

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
Государственное образовательное учреждение высшего профессионального образования
«ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Л.И. Дорофеева

**МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ
РАЗДЕЛИТЕЛЬНЫХ ПРОЦЕССОВ**

*Рекомендовано в качестве учебного пособия
Редакционно-издательским советом
Томского политехнического университета*

Издательство
Томского политехнического университета
2008

УДК 519.237.7 + 621.039.33

ББК

Д

Дорофеева Л.И.

Д Моделирование и оптимизация разделительных процессов: учебное пособие / Л.И. Дорофеева. – Томск: Изд-во Томского политехнического университета, 2008. – 128 с.

В учебном пособии изложены основы математических методов планирования эксперимента, градиентные методы поиска оптимальной области, ортогональное композиционное планирование второго порядка при определении оптимальных условий разделения веществ. Рассмотрены методы физического и математического моделирования с применением теории подобия. Отражены проблемы информационного моделирования на основе искусственных нейронных сетей. Приведены примеры использования теоретического материала для расчёта процессов, протекающих при разделении изотопов и очистки веществ.

Учебное пособие предназначено для магистрантов программы «Физика кинетических явлений», а также студентов специальности 140303 физико-технического факультета.

УДК 519.237.7 + 621.039.33

ББК

Рецензенты

Доктор физико-математических наук,
профессор ТГУ
В.Г. Бутов

Доктор физико-математических наук,
профессор НИИПММ
В.А. Архипов

© Дорофеева Л.И., 2008

© Томский политехнический университет, 2008

© Оформление. Издательство Томского
политехнического университета, 2008

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
ГЛАВА 1. МАТЕМАТИЧЕСКИЕ МЕТОДЫ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА.....	7
§ 1. Разработка регрессионных моделей разделительных процессов	9
§ 2. Оценка адекватности математической модели.....	21
§ 3. Методы оптимизации эксперимента.....	23
§ 3.1. Планирование экстремальных экспериментов	29
§ 3.2. Симплексное планирование экспериментов при определении оптимальных условий разделения веществ	38
§ 3.3. Оптимизация разделительных процессов методом крутого восхождения	45
§ 3.4. Ортогональное планирование 2-го порядка при поиске оптимальных условий	48
§ 4. Моделирование и оптимизация процессов изотопного разделения	54
Контрольные вопросы и задания	57
ГЛАВА 2. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ С ПРИМЕНЕНИЕМ МЕТОДОВ ТЕОРИИ ПОДОБИЯ	59
§ 1. Теория подобия	59
§ 2. Константы подобия.....	60
§ 3. Критерии подобия.....	63
§ 3.1. Критерии подобия в тепломассообменных процессах разделения веществ	64
§ 4. Критериальные уравнения	84
Контрольные вопросы и задания	86
ГЛАВА 3. ИНФОРМАЦИОННОЕ МОДЕЛИРОВАНИЕ С ПРИМЕНЕНИЕМ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ	87
§ 1. Основы теории нейронных сетей	91
§ 1.1. Модели нейрона	92
§ 1.2. Нейронные сети	97
§ 2. Преимущества нейронных сетей.....	107
§ 3. Искусственный интеллект и нейронные сети	111
§ 4. Нейросетевое моделирование.....	116
§ 5. Применение искусственных нейронных сетей	118
Контрольные вопросы и задания	124
СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ	125

ВВЕДЕНИЕ

Моделирование является одним из универсальных и прогрессивных методов, широко применяемых в современной науке, в первую очередь, в ее прикладных областях. Моделирование позволяет ускорить технический прогресс, существенно сократить сроки освоения новых производств. В последнее время особенно активно развиваются такие направления как – математическое и компьютерное моделирование, их развитие напрямую связано с развитием вычислительной техники.

Математическое и компьютерное моделирование многостадийных процессов разделения и очистки веществ – один из основных способов получения новых научных знаний и технологических решений, который позволяет исключить аварийные ситуации и эксплуатационные расходы, связанные с отработкой различных режимов эксплуатации оборудования, поэтому играет важную роль в работе любого предприятия.

Моделирование разделительных процессов позволяет осуществлять:

- оптимальное проектирование новых и интенсификацию действующих технологических процессов;
- контроль за ходом процесса и получение необходимой информации о нем, обработку полученной информации с целью управления ходом технологического процесса;
- решение задач исследования объектов, где невозможно проводить активные эксперименты – изменение технологических режимов работы разделительного производства, отработка аварийных ситуаций и т. д.;
- максимальное ускорение переноса результатов лабораторных исследований в промышленные масштабы.

Термин «*моделирование*» предполагает изучение объекта путем построения и исследования его модели, осуществляемое с определенной целью, и состоит в замене эксперимента с оригиналом экспериментом на модели [37]. Модель должна строиться так, чтобы она наиболее полно воспроизводила те качества объекта, которые необходимо изучить в соответствии с поставленной целью. Во всех отношениях модель должна быть проще объекта и удобнее его для изучения. Таким образом, для одного и того же объекта могут существовать различные модели, классы моделей, соответствующие различным целям его изучения. Необходимым условием моделирования является подобие объекта и его модели.

