1}AX = A^{-1}b$ или $X = A^{-1}b$, где X — вектор решения системы уравнений размером $n \times 1$.

Умножение матрицы A на вектор X приводит к новому вектору AX , норма которого может сильно отличаться от нормы вектора X . Для оценки этого изменения вводится понятие нормы матрицы A

$$\|A\| = \max_X (\|AX\| / \|X\|),$$

где максимум берется по всем ненулевым векторам.

○ **Пример к 9. А.** Матрица B получена из квадратной матрицы A порядка $n = 3$ (П.10) заменой элемента $a_{31} = 60$ на $b_{31} = 40$. В этом случае 3-я строка матрицы A есть 1-я строка матрицы B , умноженная на 2. Поэтому определитель новой матрицы, согласно свойству 4 (П.8), должен быть равен нулю:

$$\det B = \begin{vmatrix} 20 & 10 & 5 \\ 10 & 10 & 0 \\ 40 & 20 & 10 \end{vmatrix} = 20 \cdot 10 \cdot 10 + 0 + 5 \cdot 10 \cdot 20 - 40 \cdot 10 \cdot 5 - 10 \cdot 10 \cdot 10 = 0,$$

следовательно, матрица B — вырожденная. Так как определитель квадратной подматрицы B 2-го порядка, полученной из B ,

$$\det B = \begin{vmatrix} 20 & 10 \\ 10 & 10 \end{vmatrix} = 20 \cdot 10 - 10 \cdot 10 = 100,$$

т. е. не равен нулю, то ранг r матрицы B равен 2.

Б. Для невырожденной матрицы A (П.10) определена, согласно (П.24), обратная матрица A^{-1} . Для ее вычисления используются значения определителя и алгебраических дополнений (П.22) и (П.23):

$$A^{-1} = \frac{1}{-1000} \begin{pmatrix} 100 & -100 & -400 \\ 0 & -100 & 200 \\ -50 & 50 & 100 \end{pmatrix} = \begin{pmatrix} -0,1 & 0,1 & 0,4 \\ 0 & 0,1 & -0,2 \\ 0,05 & -0,05 & -0,1 \end{pmatrix}. \quad (\text{П.25})$$

Нетрудно убедиться в том, что произведение матриц A (П.10) и A^{-1} (П.25) дает единичную матрицу

$$AA^{-1} = \begin{pmatrix} 20 & 10 & 5 \\ 10 & 10 & 0 \\ 60 & 20 & 10 \end{pmatrix} \begin{pmatrix} -0,1 & 0,1 & 0,4 \\ 0 & 0,1 & -0,2 \\ -0,05 & -0,05 & -0,1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad \bullet$$

10. Квадратная матрица A называется *ортогональной*, если она не вырождена и $A^T = A^{-1}$. Ортогональные матрицы обладают следующими свойствами:

1) $\det A = \pm 1$; 2) $A^T A = E$; 3) $AA^T = E$;

4) если A — ортогональная матрица, то матрицы A^{-1} и A^T также ортогональны;

5) если A и B — ортогональные матрицы, то AB — ортогональная матрица;

6) если матрица A ортогональна, то составляющие ее вектор-столбцы — ортогональные взаимно-перпендикулярные векторы.

○ Пример к 10. Матрица

$$D = \begin{pmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{pmatrix}$$

является ортогональной, так как $\det D = \cos \varphi \cos \varphi + \sin \varphi \sin \varphi = 1$. Очевидно также, что угол между векторами $d_1 = (\cos \varphi \quad -\sin \varphi)$ и $d_2 = (\sin \varphi, \cos \varphi)$ равен $(2k + 1)\pi/2$, где $k = 0, 1, 2, \dots$ ●

11. Собственным значением матрицы A размером $n \times n$ называется действительное или комплексное число λ , которое для некоторого ненулевого вектора g удовлетворяет системе алгебраических уравнений

$$(A - \lambda E)g = 0. \quad (\text{П.26})$$

Любой ненулевой вектор g , удовлетворяющий этой системе, называется собственным вектором матрицы A , соответствующим собственному значению. Условием существования нетривиального решения есть равенство нулю определителя матрицы $A - \lambda E$: $|A - \lambda E| = 0$.

Характеристическим многочленом матрицы A размером $n \times n$ называется многочлен $A - \lambda E$ степени n по переменной λ . Характеристическим уравнением матрицы A называется уравнение $|A - \lambda E| = 0$. Для матрицы A число λ является собственным значением тогда и только тогда, когда оно удовлетворяет характеристическому уравнению для A .

12. Функция $L(X) = \sum_{i=1}^k a_i X_i$, где $X = (X_1, X_2, \dots, X_k)^T$ и a_i — некоторые числа, называется линейной формой, или функционалом линейного k -мерного пространства V_k .

Функция переменных X_1, X_2, \dots, X_k называется квадратичной формой $Q(X)$, если

$$Q(X) = X^T A X = \sum_{i=1}^k \sum_{j=1}^k a_{ij} X_i X_j, \quad (\text{П.27})$$

где $A = [a_{ij}]_k^k$ и $X = (X_1, X_2, \dots, X_k)^T$. Без потери общности матрицу A можно всегда предполагать симметрической. В противном случае A следует заменить симметрической матрицей $(A + A^T)/2$; значения $Q(X)$ при этом не изменяются.

Матрица A является положительно определенной тогда и только тогда, когда значения квадратичной формы $X^T A X > 0$ для всех $X \neq 0$.

Матрица A является положительно полуопределенной тогда и только тогда, когда значения квадратичной формы $X^T A X \geq 0$ для всех X и существует такой вектор $X \neq 0$, что $X^T A X = 0$.

○ Пример к 12. А. Оптовая цена цемента марки 300 — 17 руб. за 1 т, а цемента марки 500 — 20 руб. за 1 т. Потребность в этих материалах соответственно X_1 и X_2 , т. Сумма средств, необходимая для закупки цемента обонх марок, есть линейная форма

$$L(X) = 17X_1 + 20X_2.$$

Б. Матрица

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

положительно определенная, так как значения квадратичной формы при любом ненулевом векторе X положительны. Матрица

$$B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

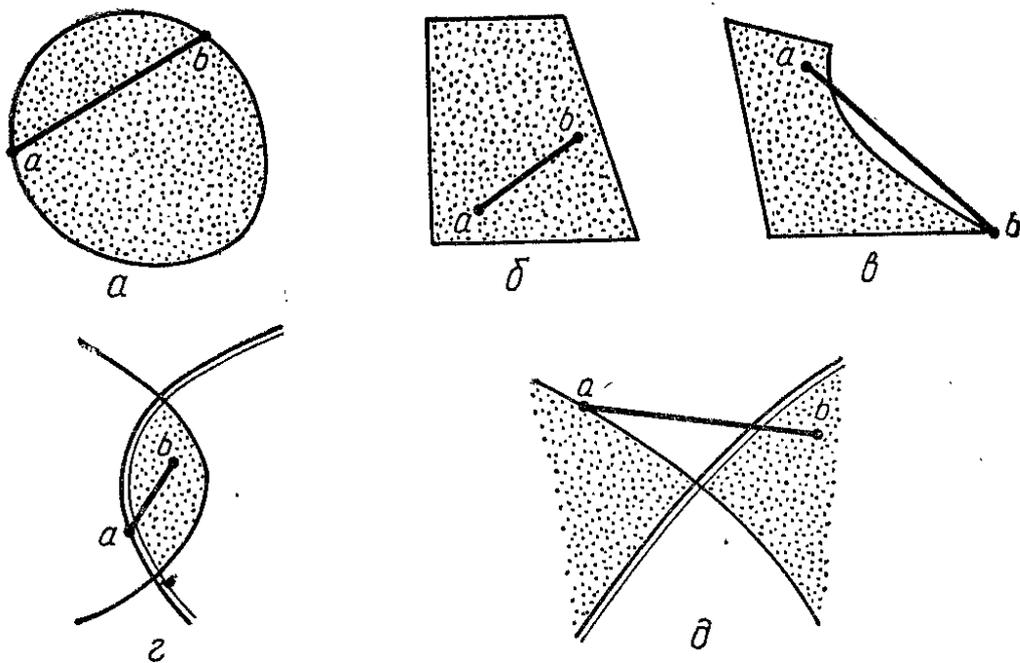


Рис. П. 1 Множества (а, б, в) и их комбинации (г, д) выпуклые (а, б, г) и невыпуклые (в и д)

положительно полуопределенная, так как при $X \neq 0$ квадратичная форма $X^T A X = 0$. ●

13. Некоторые результаты дифференциального исчисления удобно представляются в матрично-векторной форме. Если $f(X)$ — некоторая скалярная функция векторного аргумента $X = (X_1, X_2, \dots, X_k)$, то можно ввести понятие *градиента функции* как вектора, составляющие которого суть первые частные производные $f(X)$ по каждой из пространственных координат

$$\nabla(X) = \left(\frac{\partial f}{\partial X_1}, \frac{\partial f}{\partial X_2}, \dots, \frac{\partial f}{\partial X_k} \right). \quad (\text{П.28})$$

Вторые частные производные функции $f(X)$ являются элементами *матрицы Гессе*

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial X_1^2} & \frac{\partial^2 f}{\partial X_1 \partial X_2} & \dots & \frac{\partial^2 f}{\partial X_1 \partial X_k} \\ \frac{\partial^2 f}{\partial X_2 \partial X_1} & \frac{\partial^2 f}{\partial X_2^2} & \dots & \frac{\partial^2 f}{\partial X_2 \partial X_k} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial X_k \partial X_1} & \frac{\partial^2 f}{\partial X_k \partial X_2} & \dots & \frac{\partial^2 f}{\partial X_k^2} \end{pmatrix}, \quad (\text{П.29})$$

которая является симметрической матрицей.

14. Множество S в пространстве R^k называется *выпуклым*, если все точки прямой, соединяющей любые две точки a и b из S принадлежат данному множеству. Так, множества на рис. П.1, а, б — выпуклые, а множество на рис. П. 1, в не является выпуклым.

Пересечение выпуклых множеств есть выпуклое множество (рис. П.1, г)

Объединение выпуклых множеств не всегда является выпуклым (рис. П.1, д).

Выпуклой комбинацией векторов $X^{(1)}, X^{(2)}, \dots, X^{(n)}$ называется вектор $X = \alpha_1 X^{(1)} + \alpha_2 X^{(2)} + \dots + \alpha_n X^{(n)}$ при условии, что $\alpha_1 + \alpha_2 + \dots + \alpha_n = 1$ и все $\alpha_i \geq 0$ ($i = \overline{1, n}$).

15. *Вершина выпуклого множества (экстремальная точка)* — это точка, которая принадлежит данному множеству и не может располагаться на отрезке прямой, заключенном между любыми двумя точками множества.

Гиперплоскостью называется множество точек X , удовлетворяющих равенству $pX = C$ при заданном векторе $p \neq 0$ и заданной скалярной величине (уровне) C . Вектор p называется *нормалью* к гиперплоскости.

Полупространством называется множество всех точек X , удовлетворяющих неравенству $pX \leq C$ или $pX \geq C$ при заданном векторе $p \neq 0$ и заданной скалярной величине (уровню) C .

Гиперплоскость и полупространство являются выпуклыми множествами.

○ **Пример к 15. А.** Выпуклое множество $S = \{(x_1, x_2) : 0 \leq x_1 \leq 2, 0 \leq x_2 \leq 2\}$, представляющее собой квадрат, имеет четыре вершины: $(0;0)$, $(0;2)$, $(2;0)$ и $(2;2)$.

Б. В частном случае трехмерного пространства множество $T = \{(X_1, X_2, X_3) : 3X_1 + 7X_2 - 4X_3 = 1000\}$ представляет собой плоскость (здесь $p_1 = 3, p_2 = 7, p_3 = -4, C = 1000$). Полупространством в данном случае является множество точек X , расположенных по одну из сторон плоскости, включая и саму плоскость. ●

1.2. Аналитическая геометрия

1. *Аналитическая геометрия* изучает геометрические объекты (линии, поверхности и другие образы) средствами алгебры и методом координат. *Методом координат* называется способ определения одного геометрического образа относительно другого при помощи чисел.

Аналитическая геометрия *на плоскости* (в двумерном пространстве) в узком смысле — теория кривых первого (прямые), второго и более высоких порядков. Аналитическая геометрия в *трехмерном пространстве* — теория поверхностей первого (плоскость), второго и более высоких порядков. При $k \geq 3$ аналитическая геометрия решает аналогичные задачи в *гиперпространстве*, которое является математическим распространением геометрических понятий физически воспринимаемого трехмерного пространства на систему координат с числом координатных осей более трех.

Математические модели рецептурно-технологических ситуаций (где векторы формируются из расходов материалов, температур и продолжительностей этапов процесса и т. п.) тождественны описанию гиперповерхностей, поэтому для их интерпретация необходимы методы аналитической геометрии.

2. *Координаты точки* M_1 в k -мерной системе координат образуют вектор $X_{M_1} = (X_{11}, \dots, X_{i1}, \dots, X_{k1})$; его длина ρ_{M_1} (расстояние от точки M до начала координат) в соответствии с (П.3)

$$\rho_{M_1} = \|X_{M_1}\| = \sqrt{\sum_{i=1}^k X_{i1}^2}. \quad (\text{П.30})$$

Расстояние между двумя точками — M_1 с координатами X_{M_1} и M_2 с координатами X_{M_2} определяется как

$$\rho_{12} = \|X_{M_1} - X_{M_2}\| = \sqrt{\sum_{i=1}^k (X_{i1} - X_{i2})^2}.$$

○ **Пример к 2.** Трехмерное пространство образовано факторами, характеризующими время процесса τ_i , ч: X_1 — разогрев, X_2 — изотермическая выдержка, X_3 — охлаждение. Первый режим соответствует вектору $X_{1M} = (3, 10, 12)$, а второй — $X_{2M} = (2, 8, 15)$.

Длины векторов соответственно

$$\rho_1 = \sqrt{3^2 + 10^2 + 12^2} = 15,9 \text{ ч}; \quad \rho_2 = \sqrt{2^2 + 8^2 + 15^2} = 17,1 \text{ ч},$$

расстояние между точками M_1 и M_2

$$\rho_{12} = \sqrt{(3-2)^2 + (10-8)^2 + (12-15)^2} = 3,7 \text{ ч}. \quad \bullet$$

3. Общее уравнение кривых второго порядка на плоскости

$$F(x_1, x_2) = b_0 + b_1x_1 + b_2x_2 + b_{11}x_1^2 + b_{22}x_2^2 + b_{12}x_1x_2 = 0 \quad (\text{П.31})$$

определяет *линию второго порядка*, записанную в обозначениях, принятых в экспериментально-статистическом моделировании. Если квадратичные члены отсутствуют, т. е.

$$F(x_1, x_2) = b_0 + b_1x_1 + b_2x_2 = 0, \quad (\text{П.32})$$

то такое уравнение описывает *прямую*.

Из (П.32) можно получить *уравнение прямой с угловым коэффициентом* $k_1 = -b_1/b_2 = \text{tg}\alpha$:

$$x_2 = -b_0/b_2 - b_1/b_2 \cdot x_1 = k_0 + k_1x_1, \quad (\text{П.33})$$

эта прямая отсекает на оси ординат x_2 отрезок $k_0 = -b_0/b_2$, при этом угол отсчитывается от положительного направления оси x_1 до прямой в направлении против часовой стрелки.

Уравнение *семейства прямых*, проходящих через заданную точку с координатами $\{c_1; c_2\}$ (или одной прямой, если $k_1 = \text{const}$), записывается как

$$x_2 = (c_2 - k_1c_1) + k_1x_1. \quad (\text{П.34})$$

Уравнение *единственной прямой*, проходящей через две точки с координатами $\{c_1; c_2\}$ и $\{d_1; d_2\}$, имеет вид

$$\frac{x_1 - c_1}{d_1 - c_1} = \frac{x_2 - c_2}{d_2 - c_2}. \quad (\text{П.35})$$

Две прямые

$$\left. \begin{aligned} x_2 &= k_0^{(1)} + k_1^{(1)}x_1; \\ x_2 &= k_0^{(2)} + k_1^{(2)}x_1 \end{aligned} \right\} \quad (\text{П.36})$$

параллельны, если равны их угловые коэффициенты $k_1^{(1)} = k_1^{(2)}$, и они *перпендикулярны*, если произведение $k_1^{(1)}k_1^{(2)} = -1$.

Координаты *точки пересечения прямых* (П.36) определяются из равенства

$$k_0^{(1)} + k_1^{(1)}x_1 = k_0^{(2)} + k_1^{(2)}x_1. \quad (\text{П.37})$$

○ **Пример к 3.** Два нормализованных технологических фактора (x_1 — количество добавки ускорителя твердения и x_2 — то же, пассиватора коррозии) для некоторого свойства материала связаны соотношением $9x_1 + 5x_2 = 30$, из которого получается общее уравнение прямой

$$30 - 9x_1 - 5x_2 = 0, \quad (\text{П.38})$$

и уравнение с угловым коэффициентом $k_1 = -1,8$:

$$x_2^{(1)} = 6 - 1,8x_1^{(1)}. \quad (\text{П.39})$$

Для другого свойства соотношение между компонентами иное: $x_2^{(2)} = 0,5 + x_1^{(2)}$; прямые не параллельны, так как $k_1^{(1)} \neq k_1^{(2)}$; не перпендикулярны, так как $k_1^{(1)}k_1^{(2)} = (-1,8) \cdot 1 \neq -1$; точка их пересечения имеет координаты $x_1 = 2,82$ и $x_2 = 0,91$, определенные из равенств $6 - 1,8x_1 = 0,5 + x_1$ и (П.39). ●

4. Уравнение (П.31) для заданного уровня функции $f(x_1, x_2) = Y_{\text{тр}} = \text{const}$ после несложных преобразований приводится к общему уравнению *линии второго порядка*:

$$F(x_1, x_2, Y_{\text{тр}}) = b_{11}x_1^2 + 2(0,5b_{12})x_1x_2 + b_{22}x_2^2 + 2(0,5b_1)x_1 + 2(0,5b_2)x_2 + (b_0 - Y_{\text{тр}}) = 0. \quad (\text{П.40})$$

Для облегчения интерпретации математической модели в виде (П.31) уравнение (П.40) приводится к *канонической форме* (при этом сокращается число параметров модели (П.31)). Для этого выполняется преобразование декартовой системы координат: перенос начала координат в центр линии второго порядка и поворот старых осей координат (осей, в которых ставился эксперимент) в направлении главных осей кривой.

Относительно преобразования декартовой системы координат существуют *инварианты* I_i — некоторые функции коэффициентов b_i, b_{ij}, b_{ij} , которые не меняют своего значения при преобразовании координат:

$$I_1 = b_{11} + b_{22}; \quad (\text{П.41})$$

$$I_2 = \begin{vmatrix} b_{11} & 0,5b_{12} \\ 0,5b_{12} & b_{22} \end{vmatrix} = b_{11}b_{22} - 0,25b_{12}^2; \quad (\text{П.42})$$

$$I_3 = \begin{vmatrix} b_{11} & 0,5b_{12} & 0,5b_1 \\ 0,5b_{12} & b_{22} & 0,5b_2 \\ 0,5b_1 & 0,5b_2 & b_0 - Y_A \end{vmatrix} = (b_0 - Y_A) b_2 + \Delta; \quad (\text{П.43})$$

$$\Delta = 0,5b_1\delta_1 + 0,5b_2\delta_2; \quad (\text{П.44})$$

$$\delta_1 = 0,25b_2b_{12} - 0,5b_1b_{22}; \quad (\text{П.45})$$

$$\delta_2 = 0,25b_1b_{12} - 0,5b_2b_{11}. \quad (\text{П.46})$$

В зависимости от значения инварианта I_2 линии второго порядка бывают центральными или нецентральными: при $I_2 \neq 0$ линия является *центральной кривой*, при $I_2 = 0$ — *нецентральной*.

Если преобразование координат включает в себя только поворот осей, то можно использовать *полуинвариант* уравнения (П.40) относительно этого преобразования

$$P = \begin{vmatrix} b_{22} & 0,5b_2 \\ 0,5b_2 & b_0 - Y_A \end{vmatrix} + \begin{vmatrix} b_{11} & 0,5b_1 \\ 0,5b_1 & b_0 - Y_A \end{vmatrix}. \quad (\text{П.47})$$

В зависимости от значений инвариантов I_1, I_2, I_3, P уравнение (П.40) определяет одну из следующих действительных линий:

$$I_2 > 0, I_3 \neq 0, I_1 \cdot I_3 < 0 \text{ — эллипс;}$$

$$I_2 < 0, I_3 \neq 0 \text{ — гипербола;}$$

$$I_2 < 0, I_3 = 0 \text{ — пара пересекающихся прямых;}$$

$$I_2 = 0, I_3 \neq 0 \text{ — парабола;}$$

$$I_2 = 0, I_3 = 0, P < 0 \text{ — пара параллельных прямых;}$$

$$I_2 = 0, I_3 = 0, P = 0 \text{ — пара совпадающих прямых.}$$

5. Функция $Q(x) = b_{11}x_1^2 + 2(0,5b_{12})x_1x_2 + b_{22}x_2^2$ называется *характеристической квадратичной формой* уравнения (П.40), а уравнение

$$\begin{vmatrix} b_{11} - \lambda & 0,5b_{12} \\ 0,5b_{12} & b_{22} - \lambda \end{vmatrix} = 0 \quad (\text{П.48})$$

— ее *характеристическим уравнением*. Числа $\lambda_i, i = 1, 2$, называются *корнями* характеристического уравнения.

Центральные кривые имеют каноническое уравнение

$$f(x_1, x_2) - Y_S = \lambda_1\tilde{x}_1^2 + \lambda_2\tilde{x}_2^2. \quad (\text{П.49})$$

Центр кривой S имеет координаты x_{1S} и x_{2S} , в которых $f(x_1, x_2) = Y_S$:

$$x_{1S} = \delta_1/I_2; \quad x_{2S} = \delta_2/I_2. \quad (\text{П.50})$$

Если модель или уравнение (П.40) содержит ненулевой коэффициент b_{12} , то новые оси \tilde{x}_1 и \tilde{x}_2 не только перенесены в точку S , но и повернуты по отношению к осям x_1 и x_2 на угол α , который определяется из выражения

$$\operatorname{tg} 2\alpha = b_{12}/(b_{11} - b_{22}), \quad (\text{П.51})$$

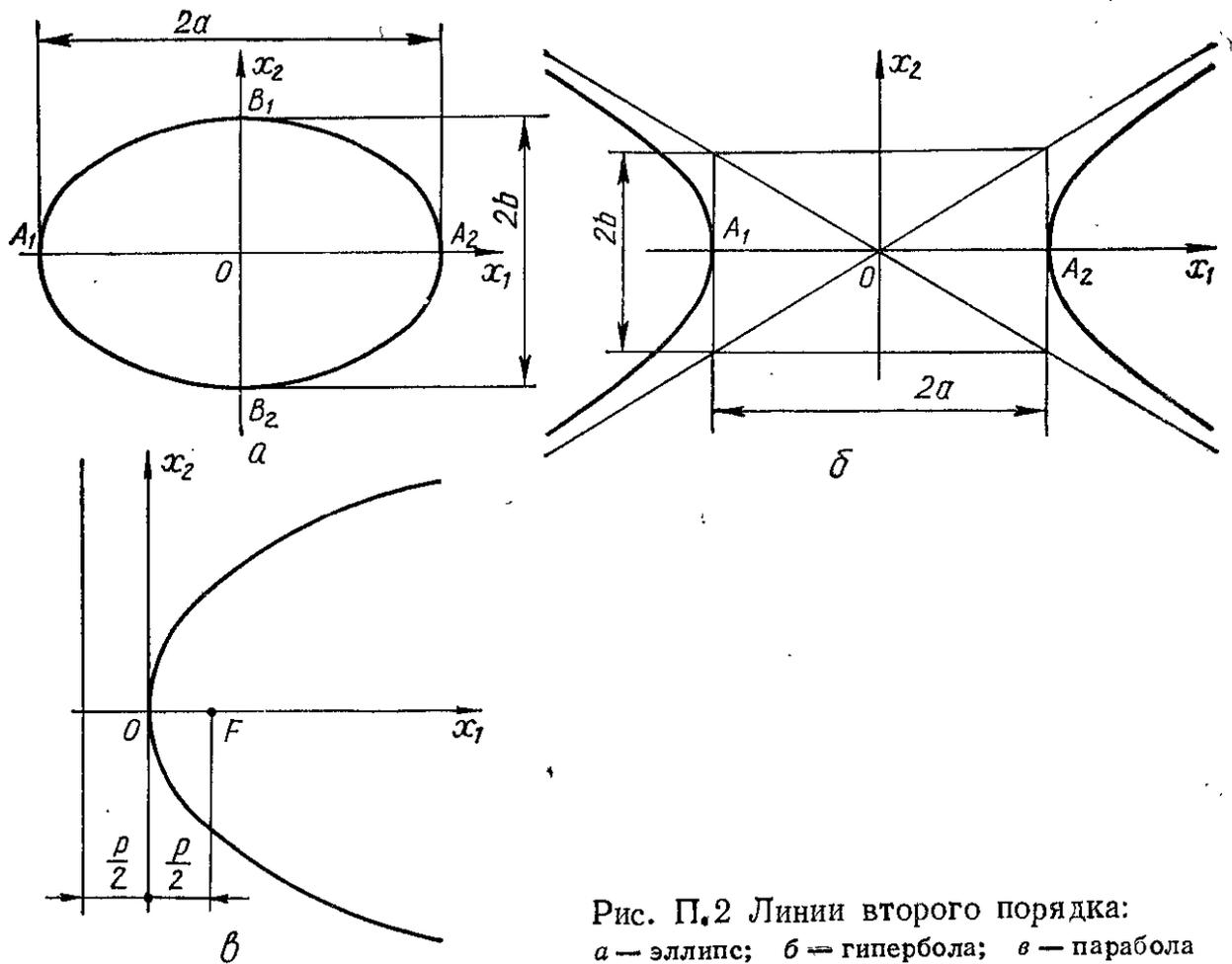


Рис. П.2 Линии второго порядка:
а — эллипс; б — гипербола; в — парабола

Примерами центральных кривых являются эллипс, окружность и гипербола.

Каноническое уравнение эллипса (рис. П.2, а)

$$\frac{\tilde{x}_1^2}{c_1^2} + \frac{\tilde{x}_2^2}{c_2^2} = 1; \quad c_1^2 = -\frac{I_3}{\lambda_2 I_2}; \quad c_2^2 = -\frac{I_3}{\lambda_1 I_2}, \quad (\text{П.52})$$

где λ_1 и λ_2 — корни характеристического уравнения (П.48), $\lambda_1 \geq \lambda_2$; $c_1 > c_2 > 0$. Точки пересечения эллипса с осями координат $A_1 \{-c_1, 0\}$, $A_2 \{c_1, 0\}$, $B_1 \{0, c_2\}$, $B_2 \{0, -c_2\}$ называются вершинами. Расстояния $2c_1$ и $2c_2$ между вершинами называются соответственно большой и малой осями эллипса.

Каноническое уравнение окружности радиуса R с центром в начале координат

$$\tilde{x}_1^2 + \tilde{x}_2^2 = R^2, \quad (\text{П.53})$$

с центром в точке $\{c_1, c_2\}$

$$(x_1 - c_1)^2 + (x_2 - c_2)^2 = R^2. \quad (\text{П.54})$$

Каноническое уравнение гиперболы (рис. П.2, б)

$$\frac{\tilde{x}_1^2}{c_1^2} - \frac{\tilde{x}_2^2}{c_2^2} = 1; \quad c_1^2 = -\frac{I_3}{\lambda_1 I_2}; \quad c_2^2 = -\frac{I_3}{\lambda_2 I_2}; \quad (\text{П.55})$$

λ_1, λ_2 — корни характеристического уравнения (П.48).

Ось абсцисс пересекает гиперболу в точках $A_1 \{-c_1, 0\}$ и $A_2 \{c_1, 0\}$, называемых вершинами гиперболы.

Расстояния c_1 и c_2 называются соответственно дейс. вительной и мнимой полуосями. Прямые $x_2 = \pm (c_1/c_2)x_1$ являются асимптотами гиперболы.

Центральная кривая — парабола (рис. П.2, в) — имеет каноническое уравнение

$$\tilde{x}_2^2 = 2p\tilde{x}_1, \quad (\text{П.56})$$

где $p = \frac{1}{I_1} \sqrt{-I_3/I_1}$ — фокальный параметр. Точку $O \{0, 0\}$ называют вершиной, точку $F \{p/2, 0\}$ — фокусом, прямую $x_1 = -p/2$ — директрисой параболы.

Главная ось параболы определяется уравнением

$$b_{11}^2 + 0,25b_{12}^2 x_1 + 0,5b_{12}(b_{11} + b_{22})x_2 + 0,25(2b_{11}b_1 + b_{12}b_2) = 0. \quad (\text{П.57})$$

Эта ось повернута по отношению к оси координат x_1 на угол α , если $b_{12} \neq 0$.