Основные положения теории подобия, критерии подобия, применяемые для расчёта тепломассообменных процессов, приводятся во второй главе настоящего пособия.

Существует ряд требований и условий, которым должна удовлетворять вновь создаваемая модель:

- затраты на создание модели должны быть значительно меньше затрат на создание оригинала,
- должны быть четко определены правила интерпретации результатов вычислительного эксперимента,
- модель должна быть существенной.

Данное требование заключается в том, что модель должна отражать необходимые, существенные для решения конкретной задачи свойства объекта. Для одного и того же объекта сложно создать обобщенную модель, отражающую все его свойства. Поэтому важно обеспечить существенность модели.

- Модель должна быть адекватной моделируемому объекту или явлению.

Оценка адекватности математических моделей приводится в первой главе пособия.

Математическая модель – совокупность математических зависимостей, отражающая в явной форме сущность технологического процесса, т. е. все существенные параметры технологического объекта, связанные системой математических уравнений.

Компьютерная модель – программная реализация математической модели, к которой могут дополнительно прилагаться различные программные модули, служащие, например, для графического представления изучаемых объектов. Главным преимуществом компьютерной модели является относительная простота её создания и модификации. Применение компьютерного моделирования вызвано необходимостью исследования сложных [20] (например, многостадийных) объектов, явлений и процессов.

По полноте отражения внутренних свойств объекта моделирования модели подразделяют на динамические и статические [22].

Статические модели отражают работу объекта исследований в стационарных (установившихся) условиях. Соответственно математическое описание в статистических моделях не включает время как переменную и состоит из алгебраических уравнений либо дифференциальных.

Динамические модели отражают изменение объекта во времени. Математическое описание таких моделей обязательно включает производную по времени. Часто динамическую модель процесса строят в виде передаточных функций, связывающих входные и выходные переменные.

По используемому математическому аппарату математические модели подразделяются на:

- 1) *детерминированные* – при изменении любого параметра значение выходных величин определяются однозначно.
- 2) *статистические* – результат определяется с определенной степенью достоверности (т. е. неоднозначно определяется заданными параметрами).

Процедура построения математической модели во многом зависит от ее целевого назначения, свойств объекта, от количества и качества имеющейся информации. Наличие достаточной информации о механизме процесса позволяет составить детерминированную математическую модель процесса.

Детерминированную математическую модель составляют на основе теоретического анализа физико-химических процессов, происходящих в объекте. При выводе уравнений математических моделей технологических процессов учитывают гидродинамические режимы перемещения веществ, скорости химических реакций, диффузии, теплопередачи, материальный и тепловой балансы, фазовые превращения. Все это требует углубленного изучения объекта моделирования.

Таким образом, математические модели подразделяются по назначению, виду моделируемого объекта, методу построения и решения модели. В зависимости от назначения модели укрупнено делятся на оптимизационные и информатизационные.

Методы оптимизации экспериментальных исследований с целью определения наиболее благоприятных условий осуществления процессов разделения и очистки веществ, а также построения на их основе оптимальных моделей приводятся в первой главе пособия. Информационное моделирование с применением искусственных нейронных сетей рассматривается в третьей главе учебного пособия.

ГЛАВА 1

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

Развитие методов планирования эксперимента началось с появления в 1935 г. монографии английского статистика Р. А. Фишера «Планирование эксперимента», в которой для нужд агротехнических исследований с помощью дисперсионного анализа решались задачи по планированию эксперимента с дискретными факторами.

В 1951 г. появились основополагающие работы американских учёных Дж. Бокса и Дж. Кифера, позволяющие планировать экономный и эффективный эксперимент, результаты которого можно было использовать для получения адекватного описания «поверхности отклика» в виде многофакторного уравнения.

Ряд оригинальных направлений в теории планирования эксперимента возник благодаря работам по теории вероятностей (А.Н. Колмогоров, Ю.В. Линник и их ученики) с одной стороны и вычислительной математики (Г.И. Марчук и его ученики) – с другой. Существенный вклад в развитие теории планирования эксперимента и её приложений внесли такие учёные как В.В. Налимов, В.В. Фёдоров, Г.К. Круг, Е.В. Маркова и другие.

Экспериментальное исследование является необходимым этапом процесса познания. Имея дело с различными процессами и объектами в своей деятельности, человек постепенно накапливает практические сведения об этих процессах и объектах, что помогает ему познать их и на основе этого знания усовершенствовать, оптимизировать. Зачастую сведений явно недостаточно, чтобы сделать какие-либо выводы. Тогда тот или иной объект, процесс подвергают планомерному экспериментальному исследованию.