○ Пример к 4 и 5. При математическом моделировании зависимости времени сохранения пластичности бетонной смеси τ , мин, от количества микродобавки сахара (x_1) и водоцементного отношения (x_2) получена модель

$$\tau = f(x_1, x_2) = 306 - 53x_1^2 - 108x_2^2 + 27x_1. \quad (\text{П.58})$$

Целесообразно исследовать проекцию линии пересечения поверхности $f(x_1, x_2)$ с плоскостью уровня $\tau_A = 210$ мин на координатную, плоскость $\{x_1; x_2\}$. Подстановка τ_A в (П.58) и приведение уравнения линии к виду (П.40) дает:

$$-53x_1^2 - 108x_2^2 + 2(0,5 \cdot 27)x_1 + (306 - 210) = 0. \quad (\text{П.59})$$

Инварианты уравнения (П.59)

$$I_1 = b_{11} + b_{22} = -53 - 108 = -161;$$

$$I_2 = b_{11} b_{22} - 0,25 b_{12}^2 = (-53)(-108) - 0,25 \cdot 0^2 = 5724;$$

$$\delta_1 = 0,25 b_2 b_{12} - 0,5 b_1 b_{22} = 0,25 \cdot 0 \cdot 0 - 0,5 \cdot 27 \cdot (-108) = 2916;$$

$$\delta_2 = 0,25 b_1 b_{12} - 0,5 b_2 b_{11} = 0,25 \cdot 27 \cdot 0 - 0,5 \cdot 0 \cdot (-53) = 0;$$

$$\Delta = 0,5 b_1 \delta_1 + 0,5 b_2 \delta_2 = 0,5 \cdot 27 \cdot 2916 = 39366;$$

$$I_3 = (b_0 - \tau_A)I_2 + \Delta = (306 - 210) \cdot 5724 + 39366 = 582870.$$

Так как инвариант $I_2 = 5724 > 0$, то исследуемая кривая является центральной. Кроме того, $I_2 > 0$, $I_3 = 582870 \neq 0$ и произведение $I_1 I_3 = -161 \cdot 582870 = -93842070 < 0$, поэтому линия (П.59) определяет эллипс.

Корни характеристического уравнения (П.48), представленного в виде

$$\lambda^2 - I_1 \lambda + I_2 = 0 \quad (\text{П.60})$$

или

$$\lambda^2 + 164\lambda + 5724 = 0,$$

равны: $\lambda_1 = -53$, $\lambda_2 = -108$.

Коэффициент $b_{12} = 0$, следовательно, главные оси эллипса совпадают по направлению с осями координат x_1 и x_2 ($\alpha = 0,5 \arctg [b_{12}/2(b_{11} - b_{22})] = 0$). Центр эллипса имеет, согласно (П.50), следующие координаты:

$$x_{1S} = \delta_1/I_2 = 2916/5724 \approx 0,51;$$

$$x_{2S} = \delta_2/I_2 = 0/5724 = 0,$$

поэтому центр эллипса оказывается сдвинутым вправо от начала координат примерно на 0,51. Большая и малая оси эллипса, согласно (П.52), определяются из выражений:

$$a^2 = -I_3/[\lambda_2 I_2] = -582870/(-108) \cdot 5724 = 1,92;$$

$$b^2 = -I_3/[\lambda_1 I_2] = -582870/(-53) \cdot 5724 = 0,94.$$

Каноническое уравнение эллипса

$$\frac{\tilde{x}_1^2}{1,92} + \frac{\tilde{x}_2^2}{0,94} = 1. \quad (\text{П.61})$$

Полученный эллипс имеет следующие точки пересечения с осями координат

$$(c_1 = \sqrt{c_1^2} = \sqrt{1,92} \approx 1,38; c_2 = \sqrt{c_2^2} = \sqrt{0,94} \approx 0,97):$$

$$A_1 = \{-c_1 + x_{1S}; 0\} = \{-1,38 + 0,51; 0\} = \{-0,87; 0\};$$

$$A_2 = \{c_1 + x_{1S}; 0\} = \{1,38 + 0,51; 0\} = \{1,89; 0\};$$

$$B_1 = \{0; c_2 + x_{2S}\} = \{0; 0,97 + 0\} = \{0; 0,97\};$$

$$B_2 = \{0; -c_2 + x_{2S}\} = \{0; -0,97 + 0\} = \{0; -0,97\}. \bullet$$

6. Общее уравнение плоскости в декартовой системе координат $\{x_1; x_2; x_3\}$

$$b_1x_1 + b_2x_2 + b_3x_3 + b_0 - Y_A = 0, \quad (\text{П.62})$$

причем $b_1^2 + b_2^2 + b_3^2 \neq 0$.

Уравнение плоскости в трехмерном пространстве в отрезках c_1, c_2, c_3 , отсекаемых соответственно на осях координат x_1, x_2, x_3 ,

$$\frac{\tilde{x}_1}{c_1} + \frac{\tilde{x}_2}{c_2} + \frac{\tilde{x}_3}{c_3} = 1 \quad (\text{П.63})$$

при условии $c_1c_2c_3 \neq 0$.

7. Общее уравнение поверхности второго порядка в декартовой системе координат $\{x_1, x_2, x_3\}$, записанное в обозначениях, принятых в экспериментально-статистическом моделировании,

$$F(x_1, x_2, x_3, Y_{\text{тр}}) = (b_0 - Y_{\text{тр}}) + b_1x_1 + b_2x_2 + b_3x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 = 0 \quad (\text{П.64})$$

преобразуется в каноническую форму при переносе начала координат в центр поверхности второго порядка S и повороте старых осей координат x_i (в которых ставился эксперимент) в направлении главных осей поверхности \tilde{x}_i . Уравнение поверхности (П.64) по аналогии с плоскими кривыми имеет следующие инварианты:

$$\left. \begin{aligned} I_1 &= b_{11} + b_{22} + b_{33}; \\ I_2 &= \begin{vmatrix} b_{11} & 0,5b_{12} \\ 0,5b_{12} & b_{22} \end{vmatrix} + \begin{vmatrix} b_{22} & 0,5b_{23} \\ 0,5b_{23} & b_{33} \end{vmatrix} + \begin{vmatrix} b_{11} & 0,5b_{13} \\ 0,5b_{13} & b_{33} \end{vmatrix}; \\ I_3 &= \begin{vmatrix} b_{11} & 0,5b_{12} & 0,5b_{13} \\ 0,5b_{12} & b_{22} & 0,5b_{23} \\ 0,5b_{13} & 0,5b_{23} & b_{33} \end{vmatrix}; \\ I_4 &= \begin{vmatrix} b_{11} & 0,5b_{12} & 0,5b_{13} & 0,5b_1 \\ 0,5b_{12} & b_{22} & 0,5b_{23} & 0,5b_2 \\ 0,5b_{13} & 0,5b_{23} & b_{33} & 0,5b_3 \\ 0,5b_1 & 0,5b_2 & 0,5b_3 & b_0 - Y_{\text{тр}} \end{vmatrix}. \end{aligned} \right\} \quad (\text{П.65})$$

Если $I_3 \neq 0$, то поверхность называется *центральной*. Центр поверхности — точка, в которой делятся пополам проходящие через нее хорды. Координаты центра находятся из решения системы уравнений

$$\left. \begin{aligned} b_{11}x_1 + 0,5b_{12}x_2 + 0,5b_{13}x_3 + 0,5b_1 &= 0, \\ 0,5b_{12}x_1 + b_{22}x_2 + 0,5b_{23}x_3 + 0,5b_2 &= 0, \\ 0,5b_{13}x_1 + 0,5b_{23}x_2 + b_{33}x_3 + 0,5b_3 &= 0. \end{aligned} \right\} \quad (\text{П.66})$$

Если инвариант $I_3 = 0$, то поверхность имеет не центральную точку, а *линию центров* или *плоскость центров*. В этих случаях она носит название *нецентральной* поверхности.

Характеристическая квадратичная форма уравнения (П.64)

$$Q(x) = b_{11}x_1^2 + b_{12}x_1x_2 + b_{22}x_2^2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{33}x_3^2, \quad (\text{П.67})$$

ее характеристическое уравнение

$$\begin{vmatrix} b_{11} - \lambda & 0,5b_{12} & 0,5b_{13} \\ 0,5b_{12} & b_{22} - \lambda & 0,5b_{23} \\ 0,5b_{13} & 0,5b_{23} & b_{33} - \lambda \end{vmatrix} = 0. \quad (\text{П.68})$$

Корни λ_1 , λ_2 и λ_3 характеристического уравнения квадратичной формы и инварианты связаны зависимостями:

$$I_1 = \lambda_1 + \lambda_2 + \lambda_3; \quad I_2 = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3; \quad I_3 = \lambda_1\lambda_2\lambda_3. \quad (\text{П.69})$$

В зависимости от значений инвариантов I_1 , I_2 , I_3 , I_4 уравнение (П.63) определяет одну из действительных поверхностей второго порядка (рис. П.3)

8. В результате переноса начала координат в центр и поворота осей общее уравнение (П.63) преобразуется к каноническому виду

$$\lambda_1\tilde{x}_1^2 + \lambda_2\tilde{x}_2^2 + \lambda_3\tilde{x}_3^2 + I_4/I_3 = 0, \quad (\text{П.70})$$

где λ_i — корни характеристического уравнения (П.68).

В случае нецентральной поверхности при $I_4 \neq 0$ (параболоиды) общее уравнение приводится к каноническому виду

$$\lambda_1\tilde{x}_1^2 + \lambda_2\tilde{x}_2^2 \pm (2\sqrt{-I_4/I_2})\tilde{x}_3 = 0. \quad (\text{П.71})$$

В этом уравнении знак у переменной \tilde{x}_3 может быть любой, а λ_i — корни характеристического уравнения

$$\lambda^3 - I_1\lambda^2 + I_2\lambda = 0, \quad (\text{П.72})$$

в котором λ_3 всегда равно нулю.

Случай, описываемый уравнениями (П.71) и (П.72), в практике встречается реже, чем центральные изоповерхности, уравнения которых в обозначениях регрессионного анализа имеют вид

$$B_1\tilde{x}_1^2 + B_2\tilde{x}_2^2 + B_3\tilde{x}_3^2 = Y_{\text{тр}} - Y_S, \quad (\text{П.73})$$

где B_i — коэффициенты канонического уравнения; $Y_{\text{тр}}$ — выход ($Y_{\text{тр}} = \text{const}$); Y_S — выход в центральной точке; \tilde{x}_i — кодированные переменные в новой системе координат.

Переменные \tilde{x}_i связаны с нормализованными переменными в старой системе координат x_i соотношением

$$x_i = \sum_{i=1}^k K_i (\tilde{x}_i - x_{iS}).$$

Коэффициенты K_i определяются из системы уравнений, включающих корни λ_i характеристического уравнения (П.68).

1.3. Представление функций с помощью разложения Тейлора

1. Разложение функций $f(X)$ в бесконечный ряд вида $\sum_{i=0}^{\infty} a_i \Phi_i(X)$ в технических приложениях преследует обычно следующее:

а) операции над функцией $f(X)$ оказывается возможным заменить более простыми операциями над функциями $\Phi_i(X)$ или над коэффициентами a_i ;

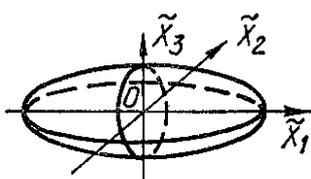
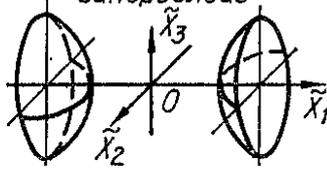
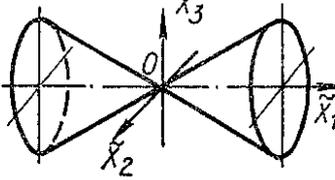
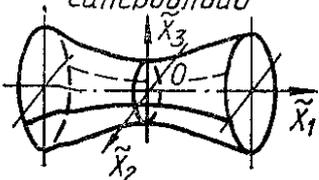
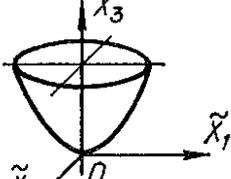
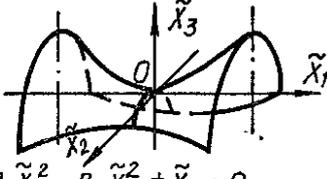
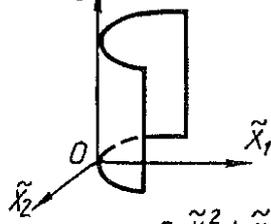
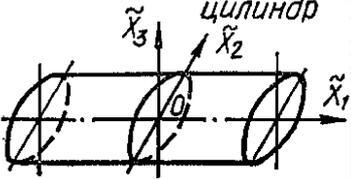
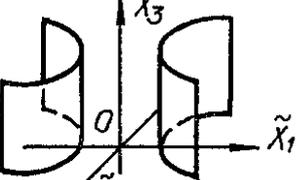
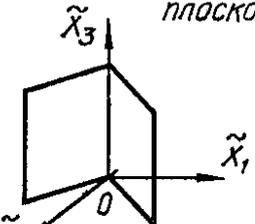
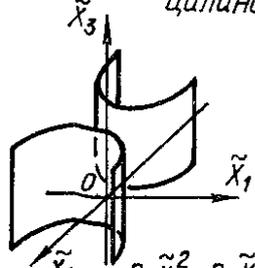
Центральные поверхности $I_3 \neq 0$	$I_4 < 0$ Эллипсоид $I_1 I_3 > 0; I_2 > 0$  $B_1 \tilde{x}_1^2 + B_2 \tilde{x}_2^2 + B_3 \tilde{x}_3^2 = 1$	$I_4 = 0$ Мнимый конус $B_1 \tilde{x}_1^2 + B_2 \tilde{x}_2^2 + B_3 \tilde{x}_3^2 = 0$	$I_4 > 0$ Мнимый эллипсоид $B_1 \tilde{x}_1^2 + B_2 \tilde{x}_2^2 + B_3 \tilde{x}_3^2 = -1$
	$I_4 < 0$ Двухполостный гиперboloид $I_1 I_3 \leq 0$ или $I_2 \leq 0$  $-B_1 \tilde{x}_1^2 + B_2 \tilde{x}_2^2 + B_3 \tilde{x}_3^2 = -1$	$I_4 = 0$ Конус  $-B_1 \tilde{x}_1^2 + B_2 \tilde{x}_2^2 + B_3 \tilde{x}_3^2 = 0$	$I_4 > 0$ Однополостный гиперboloид  $-B_1 \tilde{x}_1^2 + B_2 \tilde{x}_2^2 + B_3 \tilde{x}_3^2 = 1$
Нецентральные поверхности $I_3 = 0$	$I_4 > 0$ Эллиптический параболоид  $B_1 \tilde{x}_1^2 + B_2 \tilde{x}_2^2 \pm \tilde{x}_3 = 0$	$I_4 < 0$ Гиперболический параболоид  $B_1 \tilde{x}_1^2 - B_2 \tilde{x}_2^2 \pm \tilde{x}_3 = 0$	$I_4 = 0$ Параболический цилиндр  $B_1 \tilde{x}_1^2 \pm \tilde{x}_3 = 0$
	$I_4 = 0$ Эллиптический цилиндр  $B_2 \tilde{x}_2^2 + B_3 \tilde{x}_3^2 = 1$	$I_4 = 0$ Две мнимые плоскости $B_2 \tilde{x}_2^2 + B_3 \tilde{x}_3^2 = 0$	$I_4 = 0$ Мнимый эллиптический цилиндр $B_2 \tilde{x}_2^2 + B_3 \tilde{x}_3^2 = -1$
	$I_4 = 0$ Гиперболический цилиндр  $B_1 \tilde{x}_1^2 - B_2 \tilde{x}_2^2 = 1$	$I_4 = 0$ Две пересекающиеся плоскости  $B_1 \tilde{x}_1^2 - B_2 \tilde{x}_2^2 = 0$	$I_4 = 0$ Гиперболический цилиндр  $B_1 \tilde{x}_1^2 - B_2 \tilde{x}_2^2 = -1$
	$I_4 = 0$ Две параллельные плоскости $B_1 \tilde{x}_1^2 = +1$	$I_4 = 0$ Две совпадающие плоскости $B_1 \tilde{x}_1^2 = 0$	$I_4 = 0$ Две мнимые параллельные плоскости $B_1 \tilde{x}_1^2 = -1$

Рис. П.3. Поверхности второго порядка

б) если вид функции $f(X)$, описывающей некоторую систему, заранее неизвестен, то при постулировании априори некоторых ее свойств можно исследовать поведение системы, пользуясь представлением $f(X)$ в виде конкретных аналитических функций (в частности, полиномов).