Экспериментом называется совокупность опытов, объединённых единой целью, единой системой ограничений в пространстве и времени (u – номер опыта, $u = 1 \div N$).

Опыт можно считать реализацию на каком-либо объекте некоторых условий, правил. В результате опыта появляется то или иное событие (например, событием будет обогащение целевого изотопа либо получение особо чистого продукта). Появление события регистрируется при помощи какого-либо параметра, имеющего, как правило, численное выражение и наиболее полно характеризующего эффективность исследуемого процесса в u -м опыте с позиции цели исследования. Такой па-

раметр обычно обозначается буквой y_i и называется *критерием оптимальности*.

Всеобъемлющим критерием оптимальности является результат, конечная цель эксперимента. Частным критерием оптимальности может быть любой технологический параметр (поток отбора (отвала), коэффициент разделения (обогащения) и т. д.) – такой критерий обычно называют *выходным параметром процесса*.

Результат процесса зависит от условий его протекания, характеризуемых значениями параметров, влияющих на процесс (например, температуры, давления, величины pH , скорости движения фаз и т. д.). Эти параметры называют *факторами* или *входными параметрами процесса*. Численное значение любого фактора x_i должно устанавливаться и реализовываться независимо от значений других факторов (i – номер фактора, $i = 1 \div k$).

Условия проведения каждого опыта эксперимента в виде конкретного значения исследуемых факторов, а иногда и очередность проведения опытов регламентируются *планом эксперимента*. По результатам опытов, следуя определенному алгоритму, получают соответствующее *уравнение*, характеризующее влияние факторов на эффективность исследуемого процесса.

При выборе типа плана следует учитывать его возможные недостатки или преимущества перед другими типами планов, имеющих одно и то же число опытов [18], один и тот же диапазон изменения факторов, но разное расположение экспериментальных точек.

Планы, обеспечивающие получение наименьшей величины максимальной дисперсии предсказания, называют *G-оптимальными*.

Планы, обеспечивающие получение одинаковых дисперсий предсказания для точек, равноотстоящих от центра эксперимента, называют *ротатабельными*.

Планы, обеспечивающие получение одинаковой величины дисперсии предсказания для любой точки в пределах изучаемой области, называются *униформ-ротатабельными*.

D-оптимальными называют планы, обеспечивающие получение минимальной обобщенной дисперсии коэффициентов уравнения.

Планы, обеспечивающие получение уравнения, коэффициенты которого ортогональны друг другу (независимы друг от друга) и имеют одинаковую дисперсию, называются *ортогональными* [23].

Оптимизация процесса (объекта) сводится к отысканию таких условий его протекания (таких конструктивных размеров, технологических параметров и т. д.), при которых критерий оптимальности будет иметь экстремальное значение: максимум, если увеличение численного

значения критерия оптимальности означает увеличение эффективности процесса (объекта), или минимум, если его увеличение означает снижение эффективности процесса (объекта).

Изучение всех влияющих на исследуемый объект факторов одновременно провести невозможно, поэтому в эксперименте рассматривается их ограниченное число. Остальные активные факторы стабилизируются, т. е. устанавливаются на каких-то одинаковых для всех опытов уровнях.

Эффективность экспериментов в большой степени зависит от методов их проведения. Различают *пассивный* и *активный* эксперименты. Пассивный эксперимент является традиционным методом, когда ставится большая серия опытов с поочередным варьированием каждой из переменных. К пассивному эксперименту относится также сбор исходного статистического материала в режиме нормальной эксплуатации на промышленном объекте. Обработка опытных данных для получения математической модели проводится статистическими методами. Методы математической статистики позволяют в этом случае извлечь максимум информации из имеющихся экспериментальных данных – оптимизировать процедуру обработки и анализа эксперимента. Используя активный эксперимент (планирование эксперимента), можно достичь существенно большего – оптимизировать и стадию постановки эксперимента.

Широкое применение математической статистики позволило осуществить статистический анализ первичных экспериментальных данных, обеспечить оценку их однородности и достоверности, определить необходимый для принятия тех или иных решений объем данных. Разработанные в рамках теории эксперимента алгоритмы формализации экспериментальных данных и использования полученных моделей для оптимизации процесса обеспечили возможность широкого внедрения в практику научных исследований математических методов планирования эксперимента.

§ 1. Разработка регрессионных моделей разделительных процессов

Современное разделительное производство включает в себя сотни разнородных процессов, между тем необходимо решать задачу оптимизации и управления этими процессами. Таким образом, специфика процессов разделения ионных и изотопных смесей требует разработки математических моделей, обеспечивающих оптимизацию и прогнозирование в поведении таких систем. Особенно это касается технологии разделения изотопов.

Для этих целей успешно применяются экспериментально-статистические методы, с помощью которых составляют математическую модель, изучая зависимость отклика системы на изменения входов.

На рис. 1.1 x_1, \dots, x_k – входные измеряемые и регулируемые параметры объекта; w_1, \dots, w_l – неконтролируемые, случайным образом изменяющиеся параметры, «шум» объекта; y_1, \dots, y_u – выходные параметры. В качестве случайных рассматриваются обычно параметры, которые по тем или иным причинам невозможно (или очень трудно) учесть, например, падение активности катализатора, изменение состояния поверхности теплообменной аппаратуры, колебания наружной температуры воздуха и т. п. Комплекс параметров x_1, \dots, x_k называют основным, он определяет условия эксперимента. Такое подразделение входных параметров на основные и случайные условно. Случайным будет любой параметр, не вошедший в основной комплекс входных параметров, даже если он хорошо изучен. В зависимости от постановки задачи и технических возможностей некоторые измеряемые параметры относят к «шуму» объекта. Однако при этом уменьшается точность математической модели [11]. В качестве выходных величин рассматривают любые технологические или экономические показатели процесса.

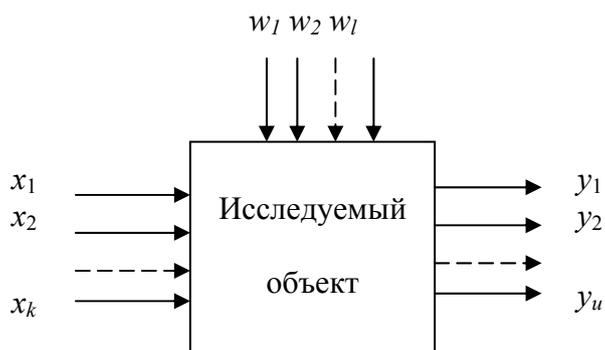


Рис. 1.1. Модель многофакторного процесса

Математической моделью объекта служит *функция отклика*, вызывающая выходной параметр, характеризующий результаты эксперимента, с переменными, которые варьируют при проведении опытов:

$$y = \varphi(x_1, \dots, x_k) \quad (1.1)$$

Принято называть независимые переменные x_1, \dots, x_k *факторами*, пространство с координатами x_1, \dots, x_k – *факторным пространством*, а геометрическое изображение функции отклика в факторном пространстве – *поверхностью отклика*.

При использовании статистических методов математическая модель статики процесса часто представляется в виде полинома – отрезка ряда Тейлора, в который разлагается неизвестная функция (1.1):

$$y = \beta_0 + \sum_{j=1}^k \beta_j x_j + \sum_{u,j=1}^k \beta_{uj} x_u x_j + \sum_{j=1}^k \beta_{jj} x_j^2 + \dots, \quad (1.2)$$

$$\text{где } \beta_0 = \varphi(0); \quad \beta_j = \frac{\partial \varphi(0)}{\partial x_j}; \quad \beta_{ij} = \frac{\partial^2 \varphi(0)}{\partial x_i \partial x_j}; \quad \beta_j = \frac{\partial^2 \varphi(0)}{2 \partial x_j^2} \quad \dots$$

Необходимость проведения эксперимента на объекте исследований связана с тем, что нужно устранить некоторую неопределённость, свойственную знаниям об объекте до проведения этого эксперимента. Результат эксперимента на сложном объекте обычно есть величина случайная. Существует много причин, приводящих к тому, что результаты наблюдения и измерения, сделанные в экспериментах оказываются случайными величинами. Иногда случайность предопределяется самой физической сущностью явлений: процессы происходят на молекулярном или атомном уровнях, а измеряются макроскопическими приборами.

Организация эксперимента и обработка экспериментальных данных, определяют степень уменьшения неопределённости знаний об объекте исследований и должны исходить из природы, существа и причины неопределённости. Говоря о случайных явлениях, прежде всего, обращают внимание на их непредсказуемость, противопоставляют случайность детерминированности, хаотичность – упорядоченности. Имеющее определенный смысл, такое противопоставление является односторонним, так как оставляет в тени тот факт, что под случайностью понимается вид неопределенности, подчиняющийся строгой закономерности, которая выражается распределением вероятностей [37]. Зная распределение (например, плотность $p(x)$) вероятностей, можно ответить на любой вопрос о случайной величине: в каком интервале находятся ее возможные значения (определим носитель распределения X -множество элементов x , для которых $p(x) > 0$); около какого значения рассеиваются ее реализующие значения (найдем параметр положения распределения, например среднее, модуль или медиану); насколько разбросаны эти значения (найдем масштабный параметр – дисперсию или стандартное отклонение, средний модуль разности, энтропию); какова связь между разными реализациями (вычислим заданную меру зависимости) и т. д.