И в том, и в другом случае следует строго вести учет возможных погрешностей $R(X)$, связанных с заменой бесконечного представления $f(X) = \sum_{i=0}^{\infty} a_i \varphi_i(X)$ конечным.

2. Если функция одной переменной $f(X)$ имеет в интервале $a \leq X \leq b$ производные до $(n+1)$ -го порядка включительно, то для нее справедлива формула Тейлора

$$f(X) = f(a) + \frac{X-a}{1!} f'(a) + \frac{(X-a)^2}{2!} f''(a) + \dots + \frac{(X-a)^n}{n!} f^{(n)}(a) + R_n(X). \quad (\text{П.74})$$

Величина $R_n(X)$ называется *остаточным членом* (в форме Лагранжа)

$$R_n = \frac{(X-a)^{n+1}}{(n+1)!} f^{(n+1)}[a + \Theta(X-a)], \quad 0 < \Theta < 1. \quad (\text{П.75})$$

Если значение остаточного члена для некоторой области значений X мало, то функцию $Y = f(X)$ можно в этой области представить в виде *многочлена* $Y = P_n(X)$ с точностью, равной значению остаточного члена $R_n(X)$.

○ **Примеры к 2.** А. Согласно формуле (П.74), функция $f(X) = e^X$ имеет следующее разложение (при $a = 0$):

$$e^X = 1 + \frac{X}{1!} + \frac{X^2}{2!} + \frac{X^3}{3!} + \dots + \frac{X^n}{n!} + \frac{X^{n+1}}{(n+1)!} e^{\Theta X}. \quad (\text{П.76})$$

Оценка точности для (П.76) по значению остаточного члена при $n = 8$ и $|X| \leq 1$

$$R_8 < \frac{3}{9!} < 10^{-5}.$$

Б. Неизвестную функцию одной переменной $Y = f(X)$ в предположении существования ее первых двух производных в интервале $a \leq X \leq b$ можно, используя формулу Тейлора (П.74), представить в виде

$$f(X) = f(a) + \frac{(X-a)}{1!} f'(a) + \frac{(X-a)^2}{2!} f''(a) + R(X-a), \quad (\text{П.77})$$

где $R(X-a)$ — остаточный член.

При $a = 0$ представление $f(X)$ имеет вид квадратичного полинома

$$f(X) = \beta_0 + \beta_1 X + \beta_{11} X^2 + \varepsilon, \quad (\text{П.78})$$

где $\beta_0 = f(0)$; $\beta_1 = f'(0)$; $\beta_{11} = f''(0)/2$; $\varepsilon = R(X)$. ● (П.79)

3. Если функция одной переменной $f(X)$ имеет производные всех порядков в интервале $a \leq X \leq b$, то n может быть сколь угодно большим. Если в рассматриваемом интервале остаточный член R_n стремится к нулю при $n \rightarrow \infty$, т. е.

$$\lim_{n \rightarrow \infty} R_n(X) = 0,$$

то функцию $f(X)$ можно разложить в ряд Тейлора

$$f(X) = f(a) + \frac{X-a}{1!} f'(a) + \dots + \frac{(X-a)^n}{n!} f^{(n)}(a) + \dots \quad (\text{П.80})$$

4. Формула Тейлора для функции $f(X_1, X_2, \dots, X_k)$ от k переменных, имеющей $(n + 1)$ частных производных в окрестности D некоторой точки a ,

$$f(X_1, X_2, \dots, X_k) = f(a_1, a_2, \dots, a_k) + \frac{(X_i - a_i)}{1!} \sum_{i=1}^k \frac{df}{\partial X_i} \Big|_{(a_1, a_2, \dots, a_k)} +$$

$$+ \frac{(X_i - a_i)(X_j - a_j)}{2!} \sum_{i=1}^k \sum_{j=1}^k \frac{\partial^2 f}{\partial X_i \partial X_j} \Big|_{(a_1, a_2, \dots, a_k)} + \dots +$$

$$+ R_n(X_1, X_2, \dots, X_k) \text{ для } (X_1, X_2, \dots, X_k) \in D. \quad (\text{П.81})$$

Остаточный член $R_n(X_1, X_2, \dots, X_k)$ удовлетворяет соотношению, аналогичному (П.75).

○ **Пример к 3 и 4.** Для решения большинства реальных технологических задач целесообразно применять полиномиальные модели порядка m , которые для k факторов записываются как приближение к неизвестным истинным моделям функционирования системы:

$$Y = f(X_1, X_2, \dots, X_k) = \alpha_0 + \sum_{i=1}^k \alpha_i X_i + \sum_{i < j} \alpha_{ij} X_i X_j +$$

$$+ \sum_{i=1}^k \alpha_{ii} X_i^2 + \dots \quad (\text{П.82})$$

Коэффициенты $\alpha_0, \alpha_i, \alpha_{ii}, \alpha_{ij}, \dots$ эквивалентны частным производным ряда Тейлора для функции k переменных, записанному по аналогии с (П.80). При этом предполагается существование всех производных в окрестности точки разложения a .

В частности, при учете погрешности ε , вносимой за счет перехода от бесконечной интерпретации функции к конечному полиному, неизвестную функцию двух переменных $Y = f(X_1, X_2)$ в предположении существования ее первых двух производных в окрестности D точки $(0; 0)$ на основании формулы Тейлора (П.81) можно представить в виде

$$f(X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \varepsilon, \quad (\text{П.83})$$

где $\beta_0 = f(0; 0); \beta_1 = \frac{\partial f}{\partial X_1}; \beta_2 = \frac{\partial f}{\partial X_2}; \beta_{11} = \frac{\partial^2 f}{\partial X_1^2};$

$$\beta_{22} = \frac{\partial^2 f}{\partial X_2^2}; \beta_{12} = \frac{\partial^2 f}{\partial X_1 \partial X_2}; \varepsilon = R(X_1, X_2). \quad \bullet \quad (\text{П.84})$$

5. Разложение Тейлора используется в задачах построения экспериментально-статистических моделей и их статистического анализа.

1.4. Основные понятия теории вероятности

1. Всякое осуществление в определенных условиях действий, направленных на обследование объекта, называется *опытом*. Опыт может иметь несколько *исходов* или *элементарных* (неразложимых) *событий* ω_i . Элементарные события несовместимы, так как никакая их пара не может произойти одновременно. Все элементарные события образуют *пространство элементарных событий* $\Omega = \{\omega_1, \omega_2, \dots, \omega_i, \dots\} = \{\omega_i\}$. Событие является *составным* (разложимым), если существуют хотя бы два элементарных события $\omega_{i.1}$ и $\omega_{i.2}$, осуществление любого из которых приводит к осуществлению составного.

Событие *достоверно*, если оно в результате данного опыта обязательно происходит, и событие *невозможно*, если в результате данного опыта оно

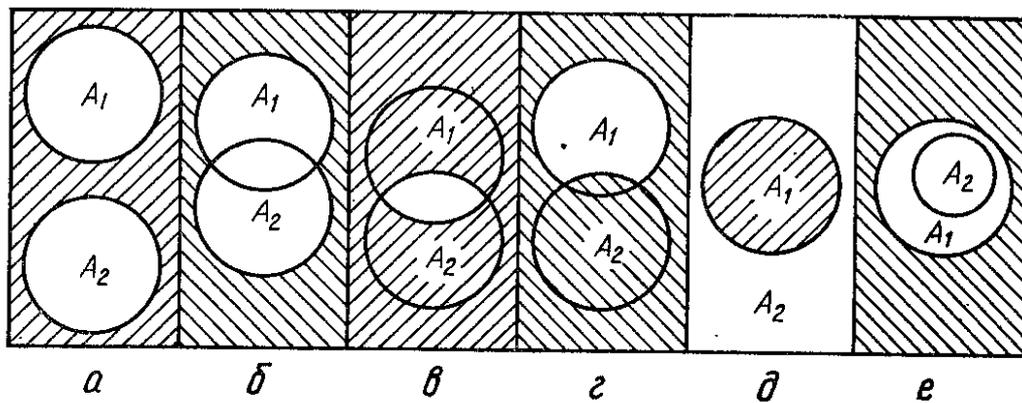


Рис. П.4. Взаимосвязь событий A_1 и A_2 :

a — события несовместны; b — объединение событий; $в$ — пересечение событий; $г$ — разность событий $A_1 - A_2$; $д$ — событие A_2 дополняет событие A_1 ; $е$ — событие A_2 содержится в A_1

никогда не произойдет. *Случайным* (или *вероятным*) событием называется событие, которое может произойти или не произойти в результате опыта.

○ **Пример к 1.** В композиционном материале есть зерна наполнителя только четырех типов: 1 — крупные кварца (КК), 2 — крупные андезита (КА), 3 — мелкие кварца (МК), 4 — мелкие андезита (МА). Опыт — извлечение из множества зерен одного зерна. Элементарных событий в пространстве Ω — четыре (ω_1 — взять зерно КК и т. д.); все они несовместны, так как зерно МК не может быть одновременно зерном КА. Факт появления крупного зерна — событие составное, так как для этого достаточно зерну быть или КК (ω_1) или КА (ω_2).

Извлечение из массы материала зерна минерального наполнителя — событие составное достоверное, а наполнителя из металла — событие невозможное, так как таких зерен нет. Тот факт, что в опыте отобрано именно зерно КК — случайное событие. ●

2. Событие A связано с данным опытом (или элементарными исходами ω_i), если по каждому элементарному исходу можно судить, произошло данное событие или нет. Событие A — определенное множество элементарных исходов ω_i соответствующего пространства элементарных событий Ω ; *достоверное событие* A формально отождествляется со всем множеством Ω ; *невозможное событие* — пустое множество (символ \emptyset), не содержащее ни одного ω_i .

События A_1 и A_2 *равные* ($A_1 = A_2$), если осуществление события A_1 влечет осуществление A_2 и наоборот.

События A_1 и A_2 *несовместные* (рис. П.4, a , где прямоугольник — множество элементарных событий Ω) или *непересекающиеся*, если наступление одного из них *исключает* появление другого.

Суммой $A_1 + A_2$ или *объединением* $A_1 \cup A_2$ двух событий (рис. П.4, b) называется событие A , которое означает осуществление хотя бы *одного* из событий — A_1 или A_2 .

Произведением $A_1 A_2$ или *пересечением* $A_1 \cap A_2$ двух событий (рис. Г.1, $в$) называется событие A , наступающее тогда и только тогда, когда произойдут *оба* события — и A_1 , и A_2 .

Разностью событий $A_1 - A_2$, или $A_1 \setminus A_2$, называется событие A (рис. П.4, $г$), состоящее в том, что происходит событие A_1 , но не происходит событие A_2 .

Дополнительным (рис. П.4, $д$) или *противоположным* к событию A_1 называется событие $A_2 = \bar{A}_1$, которое означает, что событие A_1 не происходит ($\bar{A}_1 = \Theta / A_1$).

Если событие A_1 *влечет* за собой наступление события A_2 , то событие A_2 *содержится* (рис. П.4, $е$) в событии A_1 , т. е. $A_2 \in A_1$.

○ **Пример к 2.** А. Событие A — выбор кварцевого зерна — есть сумма двух событий: появления зерна КК (событие A_1) или зерна МК (событие A_2), так как оба зерна кварцевые.

Событие A — выбор мелкого зерна кварца MK — есть произведение двух событий: появление мелкого зерна M (событие A_1), причем одновременно это зерно кварца K (событие A_2).

Б. При испытании образцов-балочек в агрессивной среде у них может разрушиться от одной до шести вершин. Событие A_1 — разрушение одной или двух вершин; дополнительное событие \bar{A}_1 — появление образца с тремя — шестью разрушенными вершинами. Если событие A_2 — появление образцов с разрушением двух или трех вершин, то разность событий $A_1 \setminus A_2$ — появление образца с одной разрушенной вершиной. Если событие A_2 — появление образца с одной разрушенной вершиной, то $A_2 \in A_1$, так как для A_1 такая ситуация предусмотрена. ●

3. *Вероятность* в современном аксиоматическом изложении определяется как числовая функция P , определенная на классе событий \mathfrak{U} , для которой: 1) \mathfrak{U} является алгеброй событий; 2) $P(A) \geq 0$ для любого события A , принадлежащего \mathfrak{U} ($A \in \mathfrak{U}$); 3) $P(\Omega) = 1$; 4) $P(A_1 + A_2) = P(A_1) + P(A_2)$, события A_1 и A_2 несовместны. Под *алгеброй событий* \mathfrak{U} понимается некоторый класс подмножеств пространств элементарных событий, для которого при любых A_1 и A_2 : а) $\Omega \in \mathfrak{U}$; б) $A_1 A_2 \in \mathfrak{U}$; в) $A_1 + A_2 \in \mathfrak{U}$; г) $A_1/A_2 \in \mathfrak{U}$. *Вероятностным пространством* называется тройка $(\Omega, \mathfrak{U}, P)$, в которой P удовлетворяет аксиомам 1—4 и, кроме того, множество \mathfrak{U} содержит счетные суммы и произведения событий.

Для *конечного* (дискретного) вероятностного пространства $\Omega = \{\omega\} = \{\omega_1, \dots, \omega_N\}$, где N — натуральное число, числа $\{p(\omega) : \omega \in \Omega\}$, удовлетворяющие условию

$$p(\omega) \geq 0; \sum_{\omega \in \Omega} p(\omega) = 1, \quad (\text{П.85})$$

называются *элементарными вероятностями*. Если элементарные вероятности одинаковы, то

$$p(\omega) \stackrel{!}{=} 1/N. \quad (\text{П.86})$$

Вероятность события $A = \{\omega_{A.1}, \dots, \omega_{A.l}, \dots, \omega_{A.k}\}$ равна сумме тех элементарных вероятностей $p(\omega)$, у которых ω входит в A :

$$P(A) = \sum_{\omega \in A} p(\omega) = \sum_{i=1}^k p(\omega_{A.i}). \quad (\text{П.87})$$

Классическое определение вероятности — вероятность случайного события A равна отношению числа k элементарных событий, при которых событие A происходит, к общему числу элементарных событий N :

$$P(A) = k/N, \quad (\text{П.88})$$

что тождественно (П.87).

Для *непрерывного* вероятностного пространства с *равномерным* распределением точек в области G вероятность события A вычисляется как отношение соответствующих площадей (или объемов) S :

$$P(A) = S(A)/S(G). \quad (\text{П.89})$$

Соотношение (П.89) называется *геометрической вероятностью*.

○ **Пример к 3. А.** Смесь зерен наполнителя образована из $k_K = 200$ зерен кварца и $k_A = 300$ зерен андезита. Зерна распределены равномерно, т. е. для всего множества зерен $N = 500$ вероятность элементарного исхода $p(\omega) = 1/500$. Вероятность извлечь из смеси зерно андезита

$$P(A) = k_A/N = 300/500 = 0,6 < 1. \quad (\text{П.90})$$

Б. На шлифе ячеистого материала размером 15×10 мм в случайном порядке расположено 36 пор радиусом 0,5 мм. Шлиф анализируется на лазерном порометре [6]; вероятность попадания неподвижного луча в пору (событие A) составляет

$$P(A) = 36\pi 0,5^2 / (15 \cdot 10) = 28,27/150 = 0,1885. \quad (\text{П.91})$$

4. Из аксиоматики теории вероятности, в частности, следует, что

а) вероятность противоположного события \bar{A}

$$P(\bar{A}) = 1 - P(A); \quad (\text{П.92})$$

б) вероятность появления A — хотя бы одного из m попарно несовместных событий A_i — равна сумме их вероятностей

$$P(A) = P(A_1) + \dots + P(A_m) = \sum_{i=1}^m P(A_m). \quad (\text{П.93})$$

○ **Пример к 4.** Вероятность появления зерна KK (см. пр. к 1) равна $P(KK) = 0,28$, а зерна крупного андезита $P(KA) = 0,18$. Вероятность появления крупного зерна $P(K) = P(KK) + P(KA) = 0,28 + 0,18 = 0,46$. Вероятность противоположного события — извлечения мелкого зерна — $P(\bar{K}) = P(M) = 1 - 0,46 = 0,54$. ●

5. Два события *независимы*, если вероятность одного из них не зависит от появления другого. Вероятность *совместного появления* двух независимых событий ($A = A_1A_2$) равна произведению вероятностей этих событий

$$P(A) = P(A_1A_2) = P(A_1)P(A_2). \quad (\text{П.94})$$

События *независимы в совокупности*, если независимы как каждое из этих событий, так и любые их сочетания (в частности, независимы не только A_i и A_{i+1} , но также $A_{i-1}A_i$ и A_{i+1}). Вероятность совместного появления нескольких событий, независимых в совокупности, определяется как

$$P(A) = P(A_1) \dots P(A_m). \quad (\text{П.95})$$

Вероятность появления *хотя бы одного* из событий A_1, \dots, A_m независимых в совокупности, равна разности между единицей и произведением вероятностей противоположных событий $\bar{A}_1, \dots, \bar{A}_m$:

$$P(A) = 1 - P(\bar{A}_1) \dots P(\bar{A}_m). \quad (\text{П.96})$$

○ **Пример к 5.** Вероятность того, что на бетоносмесительном узле № 1 в данный момент времени происходит разгрузка готовой смеси $P(A_1) = 0,25$, а на БСУ № 2 — $P(A_2) = 0,18$. Вероятность одновременной работы узлов в режиме разгрузки $P(A) = 0,25 \times 0,18 = 0,045$. Вероятность того, что хотя бы один из БСУ работает в таком режиме $P(A) = 1 - (1 - 0,25)(1 - 0,18) = 0,385$. ●

6. Два события *зависимы*, если вероятность появления одного из них (A_2) зависит от наступления или ненаступления другого (A_1).

Условной вероятностью $P(A_2/A_1)$, или $P_{A_1}(A_2)$, называется вероятность события A_2 , вычисленная в предположении, что событие A_1 уже наступило; она определяется как

$$P(A_2/A_1) = P_{A_1}(A_2) = P(A_1A_2)/P(A_1). \quad (\text{П.97})$$

Вероятность *одновременного* появления двух *зависимых* событий A_1A_2 , исходя из (П.97),

$$P(A_1A_2) = P(A_1)P(A_2/A_1). \quad (\text{П.98})$$

○ **Пример к 6.** При испытании группы образцов в агрессивных средах обнаружено обрушение хотя бы одной из вершин куба (событие A_1) с вероятностью $P(A_1) = 0,2$. При последующем импульсном ультразвуковом контроле образцов обнаружено, что обрушение углов (A_1) и превышение скорости ультразвука против норматива (событие A_2) одновременно встречаются (событие A_1A_2) с вероятностью $P(A_1A_2) = 0,16$. Вероятность того, что образец с обрушенным углом (A_1 — наступило) окажется несоответствующим нормативу по изменению скорости ультразвука составляет $P(A_2/A_1) = 0,16/0,2 = 0,8$. ●

7. Вероятность появления хотя бы одного из двух событий (совместных или несовместных) вычисляется по *общему правилу сложения вероятностей*

$$P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1 A_2). \quad (\text{П.99})$$

○ **Пример к 7.** Если в условиях предыдущего примера вероятность отклонения от норматива по ультразвуку $P(A_2) = 0,25$, то вероятность отбраковки образца хотя бы по одному признаку $P(A_1 + A_2) = 0,20 + 0,25 - 0,16 = 0,29$. ●

8. *Случайной величиной* ξ называется числовая функция от элементарного события $\omega \in \Omega$, т. е. числовое значение величины ξ зависит от случайных исходов $\xi = \xi(\omega)$.

Закон распределения случайной величины ξ *задан*, если определены вероятности $P\{X' \leq \xi \leq X''\}$ возможных событий $\{X' \leq \xi \leq X''\}$, каждое из которых означает, что ξ принимает *одно* из значений $\xi(\omega) = X$ в соответствующих пределах от X' до X'' .

Если ξ в зависимости от элементарных исходов ω принимает *конечное* число значений $\xi = X$ с соответствующими вероятностями $P_\xi(X)$, то случайная величина ξ имеет *дискретное распределение*, т. е. каждому X_i соответствует $P(X_i)$:

$$\left. \begin{aligned} \xi: X' \leq X_1, X_2, \dots, X_i, \dots, X_k \leq X''; \\ P_\xi(X) : P(X_1), P(X_2), \dots, P(X_i), \dots, P(X_k); \\ P\{X' \leq \xi \leq X''\} = \sum_{X'}^{X''} P(X_i); \sum_{-\infty}^{\infty} P(X_i) = 1. \end{aligned} \right\} \quad (\text{П.100})$$

Функция, описывающая вероятность того, что случайная величина ξ примет значение, *меньшее или равное* X , называется *функцией распределения* случайной величины:

$$F(X) = P\{\xi \leq X\}, \quad -\infty < X < +\infty. \quad (\text{П.101})$$

Функция $F(X)$ для дискретной X является *ступенчатой*, принимая в диапазоне от 0 до 1 конечное число k различных значений:

$$F(X) = \sum_{-\infty}^X P(X_i). \quad (\text{П.102})$$

Функция $F(X)$ *полностью определяет* вероятность попадания случайной величины ξ в *заданный интервал* $[\overset{\circ}{X}', \overset{\circ}{X}'']$:

$$P\{\overset{\circ}{X}' < \xi \leq \overset{\circ}{X}''\} = F(\overset{\circ}{X}'') - F(\overset{\circ}{X}'). \quad (\text{П.103})$$

○ **Пример к 8.** В задаче моделирования случайных структур материала наполнитель может быть представлен в виде многоугольников с числом вершин от $X' = 3$ до $X'' = 8$. Вероятность появления треугольников $P(3) = 0,1$, четырехугольника $P(4) = 0,15$ и т. д. задана распределением

$$\left. \begin{array}{l} \xi = X: 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \\ P_\xi(X): 0,1 \quad 0,15 \quad 0,3 \quad 0,2 \quad 0,15 \quad 0,1 \\ F(X): 0,1 \quad 0,25 \quad 0,55 \quad 0,75 \quad 0,9 \quad 1 \end{array} \right\}. \quad (\text{П.104})$$

График функции распределения $F(X)$ показан на рис. П. 5, а. Вероятность появления элементов с числом вершин от 5 до 7 в соответствии с (П.103) составляет

$$P\{\overset{\circ}{X}' = 4 < \xi \leq \overset{\circ}{X}'' = 7\} = 0,90 - 0,25 = 0,65. \quad \bullet \quad (\text{П.105})$$

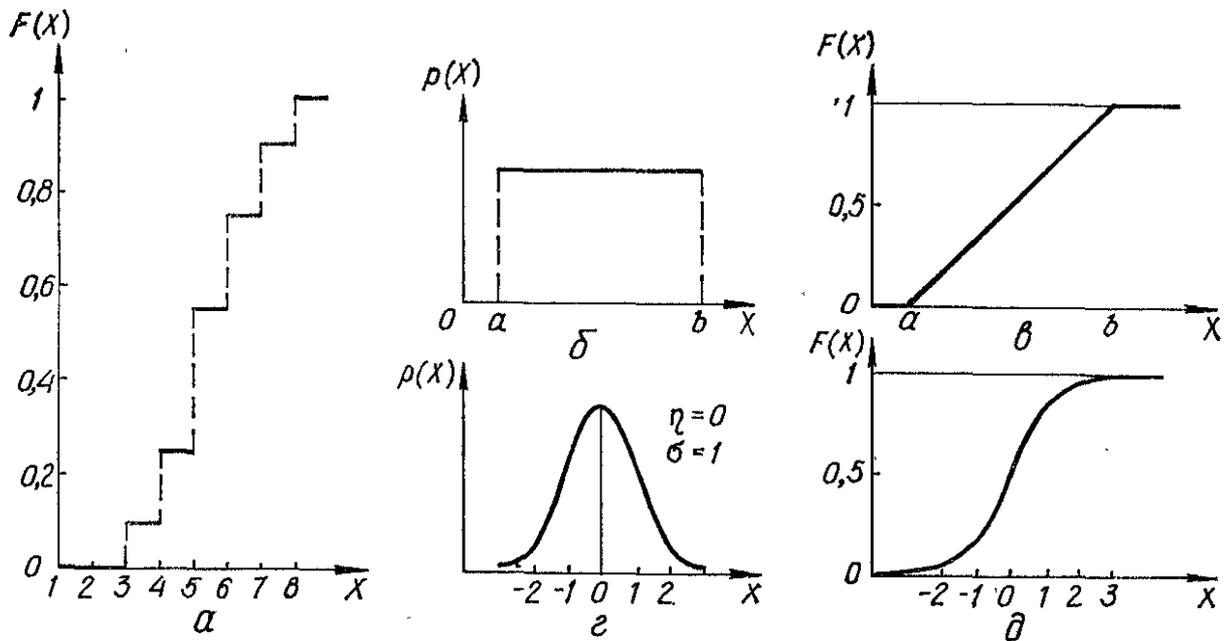


Рис. П.5. Законы распределения вероятностей:

a — функция распределения для дискретных случайных величин; *б, в* — дифференциальный; *г, д* — интегральный; *б, в* — равномерный и *г, д* — нормальные законы распределения непрерывных случайных величин

9. Случайная величина ξ имеет непрерывное распределение вероятностей, если для любых X' и X'' ($X' \leq X''$) закон распределения

$$P\{X' \leq \xi \leq X''\} = \int_{X'}^{X''} p_{\xi}(X) dX. \quad (\text{П.106})$$

где $p_{\xi}(X)$ — некоторая неотрицательная интегрируемая функция, обладающая свойством

$$\int_{-\infty}^{\infty} p_{\xi}(X) = 1. \quad (\text{П.107})$$

Функция $p_{\xi}(X)$ называется *плотностью распределения вероятностей* величины ξ . Для точки значение подынтегральной величины $p_{\xi}(X)dX$ (*элемента вероятности*) равно нулю, поэтому попадание в *точку* случайной величины ($\xi = X$) — событие *невозможное* ($P\{\xi = X\} = 0$).

Для непрерывной случайной величины ξ функция распределения $F(X)$ имеет своей производной плотность $p_{\xi}(X)$, точнее,

$$F(X) = \int_{-\infty}^X p_{\xi}(X) dX, \quad (\text{П.108})$$

причем (П.108), так же как и (п.102), обладает свойствами: 1) $0 \leq F(X) \leq 1$ для всех X ; 2) $F(X_1) \leq F(X_2)$, если $X_1 \leq X_2$; 3) $F(-\infty) = 0$ и $F(\infty) = 1$; 4) $F(X+0) = F(X)$. Вероятность попадания ξ в заданный интервал $[X', X'']$ вычисляется по (П.103).

10. *Равномерное распределение* на отрезке $[a, b]$ задается плотностью

$$p(X) = \begin{cases} \frac{1}{b-a}, & \text{если } a \leq X \leq b, \\ 0, & \text{если } X < a \text{ или } X > b, \end{cases} \quad (\text{П.109})$$

график изменения которой показан на рис. П.5, б.

Функция равномерного распределения, вычисленная по (П.108), имеет вид

$$F(X) = \begin{cases} 0, & \text{если } X < a, \\ \frac{X-a}{b-a}, & \text{если } a \leq X \leq b, \\ 1, & \text{если } X > b, \end{cases} \quad (\text{П.110})$$

ее график показан на рис. П.5, в.

11. *Нормальное (гауссовское) распределение* задается плотностью

$$p_N(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\eta)^2}{2\sigma^2}}, \quad (\text{П.111})$$

график изменения которой показан на рис. П.5, г, а соответствующая функция распределения $F(X)$ — на рис. П.5, д.

Нормальное распределение зависит от двух параметров: величины η (изменяется в диапазоне $]-\infty, +\infty[$), которая тождественна *математическому ожиданию* нормально распределенной случайной величины ξ , характеризующему центр ее группирования; величины σ^2 , которая является *дисперсией* случайной величины ξ , характеризующей степень ее рассеяния.

12. *Математическим ожиданием*, или средним значением, случайной величины $\xi = X$ называется *числовая характеристика* $E\{X\}$ распределения ее вероятностей:

для непрерывных величин

$$E\{X\} = \int_{-\infty}^{+\infty} X p_{\xi}(X) dX; \quad (\text{П.112})$$

для дискретных

$$E\{X\} = \sum_{-\infty}^{+\infty} X_i P(X_i). \quad (\text{П.113})$$

Математическое ожидание обладает следующими *свойствами*: а) для любой действительной константы C $E\{C\} = C$; б) $E\{X + C\} = E\{X\} + C$; в) $E\{C_1 X_1 + C_2 X_2\} = C_1 E\{X_1\} + C_2 E\{X_2\}$.

13. *Дисперсией* случайной величины $\xi = X$ называется *числовая характеристика* $D\{X\}$, являющаяся *мерой отклонения* случайной величины от ее математического ожидания:

$$D\{X\} = E(X - E\{X\})^2. \quad (\text{П.114})$$

Дисперсия обладает *свойствами*: а) $D\{C\} = 0$; б) $D\{C + X\} = D\{X\}$; в) $D\{CX\} = C^2 D\{X\}$, а также

$$D\{X\} = E\{X^2\} - (E\{X\})^2. \quad (\text{П.115})$$

14. Теория вероятностей тесно связана с *математической статистикой* — разделом математики, посвященным (по акад. А. Н. Колмогорову) математическим методам систематизации, обработки и использования статистических данных (т. е. сведений о числе объектов в какой-либо более или менее обширной совокупности, обладающей теми или иными признаками) для научных и практических *выводов*.