Неучтенные факторы, «шум» объекта приводят к тому, что в результате повторных измерений в большинстве реальных экспериментов могут получиться отличающиеся друг от друга значения измеряемых величин. Как и любые эксперименты, измерения случайных величин и процессов выполняются для уточнения их моделей, снятия или уменьшения неопределенности незнания. Обычно достаточно знать не всё распределение, а лишь какой-то из его параметров, и тогда задача сводится к оценке этого параметра по наблюдаемой выборке. Хотя это уже «вторичная» обработка данных, измерение выборочных значений и вы-

числение оценки в совокупности можно трактовать как «измерение параметра». То же относится и к определению по выборке более сложных характеристик – самих распределений, регрессий, корреляций, спектров и т. д. Такое совместное рассмотрение непосредственных измерений и их обработки оказывается полезным еще и потому, что можно проводить общую оптимизацию этого процесса, и она далеко не всегда совпадает с оптимизацией компонент в отдельности [11].

Поэтому при обработке и анализе экспериментальных данных используют методы математической статистики. Так, для полиномиальной модели (1.2) получают так называемые *выборочные коэффициенты* регрессии b_0, b_j, b_{uj}, b_{jj} , являющиеся оценками *теоретических коэффициентов* $\beta_0, \beta_j, \beta_{uj}, \beta_{jj}$. Уравнение регрессии, полученное на основании экспериментальных данных, запишется следующим образом:

$$y = b_0 + \sum_{j=1}^k b_j x_j + \sum_{u,j=1}^k b_{uj} x_u x_j + \sum_{j=1}^k b_{jj} x_j^2 + + \sum_{\substack{i,u,j=1 \\ i \neq j \neq u}}^k b_{iuj} x_i x_u x_j \dots, \quad (1.3)$$

где b_0 – свободный член уравнения регрессии; b_j – линейные эффекты, $j = 1, 2, \dots, k$; b_{uj} – квадратичные эффекты; b_{jj} – эффекты парного взаимодействия; b_{iuj} – эффекты тройного взаимодействия.

Полиномиальные (регрессионные модели) модели справедливы только для объекта, на котором проводился эксперимент. Зная оценки коэффициентов отрезка ряда Тейлора, нельзя восстановить исходную функцию, аналитическое выражение которой остается неизвестным исследователю, и, следовательно, невозможно получить информацию о механизме процесса. В практическом отношении полиномиальные модели очень полезны и широко используются при решении задач оптимизации и управления технологическими процессами. При применении экспериментально-статистических методов, в ряде случаев экспериментатор располагает определенной априорной информацией о физической сущности исследуемого процесса, пользуясь которой можно получить представление о структуре модели.

Построение математической модели и её анализ в процессах разделения изотопов осуществляется в следующей последовательности [12]:

- 1) предварительное исследование процесса,
- 2) составление по результатам исследований модели,
- 3) интерпретация модели и её использование.

При планировании по схеме *полного факторного эксперимента* (ПФЭ) реализуются все возможные комбинации факторов на всех выбранных для исследования уровнях. Необходимое количество опытов N при ПФЭ определяется по формуле

$$N = n^k \quad (1.4)$$

где n – количество уровней; k – число факторов.

Если эксперименты проводятся только на двух уровнях, при двух значениях факторов и при этом в процессе эксперимента осуществляются все возможные комбинации из k факторов, то постановка опытов по такому плану называется полным факторным экспериментом типа 2^k . Уровни факторов представляют собой границы исследуемой области по данному технологическому параметру.

Например, рассмотрим процесс переноса вещества в двухфазной системе ионит – раствор. Эффективность работы разделительной установки будет зависеть от величины эквивалентной теоретической ступени (ВЭТС) ионообменной колонны, под которой подразумевается участок колонны, удовлетворяющий определению равновесной ступени разделения. Определим влияние на ВЭТС [см] в ионообменной колонне (y) трех факторов:

1. скорости движения раствора (в диапазоне $0,07 \div 0,13$) – z_1 , [см/с];
2. скорости движения ионита (в диапазоне $0,04 \div 0,06$) – z_2 , [см/с];
3. радиуса зерна ионита (в диапазоне $0,02 \div 0,03$) – z_3 , [см].

Верхний уровень по первому фактору z_1^{max} равен 0,13 см/с; нижний z_1^{min} равен 0,07 см/с. Тогда для z_1 имеем

$$z_1^0 = \frac{z_1^{max} + z_1^{min}}{2} = 0,1 \text{ см/с} \quad (1.5)$$

$$\Delta z_1^0 = \frac{z_1^{max} - z_1^{min}}{2} = 0,03 \text{ см/с} \quad (1.6)$$

Тогда для любого фактора z_j

$$z_j^0 = \frac{z_j^{max} + z_j^{min}}{2}, \quad j = 1, 2, \dots, k; \quad (1.7)$$

$$\Delta z_j^0 = \frac{z_j^{max} - z_j^{min}}{2}. \quad (1.8)$$

Точка с координатами (z_1^0, \dots, z_k^0) называется центром плана, иногда ее называют основным уровнем; Δz_j^0 – интервал варьирования по оси z_j . От переменных z_1^0, \dots, z_k^0 перейдем к новым – x_1, \dots, x_k путем следующего линейного преобразования:

$$x_j = \frac{z_j - z_j^0}{\Delta z_j^0}, \quad j = 1, 2, \dots, k. \quad (1.9)$$

Для переменных x_1, \dots, x_k верхний уровень (рис. 1.2) равен (+1), нижний уровень (-1), координаты центра плана равны нулю и совпадают с началом координат. В рассматриваемом примере $k = 3$.