Приложение II. СПРАВОЧНЫЕ ТАБЛИЦЫ

II.1. Плотность нормированного нормального распределения

$$\varphi_N \{\varepsilon\} = (2\pi)^{-2} \exp \{-\varepsilon^2/2\}$$

ε	$\varphi_N \{\varepsilon\}$	ε	$\varphi_N \{\varepsilon\}$	ε	$\varphi_N \{\varepsilon\}$
0,0	0,399	—	—	—	—
0,1	0,397	1,1	0,218	2,1	0,044
0,2	0,391	1,2	0,194	2,2	0,036
0,3	0,381	1,3	0,171	2,3	0,028
0,4	0,368	1,4	0,150	2,4	0,022
0,5	0,352	1,5	0,130	2,5	0,018
0,6	0,333	1,6	0,111	2,6	0,014
0,7	0,312	1,7	0,094	2,7	0,010
0,8	0,290	1,8	0,079	2,8	0,008
0,9	0,266	1,9	0,066	2,9	0,006
1,0	0,242	2,0	0,054	3,0	0,004

II.2. Функция нормированного нормального распределения

$$F_N \{\varepsilon\} = \int_{-\infty}^{\varepsilon} \varphi_N \{\varepsilon\} d\varepsilon$$

ε	$F_N \{\varepsilon\}$	ε	$F_N \{\varepsilon\}$	ε	$F_N \{\varepsilon\}$
0,0	0,500	—	—	—	—
0,1	0,540	1,1	0,864	2,1	0,982
0,2	0,579	1,2	0,885	2,2	0,986
0,3	0,618	1,3	0,903	2,3	0,989
0,4	0,655	1,4	0,919	2,4	0,992
0,5	0,692	1,5	0,933	2,5	0,994
0,6	0,726	1,6	0,945	2,6	0,995
0,7	0,758	1,7	0,955	2,7	0,996
0,8	0,788	1,8	0,964	2,8	0,997
0,9	0,816	1,9	0,971	2,9	0,998
1,0	0,841	2,0	0,977	3,0	0,999

II.3. Удвоенная нормированная функция нормального распределения

$$\Phi_N \{\varepsilon\} = 2 \int_0^{\varepsilon} \varphi_N \{\varepsilon\} d\varepsilon$$

ε	$\Phi_N \{\varepsilon\}$	ε	$\Phi_N \{\varepsilon\}$	ε	$\Phi_N \{\varepsilon\}$
0,0	0,0000	—	—	—	—
0,1	0,0797	1,1	0,7287	2,1	0,9643
0,2	0,1585	1,2	0,7699	2,2	0,9722
0,3	0,2358	1,3	0,8064	2,3	0,9786

ε	$\Phi_N \{\varepsilon\}$	ε	$\Phi_N \{\varepsilon\}$	ε	$\Phi_N \{\varepsilon\}$
0,4	0,3108	1,4	0,8385	2,4	0,9836
0,5	0,3829	1,5	0,8664	2,5	0,9876
0,6	0,4515	1,6	0,8904	2,6	0,9907
0,7	0,5161	1,7	0,9109	2,7	0,9931
0,8	0,5763	1,8	0,9281	2,8	0,9949
0,9	0,6319	1,9	0,9426	2,9	0,9963
1,0	0,6827	2,0	0,9545	3,0	0,9973

II.4. Значения χ^2 при числе степеней свободы f , вероятность превышения которых α

f	χ^2 при α						
	0,99	0,95	0,90	0,50	0,10	0,05	0,01
1	0,0002	0,004	0,016	0,46	2,71	3,84	6,63
2	0,020	0,103	0,211	1,39	4,61	5,99	9,21
3	0,115	0,352	0,584	2,37	6,25	7,82	11,3
4	0,297	0,711	1,06	3,36	7,78	9,49	13,3
5	0,554	1,15	1,61	4,35	9,24	11,1	15,1
6	0,872	1,64	2,20	5,35	10,6	12,6	16,8
8	1,65	2,18	3,49	7,34	13,4	15,5	20,1
10	2,56	3,25	4,87	9,34	16,0	18,3	23,2
15	5,23	7,26	8,55	14,3	22,3	25,0	30,6
20	8,26	10,9	12,4	19,3	28,4	31,4	37,6
25	11,5	14,6	16,5	24,3	34,4	37,7	44,3
30	15,0	18,5	15,6	29,3	40,3	43,8	50,9

II.5. Абсолютные значения t , вероятность превышения которых α (двусторонний уровень)

f_1	t при α							
	0,5	0,4	0,2	0,1	0,05	0,02	0,01	0,005
1	1,000	1,376	3,078	6,314	12,706	31,821	63,657	127,32
2	0,816	1,061	1,886	2,920	4,303	6,965	9,925	14,089
3	0,765	0,978	1,638	2,353	3,182	4,541	5,841	7,453
4	0,741	0,941	1,533	2,132	2,776	3,747	4,604	5,598
5	0,727	0,920	1,476	2,015	2,571	3,365	4,032	4,773
6	0,718	0,906	1,440	1,943	2,447	3,143	3,707	4,317
7	0,711	0,896	1,415	1,895	2,365	2,998	3,499	4,029
8	0,706	0,889	1,397	1,860	2,306	2,896	3,355	3,833
9	0,703	0,883	1,383	1,833	2,262	2,821	3,250	3,690
10	0,700	0,879	1,372	1,812	2,228	2,764	3,169	3,581
11	0,697	0,876	1,363	1,796	2,201	2,718	3,106	3,500
12	0,695	0,873	1,356	1,782	2,179	2,681	3,055	3,428

f_1	t при α							
	0,5	0,4	0,2	0,1	0,05	0,02	0,01	0,005
13	0,694	0,870	1,350	1,771	2,160	2,650	3,012	3,372
14	0,692	0,868	1,345	1,761	2,145	2,624	2,977	3,326
15	0,691	0,866	1,341	1,753	2,131	2,602	2,947	3,286
16	0,690	0,865	1,337	1,746	2,120	2,583	2,921	3,252
17	0,689	0,863	1,333	1,740	2,110	2,567	2,898	3,222
18	0,688	0,862	1,330	1,734	2,101	2,552	2,878	3,197
19	0,688	0,861	1,328	1,729	2,093	2,539	2,861	3,174
20	0,687	0,860	1,325	1,725	2,086	2,528	2,845	3,153
22	0,686	0,858	1,321	1,717	2,074	2,508	2,819	3,119
24	0,685	0,857	1,318	1,711	2,064	2,492	2,797	3,090
26	0,684	0,856	1,315	1,706	2,056	2,479	2,779	3,067
28	0,683	0,855	1,313	1,701	2,048	2,467	2,763	3,047
30	0,683	0,854	1,310	1,697	2,042	2,457	2,750	3,030
60	0,679	0,848	1,296	1,671	2,000	2,390	2,660	2,915
120	0,677	0,845	1,289	1,658	1,980	2,358	2,617	2,860
∞	0,674	0,842	1,282	1,645	1,960	2,326	2,576	2,807

II.6. Значения F , вероятность превышения которых $\alpha = 0,05$
(число степеней свободы в числителе f_1)

f_2	Значения F при числе степеней свободы f_1										
	1	2	3	4	5	6	10	20	50	100	∞
1	161	200	216	225	230	234	242	248	252	253	254
2	18,51	19,00	19,16	19,25	19,30	19,33	19,40	19,41	19,47	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,79	8,66	8,58	8,56	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	5,96	5,80	5,70	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,74	4,56	4,44	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,06	3,87	3,75	3,71	3,67
8	5,32	4,46	4,07	3,84	3,69	3,58	3,35	3,15	3,03	2,98	2,93
10	4,96	4,10	3,71	3,48	3,33	3,22	2,98	2,77	2,64	2,59	2,54
15	4,54	3,68	3,29	3,06	2,90	2,79	2,54	2,33	2,18	2,12	2,07
20	4,35	3,49	3,10	2,87	2,71	2,60	2,38	2,12	1,96	1,90	1,84
40	4,08	3,23	2,84	2,61	2,45	2,34	2,08	1,84	1,66	1,59	1,51
100	3,94	3,09	2,70	2,46	2,30	2,19	1,92	1,68	1,48	1,39	1,28
200	3,89	3,04	2,65	2,41	2,26	2,14	1,87	1,62	1,42	1,32	1,19
∞	3,84	3,00	2,60	2,37	2,21	2,10	1,83	1,57	1,35	1,24	1,00

II.7. Критерий G при $\alpha = 0,05$ как предел отношения наибольшей эмпирической дисперсии к сумме K дисперсий

K	Значения G при числе степеней свободы, f_1										
	1	2	3	4	5	6	8	16	35	144	∞
2	0,998	0,975	0,939	0,906	0,877	0,853	0,816	0,734	0,660	0,581	0,500
3	0,967	0,871	0,798	0,746	0,707	0,677	0,633	0,547	0,475	0,403	0,333
4	0,906	0,768	0,684	0,629	0,590	0,560	0,518	0,437	0,372	0,309	0,250
5	0,841	0,684	0,598	0,544	0,506	0,478	0,439	0,364	0,307	0,251	0,200
6	0,781	0,616	0,532	0,480	0,445	0,418	0,381	0,313	0,261	0,212	0,167
8	0,680	0,516	0,438	0,391	0,360	0,336	0,304	0,246	0,202	0,162	0,125
10	0,602	0,445	0,373	0,331	0,303	0,282	0,254	0,203	0,165	0,131	0,100
15	0,471	0,335	0,276	0,242	0,220	0,203	0,181	0,143	0,114	0,089	0,067
20	0,389	0,270	0,220	0,192	0,174	0,160	0,142	0,110	0,088	0,068	0,050
30	0,293	0,198	0,159	0,138	0,124	0,114	0,100	0,077	0,060	0,046	0,033
40	0,237	0,158	0,126	0,108	0,097	0,089	0,078	0,060	0,046	0,035	0,025
60	0,174	0,113	0,090	0,076	0,062	0,062	0,055	0,041	0,032	0,023	0,017
∞	0	0	0	0	0	0	0	0	0	0	0

II.8. Абсолютные значения коэффициента корреляции $r\{Y_1; Y_2\}$ для различных уровней значимости α (двусторонний критерий)

f	$r\{Y_1; Y_2\}$ при уровне α			f	$r\{Y_1; Y_2\}$ при уровне α		
	0,10	0,05	0,01		0,10	0,05	0,01
1	0,988	0,997	1,000	16	0,400	0,468	0,590
2	0,900	0,950	0,990	18	0,378	0,444	0,561
3	0,805	0,878	0,959	20	0,360	0,423	0,537
4	0,729	0,811	0,917	25	0,323	0,381	0,487
5	0,669	0,754	0,874	30	0,296	0,349	0,449
6	0,621	0,707	0,834	40	0,257	0,304	0,393
8	0,549	0,632	0,765	50	0,231	0,273	0,354
10	0,497	0,576	0,708	60	0,211	0,250	0,325
12	0,457	0,532	0,661	80	0,183	0,217	0,283
14	0,426	0,497	0,623	100	0,164	0,195	0,254

II.9. Вероятности $P(\lambda)$ для критерия Колмогорова λ

λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$
0,30	1,000	0,55	0,923	0,80	0,544	1,10	0,178
0,35	1,000	0,60	0,864	0,85	0,465	1,20	0,112
0,40	0,997	0,65	0,792	0,90	0,393	1,30	0,068
0,45	0,987	0,70	0,711	0,95	0,328	1,40	0,040
0,50	0,964	0,75	0,627	1,00	0,270	1,50	0,022

II.10. Равномерно распределенные случайные числа

86515	90795	66155	66434	56558	12332	94377	57802
69186	03393	42502	99224	88955	53758	91641	18867
41686	42163	85181	38967	33181	72664	53807	00607
86522	47171	88059	89342	67248	09082	12311	90316
72587	93000	89688	78416	27589	99528	14480	50961
52452	42499	33346	83935	79130	90410	45420	77757
76773	97526	27256	66447	25731	37525	16287	66181
04825	82134	80317	75120	45904	75601	70492	10274
87113	84778	45863	24520	19976	04925	07824	76044
84754	57616	38132	64294	15218	49286	89571	42903

II.11. Планы ДФЭ, рекомендуемые для построения линейных и неполных квадратичных моделей

k	План ДФЭ	Строки матрицы	Смешанные эффекты	
			линейные	взаимодействия
3	1/2 реплика 2^{3-1}	<i>a, b, c, abc</i>	$b_1 \rightarrow \beta_1 + \beta_{23}$ $b_2 \rightarrow \beta_2 + \beta_{13}$ $b_3 \rightarrow \beta_3 + \beta_{12}$	Не определяются
4	1/2 реплика 2^{4-1}	<i>ab, ac, ad, (I), bc, bd, cd, abcd</i>	Не смешаны	$b_{12} \rightarrow \beta_{12} + \beta_{34}$ $b_{13} \rightarrow \beta_{13} + \beta_{24}$ $b_{14} \rightarrow \beta_{14} + \beta_{23}$
5	1/2 реплика 2^{5-1}	<i>(I), ab, acde, bcde, ac, bc, de, abde, ae, be, cd, abcd, ad, bd, ce, abce</i>	Не смешаны	Не смешаны
5	1/4 реплика 2^{5-2}	<i>ace, bce, ade, bde, ab, cd, abcd, (I)</i>	$b_1 \rightarrow \beta_1 + \beta_{25}$ $b_2 \rightarrow \beta_2 + \beta_{15}$ $b_3 \rightarrow \beta_3 + \beta_{45}$ $b_4 \rightarrow \beta_4 + \beta_{35}$ $b_5 \rightarrow \beta_5 + \beta_{12} + \beta_{34}$	$b_{13} \rightarrow \beta_{13} + \beta_{24}$ $b_{14} \rightarrow \beta_{14} + \beta_{23}$
6	1/8 реплика 2^{6-2}	<i>acf, ade, bce, bdf, abcd, abef, cdef, (I)</i>	$b_1 \rightarrow \beta_1 + \beta_{35} + \beta_{46}$ $b_2 \rightarrow \beta_2 + \beta_{36} + \beta_{45}$ $b_3 \rightarrow \beta_3 + \beta_{15} + \beta_{26}$ $b_4 \rightarrow \beta_4 + \beta_{16} + \beta_{25}$ $b_5 \rightarrow \beta_5 + \beta_{13} + \beta_{24}$ $b_6 \rightarrow \beta_6 + \beta_{14} + \beta_{23}$	$b_{12} \rightarrow \beta_{12} + \beta_{34} +$ $+ \beta_{56}$