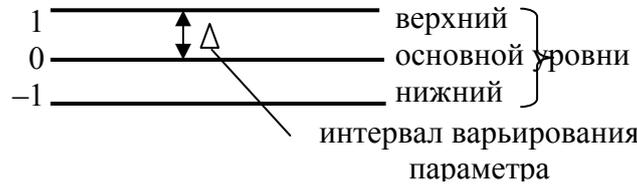


Рис. 1.2. Уровни факторов:
 $x = 0$ (средний уровень), $x = +1$ (верхний), $x = -1$ (нижний)

Число возможных комбинаций из трех факторов на двух уровнях с учётом (1.4) равно 8. План проведения экспериментов (матрица планирования) записывается в виде табл. 1.1.

Таблица 1.1

ПФЭ типа 2^3

Номер опыта	Факторы						
	в натуральном масштабе			в безразмерном масштабе			
	z_1	z_2	z_3	x_1	x_2	x_3	y
1	0,07	0,04	0,02	-1	-1	-1	3,14
2	0,13	0,04	0,02	+1	-1	-1	6,40
3	0,07	0,06	0,02	-1	+1	-1	2,86
4	0,13	0,06	0,02	+1	+1	-1	5,98
5	0,07	0,04	0,03	-1	-1	+1	6,65
6	0,13	0,04	0,03	+1	-1	+1	12,98
7	0,07	0,06	0,03	-1	+1	+1	6,36
8	0,13	0,06	0,03	+1	+1	+1	12,49

Представленный в табл. 1.1 план в безразмерном масштабе геометрически может быть интерпретирован в виде восьми вершин куба (рис. 1.3).

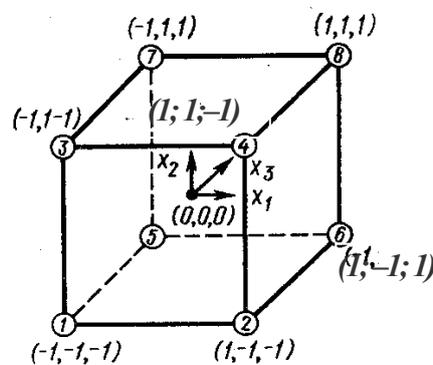


Рис. 1.3. Геометрическая интерпретация ПФЭ типа 2^3

Таблица 1.2

Матрица планирования с фиктивной переменной

Номер опыта	x_0	x_1	x_2	x_3	y
1	+1	-1	-1	-1	y_1
2	+1	+1	-1	-1	y_2
3	+1	-1	+1	-1	y_3
4	+1	+1	+1	-1	y_4
5	+1	-1	-1	+1	y_5
6	+1	+1	-1	+1	y_6
7	+1	-1	+1	+1	y_7
8	+1	+1	+1	+1	y_8

Приведенная в табл. 1.2 матрица планирования обладает следующими свойствами:

$$\begin{aligned}
 \sum_{i=1}^N x_{ui} x_{ji} &= 0; \quad u \neq j; \quad u, j = 0, 1, \dots, k; \\
 \sum_{i=1}^N x_{ji} &= 0; \quad j = 1, 2, \dots, k; \quad j \neq 0 \\
 \sum_{i=1}^N x_{ji}^2 &= N; \quad j = 0, 1, \dots, k,
 \end{aligned} \tag{1.10}$$

где k – число независимых факторов; N – число опытов в матрице планирования.

Первое свойство (1.10) – равенство нулю скалярных произведений всех вектор-столбцов – называется свойством ортогональности матрицы планирования. Это свойство резко уменьшает трудности, связанные с расчётом коэффициентов уравнения регрессии, так как матрица коэффициентов нормальных уравнений ($X^T X$) становится диагональной и её диагональные элементы равны числу опытов в матрице планирования.