И.12. Симметричные планы, рекомендуемые для построения квадратичных моделей при $k = 3..5$, и расчетные коэффициенты $T_i \cdot 10^5$ для этих планов

k	3		4		5	
Ядро плана	ПФЭ 2 ³		ПФЭ 2 ⁴		ПФЭ 2 ⁵	
Точки в середине ребер {0; ±1; 0}	6		8		10	
Точки в центре {0: 0: ... , 0}	0	1	0	1	0	1
Всего точек N	14 (B ₃)	15	24 (B ₄)	25	26	27 (Ha ₅)
T_1	40625	28889	22917	18644	16016	13805
T_2	-15625	-11111	-6250	-5082	-3516	-3030
T_3	10000	10000	5556	5556	5556	5556
T_4	50000	50000	50000	50000	50000	50000
T_5	-9375	-11111	-10417	-10734	-8984	-9091
T_6	12500	12500	6250	6250	6250	6250

И.13. Планы Рехтшафнера, рекомендуемые для построения квадратичных моделей при $k = 6$ и 7^*

$k = 6$							$k = 7$																							
Номер опыта	x_1	x_2	x_3	x_4	x_5	x_6	Номер опыта	x_1	x_2	x_3	x_4	x_5	x_6	x_7	Номер опыта	x_1	x_2	x_3	x_4	x_5	x_6	x_7								
1	-	-	-	-	-	-	18	-	-	+	-	+	-	1	-	-	-	-	-	-	-	-	18	-	+	-	-	-	+	-
2	-	+	+	+	+	+	19	-	-	+	-	-	+	2	-	+	+	+	+	+	+	+	19	-	+	-	-	-	-	+
3	+	-	+	+	+	+	20	-	-	-	+	+	-	3	+	-	+	+	+	+	+	+	20	-	-	+	+	-	-	-
4	+	+	-	+	+	+	21	-	-	-	+	-	+	4	+	+	-	+	+	+	+	+	21	-	-	+	-	+	-	-
5	+	+	+	-	+	+	22	-	-	-	-	+	+	5	+	+	+	-	+	+	+	+	22	-	-	+	-	-	+	-
6	+	+	+	+	-	+	23	+	0	0	0	0	0	6	+	+	+	+	-	+	+	+	23	-	-	+	-	-	-	+
7	+	+	+	+	+	-	24	0	+	0	0	0	0	7	+	+	+	+	+	-	+	+	24	-	-	-	+	+	-	-
8	+	+	-	-	-	-	25	0	0	+	0	0	0	8	+	+	+	+	+	+	-	-	25	-	-	-	+	-	+	-
9	+	-	+	-	-	-	26	0	0	0	+	0	0	9	+	+	-	-	-	-	-	-	26	-	-	-	+	-	-	+
10	+	-	-	+	-	-	27	0	0	0	0	+	0	10	+	-	+	-	-	-	-	-	27	-	-	-	-	+	-	+
11	+	-	-	-	+	-	28	0	0	0	0	0	+	11	+	-	-	-	+	-	-	-	28	-	-	-	-	-	+	+
12	+	-	-	-	-	+								12	+	-	-	-	-	+	-	-	30	+	0	0	0	0	0	0
13	-	+	+	-	-	-								13	+	-	-	-	-	-	+	-	31	0	+	0	0	0	0	0
14	-	+	-	+	-	-								14	+	-	-	-	-	-	+	-	32	0	0	+	0	0	0	0
15	-	+	-	-	+	-								15	-	+	+	-	-	-	-	-	33	0	0	0	+	0	0	0
16	-	+	-	-	-	+								16	-	+	-	+	-	-	-	-	34	0	0	0	0	+	0	0
17	-	-	+	+	-	-								17	-	+	-	-	+	-	-	-	35	0	0	0	0	0	+	0
																							36	0	0	0	0	0	0	+

* Оценки рассчитываются на ЭВМ по программам МНК или с помощью L-матриц [4].

Ц.14. Симплекс-решетчатые планы для построения моделей трехкомпонентных систем «состав—свойство»

Номер опыта	Модель												y
	линейная			квадратичная			неполная кубическая			кубическая			
	v ₁	v ₂	v ₃	v ₁	v ₂	v ₃	v ₁	v ₂	v ₃	v ₁	v ₂	v ₃	
1	1	0	0	1	0	0	1	0	0	1	0	0	y ₁
3	0	1	0	0	1	0	0	1	0	0	1	0	y ₂
3	0	0	1	0	0	1	0	0	1	0	0	1	y ₃
4				1/2	1/2	0	1/2	1/2	0	—	—	—	y ₁₂
5				1/2	0	1/2	1/2	0	1/2	—	—	—	y ₁₃
6				0	1/2	1/2	0	1/2	1/2	—	—	—	y ₂₃
7							1/3	1/3	1/3	1/3	1/3	1/3	y ₁₂₃
8										1/3	2/3	0	y ₁₂₂
9										2/3	1/3	0	y ₁₁₂
10										1/3	0	2/3	y ₁₃₃
11										2/3	0	1/3	y ₁₁₃
12										0	1/3	2/3	y ₂₃₃
13										0	2/3	1/3	y ₂₂₃
A _i	y _i			y _i			y _i			y _i			
A _{ij}				4y _{ij} - 2(y _i + y _j)			4y _{ij} - 2(y _i + y _j)			$\frac{9}{4}(y_{iij} + y_{iij} - y_i - y_j)$			
A ₁₂₃				$27y_{123} - 12 \sum y_{ij} + 3 \sum y_i$			$27y_{123} - 12 \sum y_{ij} + 3 \sum y_i$			$\frac{27}{4} y_{123} - \frac{9}{2} \sum y_i$			
A _{i-j}										$\frac{27}{4}(y_{iij} - y_{iji}) - \frac{9}{4}(y_i - y_j)$			

Оценки коэффициентов

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

Основная литература

1. Ахназарова С. Л., Кафаров В. В. Методы оптимизации эксперимента в химической технологии.— 2-е изд.— М.: Высш. шк., 1985.— 327 с.
2. Баженов Ю. М., Вознесенский В. А. Перспективы применения математических методов в технологии сборного железобетона. — М.: Стройиздат, 1974.— 192 с.
3. Брановицкая С. В., Медведев Р. Б., Фиалков Б. Я. Вычислительная математика в химии и химической технологии.— К.: Вища шк. Головное изд-во, 1986.— 216 с.
4. Вознесенский В. А. Статистические методы планирования эксперимента в технико-экономических исследованиях.— 2-е изд.— М.: Финансы и статистика, 1981 — 263 с.
5. Ногин В. Д., Протодьяконов И. О., Евлампиев И. И. Основы теории оптимизации.— М.: Высш. шк., 1986.— 384 с.
6. Современные методы оптимизации композиционных материалов / В. А. Вознесенский, В. Н. Выровой, В. Я. Керш и др.— К.: Будівельник, 1983.— 144 с.
7. Турчак Л. И. Основы численных методов.— М.: Наука, 1987.— 318 с.
8. ЭВМ и оптимизация композиционных материалов / В. А. Вознесенский, Т. В. Ляшенко, Я. П. Иванов, И. И. Николов.— К.: Будівельник, 1989.— 240 с.

Дополнительная литература

9. Адлер Ю. П., Маркова Е. В., Грановский Ю. В. Планирование эксперимента при поиске оптимальных условий.— 2-е изд.— М.: Наука, 1976.— 279 с.
10. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Основы моделирования и первичная обработка данных: Справочник.— М.: Финансы и статистика, 1983.— 471 с.
11. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Исследование зависимостей.— М.: Финансы и статистика, 1985.— 487 с.
12. Акулич И. Л. Математическое программирование в примерах и задачах.— М.: Высш. шк., 1986.— 319 с.
13. Анализ эффективной вязкости полимерной системы на основе модели «смесь I, смесь II, технология — свойства»/В. А. Вознесенский, Я. П. Иванов, Т. В. Ляшенко, В. И. Соломатов // Физико-химическая механика дисперсных систем.— К., 1986.— С. 122—128.
14. Антоненко Г. Я. Организация, планирование и управление предприятиями строительных изделий и конструкций.— К.: Вища шк. Головное изд-во, 1988.— 375 с.
15. Баженов Ю. М., Комар А. Г. Технология бетонных и железобетонных изделий.— М.: Стройиздат, 1984.— 672 с.
16. Барановская Г. Г., Любченко И. Н. Микрокалькуляторы в курсе высшей математики: Практикум.— К.: Вища шк. Головное изд-во, 1987.— 288 с.
17. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики.— 3-е изд.— М.: Наука, 1983.— 416 с.
18. Бондарь А. Г., Статюха Г. А., Потяженко И. А. Планирование эксперимента при оптимизации химической технологии.— К.: Вища шк. Головное изд-во, 1980.— 263 с.

19. *Бояринов А. И., Кафаров В. В.* Методы оптимизации в химической технологии.— М.: Химия, 1975.— 576 с.
20. *Браунли К. А.* Статистическая теория и методология в науке и технике.— М.: Наука, 1977.— 407 с.
21. *Бронштейн И. Н., Семендяев К. А.* Справочник по математике для инженеров и учащихся вузов.— М.: Наука, 1981.— 720 с.
22. *Вознесенский В. А.* Оптимизация состава многокомпонентных добавок в композиты.— К.: О-во «Знание» УССР, 1981.— 20 с.
23. *Вознесенский В. А., Ковальчук А. Ф.* Принятие решений по статистическим моделям.— М.: Статистика, 1978.— 192 с.
24. *Волков Е. А.* Численные методы.— М.: Наука, 1982.— 256 с.
25. *Воробьев В. А., Кивран В. К., Корякин В. Н.* Применение физико-математических методов в исследовании свойств бетона.— М.: Высш. шк., 1977.— 272 с.
26. *Геминтерн В. И., Каган Б. М.* Методы оптимального проектирования.— М.: Энергия, 1980.— 160 с.
27. *Дворкин Л. И.* Оптимальное проектирование составов бетона.— Львов: Вища шк. Изд-во при Львов. ун-те, 1981.— 160 с.
28. *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ: В 2 кн. —2-е изд.— М.: Финансы и статистика, 1986.— Кн. 1—2.
29. *Еремин Н. Ф.* Процессы и аппараты в технологии строительных материалов.— М.: Высш. шк., 1986.— 280 с.
30. *Ермаков С. М., Жигляевский А. А.* Математическая теория оптимального эксперимента.— М.: Наука, 1987.— 320 с.
31. *Зайцев Ю. В.* Моделирование деформаций и прочности бетона методами механики разрушений.— М.: Стройиздат, 1982.— 196 с.
32. *Захаров В. К., Севастьянов Б. А., Чистяков В. П.* Теря вероятностей.— М.: Наука, 1983.— 159 с.
33. *Кафаров В. В.* Методы кибернетики в химии и химической технологии.— М.: Химия, 1985.— 448 с.
34. *Клименюк Н. Н.* Автоматизированные системы управления предприятиями сборного железобетона.— К.: Вища шк. Головное изд-во, 1984.— 215 с.
35. *Козин А. С., Ляшенко Н. Я.* Вычислительная математика.— К.: Рад. шк., 1983.— 191 с.
36. *Коршунов Ю. М.* Математические основы кибернетики.— 3-е изд.— М.: Энергоатомиздат, 1987.— 496 с.
37. *Краскевич В. Е., Зеленский К. Х., Гречко В. И.* Численные методы в инженерных исследованиях.— К.: Вища шк. Головное изд-во, 1986.— 263 с.
38. *Кудрявцев Е. М.* Исследование операций в задачах, алгоритмах и программах.— М.: Радио и связь, 1984.— 184 с.
39. *Кузьмик П. К., Маничев В. Б.* САПР-5: Автоматизация функционального проектирования.— М.: Высш. шк., 1986.— 143 с.
40. *Математическая теория планирования эксперимента / Под ред. С. М. Ермакова.*— М.: Наука, 1983.— 392 с.
41. *Методические указания по моделированию систем «смеси; технология — свойства» / В. А. Вознесенский, Т. В. Ляшенко, В. В. Абакумов, А. Абдыкалыков.*— Одесса: ОИСИ, 1985.— 64 с.
42. *Моделирование технологии и свойств газосиликата: исследование взаимосвязей между параметрами технологии и показателями свойств газосиликата по девятифакторному плану Рехтшафнера/К. К. Эскуссон, А. Э. Кильксон, И. Ю. Эскуссон и др.// Производство и применение силикатных бетонов: Сб. тр. НИПИ силикатобетон Минстройматериалов СССР.— Таллин, 1985.— С. 5—34.*
43. *Моисеев Н. Н.,* Математические методы системного анализа.— М.: Наука, 1981.— 487 с.
44. *Моисеев Н. Н., Иванюков Ю. П., Столярова Е. М.* Методы оптимизации.— М.: Наука, 1978.— 351 с.
45. *Налимов В. В.* Теория эксперимента.— М.: Наука, 1971.— 207 с.
46. *Налимов В. В., Голикова Т. И.* Логические основания планирования эксперимента — 2-е изд.— М.: Металлургия, 1981.— 152.

III.16.	Решение с помощью полиномиальных моделей задач оптимизации второго вида — минимизации расхода ресурса	257
III.17.	Принятие инженерных решений по комплексу экспериментально-статистических моделей	262
<i>Глава IV.</i>	<i>Принципы имитационного моделирования</i>	<i>269</i>
IV.1.	Использование моделей в режиме имитации	269
IV.2.	Генерирование случайных чисел на ЭВМ	272
IV.3.	Анализ методом Монте-Карло устойчивости технологических решений, принимаемых по экспериментально-статистическим моделям	274
IV.4.	Метод Монте-Карло при моделировании структуры и свойств композиционных материалов	278
IV.5.	Решение рецептурно-технологических задач на ЭВМ в режиме диалога	282
<i>Приложение I.</i>	<i>Некоторые сведения по математике</i>	<i>285</i>
I.1.	Элементы линейной алгебры	285
I.2.	Аналитическая геометрия	294
I.3.	Представление функций с помощью разложения Тейлора	300
I.4.	Основные понятия теории вероятности	303
<i>Приложение II.</i>	<i>Справочные таблицы</i>	<i>310</i>
	<i>Список использованной литературы</i>	<i>318</i>
	<i>Предметный указатель</i>	<i>321</i>

А

Адекватность 236
 Алгоритмизированное планирование эксперимента 201
 Алгоритм 21
 Альтернативная гипотеза 144, 173
 Антагонизм 233, 239
 Антиградиент 81
 Апостериорная информация 119, 122, 177
 Аппроксимация 23, 31, 161
 Априорная информация 119, 177

Б

База данных 269
 Бланк-алгоритм 228

В

Веер моделей 120, 191
 Вектор критериев 57
 — оценок коэффициентов 177
 — результатов 124
 Вершина многогранника 86, 98
 — отраженная 76
 — симплекса 76
 Воспроизведение испытаний 185
 Вход системы 8
 Выборка большая 155
 — малая 136, 149
 — объем 146
 Выход системы 10, 119

Г

Генератор случайных чисел 16, 272
 Генеральная совокупность 136
 Генерирующее соотношение 210
 Гиперкуб 12, 125
 Гипотеза статистическая 142
 — — проверка 145
 Гистограмма 160
 Градиент 62, 80
 Граница области 19, 83, 119, 241, 248
 Графическое отображение модели 16