Коэффициенты уравнения регрессии по методу наименьших квадратов определяются следующим образом:

$$\mathbf{B} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix} = (X^T X)^{-1} X^T Y. \tag{1.11}$$

Матрица моментов ($X^T X$), соответствующая табл. 1.2, имеет вид:

$$(X^T X) = \begin{bmatrix} \sum_{i=1}^8 x_{0i}^2 & \sum_{i=1}^8 x_{0i}x_{1i} & \sum_{i=1}^8 x_{0i}x_{2i} & \sum_{i=1}^8 x_{0i}x_{3i} \\ \sum_{i=1}^8 x_{1i}x_{0i} & \sum_{i=1}^8 x_{1i}^2 & \sum_{i=1}^8 x_{1i}x_{2i} & \sum_{i=1}^8 x_{1i}x_{3i} \\ \sum_{i=1}^8 x_{2i}x_{0i} & \sum_{i=1}^8 x_{2i}x_{1i} & \sum_{i=1}^8 x_{2i}^2 & \sum_{i=1}^8 x_{2i}x_{3i} \\ \sum_{i=1}^8 x_{3i}x_{0i} & \sum_{i=1}^8 x_{3i}x_{1i} & \sum_{i=1}^8 x_{3i}x_{2i} & \sum_{i=1}^8 x_{3i}^2 \end{bmatrix}. \quad (1.12)$$

Учитывая свойства (1.12), получаем:

$$(X^T X) = \begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}. \quad (1.13)$$

Обратная матрица получается путём деления союзной матрицы (или присоединённой) на детерминант (или определитель матрицы):

$$(X^T X)^{-1} = \begin{bmatrix} \frac{1}{8} & 0 & 0 & 0 \\ 0 & \frac{1}{8} & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 \\ 0 & 0 & 0 & \frac{1}{8} \end{bmatrix}. \quad (1.14)$$

Произведение транспонированной матрицы на вектор-столбец результатов:

$$X^T Y = \begin{bmatrix} \sum_{i=1}^N x_{0i}y_i \\ \sum_{i=1}^N x_{1i}y_i \\ \sum_{i=1}^N x_{2i}y_i \\ \sum_{i=1}^N x_{3i}y_i \end{bmatrix}. \quad (1.15)$$

Коэффициенты искомого уравнения:

$$B = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{bmatrix} \frac{1}{8} & 0 & 0 & 0 \\ 0 & \frac{1}{8} & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 \\ 0 & 0 & 0 & \frac{1}{8} \end{bmatrix} \times \begin{bmatrix} \sum_{i=1}^8 x_{0i}y_i \\ \sum_{i=1}^8 x_{1i}y_i \\ \sum_{i=1}^8 x_{2i}y_i \\ \sum_{i=1}^8 x_{3i}y_i \end{bmatrix}. \quad (1.16)$$

Следовательно, любой коэффициент уравнения регрессии b_j определяется скалярным произведением столбца y на соответствующий столбец x_j , делённым на число опытов в матрице планирования N :

$$b_j = \frac{1}{N} \sum_{i=1}^N x_{ji}y_i. \quad (1.17)$$

Используя план эксперимента в табл. 1.1 вычисляют коэффициенты уравнения регрессии:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \quad (1.18)$$

Таким образом, для определения коэффициента b_1 при x_1 вычисляют сумму произведений:

$$\begin{bmatrix} -1 \\ +1 \\ -1 \\ +1 \\ -1 \\ +1 \\ -1 \\ +1 \end{bmatrix} \times \begin{bmatrix} 3,14 \\ 6,40 \\ 2,86 \\ 5,98 \\ 6,65 \\ 12,98 \\ 6,36 \\ 12,49 \end{bmatrix} = \begin{bmatrix} -3,14 \\ +6,40 \\ -2,86 \\ +5,98 \\ -6,65 \\ +12,98 \\ -6,36 \\ +12,49 \end{bmatrix}$$

$$b_1 = \frac{1}{8} \sum_{i=1}^N x_{1i}y_i = \frac{18,84}{8} = +2,36.$$

Аналогично для остальных коэффициентов, тогда искомое уравнение регрессии будет иметь вид:

$$\hat{y} = 7,108 + 2,355x_1 - 0,185x_2 + 2,513x_3 \quad (1.19)$$

Линейные коэффициенты целиком характеризуют влияние исследуемого фактора на процесс. Если, например, $b_1 = 2,355$, то это значит, что увеличение этого фактора от центра эксперимента на 1 интервал варьирования вызывает увеличение выхода процесса на 2,355 единиц.

Анализ полученной математической модели показывает, что наибольшее влияние на ВЭТС оказывает скорость движения раствора (удельный вес фактора составляет 2,36) и радиус зерна ионита (удельный вес фактора составляет 2,51), увеличение скорости движения ионита приводит к снижению оптимизируемой величины и увеличению эффективности разделения. Данную модель можно использовать для определения оптимальных параметров процесса с целью достижения наиболее благоприятных условий проведения процесса.

Однако линейное уравнение отражает с достаточной точностью поверхность отклика лишь в некоторой локальной области, соответствующей изученному диапазону изменения факторов. На некотором удалении от изученного участка характер поверхности отклика будет всё больше отходить от плоскости, описанной линейным уравнением, поэтому линейную модель постепенно усложняют, и по мере приближения к околооптимальной области уравнение будет содержать всё большее число значимых факторов [16].