Д

Декомпозиция 269
 Диалог «ЭВМ — технолог» 58, 70, 245, 269, 282
 Диаграмма 16, 125, 135, 231
 — регулировочная 258
 Дискретизация 21
 Дисперсионный анализ 169
 Дисперсия 137
 — внутригрупповая 170
 — межгрупповая 170
 — неадекватности 185
 — остаточная 174
 — оценки коэффициента ЭС-модели 183
 — предсказанного значения 187, 203, 235, 241

З

Задача инженерная типовая 18, 245, 257, 262
 — линейного программирования (ЛП) 64, 87
 — — — о загрузке оборудования 93
 — — — об использовании сырья 98
 — — — основная, каионическая (ОЗЛП) 96
 — — — планирования производства 88
 — — — размещения производства 94
 — — — составления смеси 91
 — — — транспортная 93, 110
 — математического программирования 64
 — нелинейного программирования 64
 — регрессионная типовая 182, 240
 — смесевая 230
 — экономии ресурсов 18, 257
 Значимость оценок 142, 183
 Значащая цифра 25

И

Изолиния 17, 131, 206, 221
 Изоповерхность 227, 237
 Изопараметрический анализ 264
 Индикатор 103

Интервал варьирования 11
 — группирования 155
 — доверительный 142, 146, 159, 183
 — неопределенности 71
 Интерполяция 48
 Интерпретация модели 128, 245, 257, 262
 — — при нарастающей степени риска 199
 Информационная таблица 123
 — ценность ЭС-модели 185
 Истинный коэффициент регрессии 182
 Итерационный процесс 22, 34, 37, 45, 101
 — — сходящийся 34

К

Каталог планов 218, 241
 Квантиль 140, 158
 Квантование 155
 Ковариация 150, 183
 Ковариационная матрица 178
 — — диагональный элемент 183
 Компромисс 19, 262
 Коитинуум 20
 Корреляционный анализ 153
 Коэффициент асимметрии 157
 — вариации 138, 146
 — избыточности плана 209, 218
 — корреляции 149
 — — множественной 151
 — — оценок модели 183
 — полнота 120
 — ЭС-модели 121, 196, 233
 Критерий Бартлетта 168
 — Кокрена 167
 — комплексный, интегральный 57, 262
 — оптимальности 54
 — — планов 201
 — скалярный 57
 — согласия 161
 — статистический 142
 — Стьюдента (t) 146, 165, 183
 — Фишера (F) 164, 185
 — χ^2 148, 161
 — частный 57
 Критический уровень 183

Л

Линеаризация 190
 Линейно связанные элементы 231

М

Мажоранта 31
 Максимум 18, 54, 252
 Математическое ожидание 137
 Матрица ковариационная (дисперсионно-ковариационная) 178, 200, 240

— — диагональная 201
 — — информационная 177
 — L 178, 222
 — плана эксперимента 122, 177
 Медиана 140
 Метод бисекций (дихотомии) 37
 — Бокса (комплекс-метод) 86
 — Бокса — Уэлсона 254
 — больших штрафов 107
 — Гаусса 40
 — Гаусса — Жордана 44
 — Гаусса — Зейделя 47, 73
 — деформированного многогранника (Нелдера и Мида) 78
 — диссоциативно-шаговый оптимизации (ДСМ) 249
 — градиента 80
 — итераций 37, 45
 — крутого восхождения 82, 254
 — Монте-Карло 270
 — наименьших квадратов (МНК) 33, 174, 181
 — нанскорейшего спуска 81
 — поиска оптимума 65, 67
 — — — симплексный 76
 — покоординатного спуска (Гаусса — Зейделя) 73
 — потенциалов 113
 — равномерного (общего) поиска 71
 — сканирования (перебора) 67
 — средних квадратов 273
 — случайного поиска 80
 — спуска 85
 — сравнения 273
 Метрика 31
 Минимум 18, 54, 98, 252
 — локальный 61
 Множество допустимых решений 54
 — цифр конечное 20
 Множитель Лагранжа 84
 Моделирование аналоговое 22
 — статистическое 334
 Модель имитационная 15, 269
 — исследования операций 15
 — квазиднофакторная 246
 — корреляционная 154
 — конкурирующая 120, 252
 — константа 186
 — концептуальная 15
 — линейная (первого порядка) 159, 172
 — линейная по параметрам 189
 — нелинейная 128, 189
 — однофакторная 172, 181
 — полиномиальная 194, 196
 — функциональная 15
 — экспериментально-статистическая (ЭС-модель) 14, 121
 — произведение 238
 Момент распределения 156
 Мощность критерия 143

Н

- Наблюдение 121
- Начальное приближение 68
 - задачи ЛП 101
- Невязка 173
- Норма 30, 173
 - матрицы 46
- Нормальные уравнения 175
- Нуль-гипотеза 142
 - о равенстве (однородности) дисперсий 164, 167
 - — — коэффициентов вариации 164
 - — — ЭС-модели нулю 183
 - — — средних 164, 169
 - — — частостей 164

О

- Область доверительная для оценок 184
 - допустимых (разрешенных) значений 57, 60, 86, 89, 263, 267
 - — — — объем 261, 268
 - запрещенная 86, 89, 241, 260, 264, 267
 - факторная 10, 126, 238, 240
 - — нестандартная 241
- Обобщающий показатель 261
- Объект моделирования 13
 - оптимизации 66
- Объемная диаграмма 227, 237, 267
- Ограничение 18, 58, 60, 63, 88, 267
 - активное 62
 - неравенство 59, 64
 - прямое, факторное 58
 - равенство 59
 - смешанное 64
 - функциональное 59
- Окрестность цели 68
- Округление числа 25
- Опорный план 98
- Определяющий контраст 210
- Оптимальное решение 54
 - задачи ЛП 98
- Оптимизационная задача 1-го рода 18, 130, 245
 - — 2-го рода 18, 130, 257
- Оптимизация 18, 54, 63
 - безусловная 63, 70
 - математическая задача 55
 - параметрическая 55, 57
 - условная 83
 - эксперимента 199
- Оптимум 18, 54
 - безусловный 62, 67
 - глобальный 61
- Опыт 122, 303
- Особая точка в факторной области 242
- Отклик 9, 17, 187, 240
- Оценка 137, 181
 - интервальная 137

- независимая коэффициента регрессии 183
- непараметрическая 140
- несмещенная 141
- параметрическая 139
- смешанная 210
- состоятельная 141
- точечная 141, 181
- эффективная 141
- Ошибка статистических выводов 142
 - — — 1-го и 2-го рода 143, 144
 - оценки среднего 146
 - эксперимента 183, 185

П

- Пакет программ 282
- Парабола 129, 297
- Параметр выходной 9
 - управляемый 55
 - формы кривой распределения 156
- Переменная в ЛП 88
 - — — базисная 98, 100
 - — — исключаемая 102
 - — — логическая 96
 - — — структурная 96
 - натуральная 11
 - нормализованная (безразмерная) 11, 126
 - смесевая 232
- План в ЛП 98
 - эксперимента 122, 202
 - — А-оптимальный 203
 - — Бокса 218, 220, 225, 315
 - — Бокса — Бенкина 225
 - — второго порядка 212
 - — двухуровневый 205
 - — D-оптимальный 203, 234, 240
 - — G-оптимальный 203
 - — E-оптимальный 203
 - — композиционный 204
 - — Q-оптимальный 203
 - — насыщенный 204, 234
 - — нормированный 202
 - — оптимальный 201
 - — ортогональный 201, 203
 - — Плакетта — Бермана 211
 - — Рехтшафнера 220, 225, 316
 - — ротабельный 204, 220, 225
 - — симметричный 203
 - — симплекс-решетчатый (Шеффе) 234, 317
 - — симплекс-суммируемый 220, 225
 - — униформный 204
 - — факторный дробный 209, 314
 - — — полный 205, 209, 215
 - — Хартли 220, 225, 315
- Плоскость 16, 132, 172
- Плотность распределения вероятностей 156, 308
- Погрешность 24, 34
 - абсолютная 24

- вычислительная 24
- относительная 24, 26, 28
- предельная 24, 25
- функции 26, 28
- Поиск экстремума (оптимума) 67, 70
 - — последовательный 72
- Показатель качества (технологический, основного назначения, структурный, вероятностный, технико-экономический) 10, 11
- Полигон 160
- Полином 120, 194
 - Лагранжа 49
 - приведенный 233, 238
- Правдоподобное суждение 126
- Предсказанное значение отклика 187
- Приближенное решение 30
 - число 24
- Принцип аддитивности 88
 - делимости 88
 - простоты 120
- Прогностическая возможность модели 187
 - — — мера точности (d -функция) 188
- Произведение полиномов 238
- Пространство факторное 10, 126, 236, 240, 241, 247

Р

- Размах варьирования фактора 11
 - выборки 140, 158
- Рандомизация 204
- Ранжирование 18
- Ранжированный ряд 140
- Распределение случайных величин 307
 - нормальное (Гаусса) 274, 307
 - — многомерное 149
 - равномерное 272, 308
 - эмпирическое 155, 160
- Регрессионный анализ 181
 - — последовательный (ПРА) 198
- Режим интерактивный (диалоговый) 183
 - пакетный 182
- Реплика плана 210
- Ресурс 18, 88, 257
 - ценность 109
- Рецептура 9, 230
- Риск статистического решения 141, 199
 - второго рода (производителя) 143
 - первого рода (потребителя) 142
- Ряд Тейлора 26, 52, 120, 300

С

- Свободный член 128
- Семейство параллельных прямых 132
 - непараллельных прямых 133
- Симплекс 76, 231

- Симплекс-метод в ЛП 100
 - — — — двухэтапный 107
- Симплекс-таблица 102
- Система линейных уравнений 40
 - «смесь, технология — свойства» 237
 - «состав (смесь) — свойства» 231
 - стохастическая 12
 - управляемая 9
- Синергизм 233, 239
- Синтез оптимальных планов 240
- Сканирование 67
- Сплайн-функция 16, 33
- Среднее 122
- Среднеквадратичное, среднеквадратическое, стандартное отклонение 138, 148
- Статистический анализ модели 181
- Суждение вероятностное 142

Т

- Теория подобия 14, 121
- Точность гауссовская 183
 - поиска 68
- Транзитивность модели 13
- Транспортная таблица 113

У

- Уравнение математической физики 14
 - состояния системы 13
- Уровень значимости 142
 - риска 144
 - свойства требуемый (нормативный) 21, 257, 263
 - фактора основной 11
 - — колеблющийся по заданному закону распределения 275
 - — оптимальный переменный 183
- Условие стационарности 61
- Устойчивость технологического решения 275

Ф

- Фактор 9, 119, 232
 - взаимонезависимый 232
 - нормализованный 11, 126
 - преобразованный 127
 - смесевой (взаимозависимый) 232
- Формула квадратурная 50
 - Крамера 40, 178
 - прямоугольников, трапеций 50
 - Симпсона 51
- Функция (системы) 8
 - базисная 190
 - барьерная 85
 - нелинейная, приводимая к прямой 191
 - — — к плоскости 195
 - отклика 240
 - случайных величин 26

- точности прогноза 188
- унимодальная 61
- штрафная 85

Ц

- Целевая функция 54, 88, 95
- Цель управления 9, 13, 64
- Центральный элемент в ЛП 103

Ч

- Частость 140, 159
- Частота 155
- Численное дифференцирование 52
 - интегрирование 50
- Численный метод 21
- Число опытов 122
 - — разных минимальное 200
 - параллельных измерений 146
 - псевдослучайное 272
 - степеней свободы 141, 185

- факторов 12
- эффектов в ЭС-модели 196
- Числовая характеристика случайной величины 137, 156, 309

Ш

- Шаг поиска 68, 74
- Шкала метрическая, номинальная, порядковая (райговая) 162

Э

- Эксперимент 121
 - оптимальное планирование 199
 - машинный 270
- Экстремум 54, 245
- Эксцесс 157
- Эллипс рассеяния 152, 202
- Эффект взаимодействия 131, 233
 - квадратичный 129
 - линейный 128
 - незначимый 182
 - смешанный 210
- Эффективность плана 209, 218, 225

ОГЛАВЛЕНИЕ

Предисловие	3	III.16. Решение с помощью полиномиальных моделей задач оптимизации второго вида — минимизации расхода ресурса	257
Условные обозначения и сокращения	6	III.17. Принятие инженерных решений по комплексу экспериментально-статистических моделей	262
Введение	8	Глава IV. Принципы имитационного моделирования	269
Глава I. Введение в численные методы решения математических задач	24	IV.1. Использование моделей в режиме имитации	269
I.1. Приближенные числа	24	IV.2. Генерирование случайных чисел на ЭВМ	272
I.2. Приближенные решения	29	IV.3. Анализ методом Монте-Карло устойчивости технологических решений, принимаемых по экспериментально-статистическим моделям	274
I.3. Методы решения уравнений	35	IV.4. Метод Монте-Карло при моделировании структуры и свойств композиционных материалов	278
I.4. Численные методы решения задач линейной алгебры	40	IV.5. Решение рецептурно-технологических задач на ЭВМ в режиме диалога	282
I.5. Интерполяция, численное интегрирование и дифференцирование	48	Приложение I. Некоторые сведения по математике	285
Глава II. Методы поиска оптимума в строительно-технологических задачах	54	I.1. Элементы линейной алгебры	285
II.1. Общая постановка и классификация задач оптимизации	54	I.2. Аналитическая геометрия	294
II.2. Основные этапы и классификация методов поиска экстремума	65	I.3. Представление функций с помощью разложения Тейлора	300
II.3. Поиск безусловного экстремума	70	I.4. Основные понятия теории вероятности	303
II.4. Методы условной оптимизации	83	Приложение II. Справочные таблицы	310
II.5. Постановка задач линейного программирования	87	Список использованной литературы	318
II.6. Симплекс-метод решения задач линейного программирования	100	Предметный указатель	321
II.7. Транспортная задача как линейная оптимизационная модель	110		
Глава III. Элементы математической теории эксперимента	119		
III.1. Экспериментально-статистические модели и информационная таблица для их построения	119		
III.2. Экспериментально-статистические модели в нормализованном факторном пространстве	125		
III.3. Эксперимент в одной точке факторного пространства и статистические характеристики его результатов (малая выборка, одно свойство)	136		
III.4. Эксперимент в одной точке факторного пространства и его статистические характеристики (малая выборка, несколько свойств)	149		
III.5. Эксперимент в одной точке факторного пространства (большая выборка)	155		
III.6. Сравнение результатов эксперимента в двух и нескольких точках факторного пространства	163		
III.7. Построение и статистический анализ линейных моделей	172		
III.8. Основные этапы регрессионного анализа линейных моделей	181		
III.9. Построение и статистический анализ нелинейных моделей (линейных по параметрам)	189		
III.10. Основные принципы оптимального планирования экспериментов	199		
III.11. Планы полного и дробного факторного эксперимента для построения линейных и неполных квадратичных моделей	205		
III.12. Планы для построения квадратичных моделей	212		
III.13. Моделирование свойств смесей	230		
III.14. Синтез оптимальных планов эксперимента	240		
III.15. Решение с помощью полиномиальных моделей задач оптимизации первого вида — поиска экстремумов выхода	245		