Для учёта влияния парного и тройного взаимодействия определяются коэффициенты b_{12} , b_{23} , b_{13} и b_{123} для этой матрицы, приведённая в табл. 1.2, расширяется (табл. 1.3).

Таблица 1.3

Расширенная матрица планирования ПФЭ типа 2^3

Номер опыта	Факторы				Эффекты межфакторных взаимодействий				у
	x_0	x_1	x_2	x_3	x_{12}	x_{13}	x_{23}	x_{123}	
1	+1	-1	-1	-1	+1	+1	+1	-1	3,14
2	+1	+1	-1	-1	-1	-1	+1	+1	6,4
3	+1	-1	+1	-1	-1	+1	-1	+1	2,86
4	+1	+1	+1	-1	+1	-1	-1	-1	5,98
5	+1	-1	-1	+1	+1	-1	-1	+1	6,65
6	+1	+1	-1	+1	-1	+1	-1	-1	12,98
7	+1	-1	+1	+1	-1	-1	+1	-1	6,36
8	+1	+1	+1	+1	+1	+1	+1	+1	12,49

С учётом табл. 1.3 полиномиальная модель (1.19) приобретёт вид неполного квадратного уравнения:

$$\hat{y} = 7,108 + 2,355x_1 - 0,185x_2 + 2,513x_3 - 0,043x_1x_2 + 0,760x_1x_3 - 0,010x_2x_3 + 0,008x_1x_2x_3 \quad (1.20)$$

Интерпретация эффектов межфакторных взаимодействий осложнено, например, если в уравнении имеется слагаемое $+0,760x_1x_3$, то об увеличении выхода процесса на 0,76 ед. можно говорить при:

- 1) $x_1 = +1$ и $x_2 = +1$,
- 2) $x_1 = -1$ и $x_2 = -1$, если $x_1 = +1$ и $x_2 = -1$, или $x_1 = -1$ и $x_2 = +1$, то выход процесса будет уменьшаться на 0,76 ед.

Эффекты межфакторных взаимодействий, оценки которых являются нелинейными членами уравнения, будут играть всё более заметную роль в описании процесса при больших интервалах варьирования и по мере приближения исследуемого участка к вершине поверхности отклика из-за увеличения кривизны последней.

Полученное уравнение регрессии (1.20) необходимо проверить на значимость входящих в него коэффициентов. Незначимые коэффициенты, как правило, исключаются из уравнения регрессии, что никак не затронет остальные коэффициенты, т. к. проверка значимости каждого коэффициента проводится независимо.

Статистический анализ значимости оценок коэффициентов уравнения имеет своей целью показать с заранее заданной вероятностью $p(x)$, что полученные оценки коэффициентов по модулю либо больше (значимо отличаются от нуля), либо меньше ошибки в их определении (незначимо отличаются от нуля и должны быть исключены из уравнения)

Статистическую оценку значимости коэффициентов можно проводить двумя способами:

- построением доверительного интервала,
- проверкой по критерию Стьюдента.

При использовании полного факторного эксперимента или регулярных дробных реплик доверительные интервалы для всех коэффициентов (в том числе и эффектов взаимодействия) равны друг другу. Необходимо вначале вычислить дисперсию коэффициента регрессии, исходя из величины дисперсии опытов:

$$S_{b_j}^2 = \frac{S^2\{y\}}{N}, \quad (1.21)$$

Если параллельные опыты отсутствуют, то можно использовать выражение (1.21). Из формулы видно, что дисперсии всех коэффициентов равны друг другу, так как они зависят только от ошибки опыта и числа опытов.

Для определения значимости каждого коэффициента уравнения регрессии при заданном уровне значимости строится доверительный интервал:

$$\Delta b_j = \pm t_{\alpha, f} S_{b_j}, \quad (1.22)$$

где $t_{\alpha, f}$ – критерий Стьюдента при заданном уровне значимости α и числе степеней свободы f

Коэффициент значим, если его абсолютная величина больше доверительного интервала.

Для проверки по критерию Стьюдента используют формулу (1.23):

$$t_{\alpha, f} = \frac{|b_j|}{S_{b_j}} \quad (1.23)$$

Так как диагональные элементы ковариационной матрицы (1.14) равны между собой, поэтому коэффициенты уравнений (1.19 ÷ 1.20) определяются с одинаковой точностью:

$$S_{b_j} = \frac{S \{y\}}{\sqrt{N}} \quad (1.24)$$

$$S_{b_j} = \frac{14,254}{\sqrt{8}} = 1,335$$

$$t_0 = \frac{|b_0|}{S_{b_j}} = \frac{7,108}{1,335} = 3,750$$

$$t_1 = \frac{|b_1|}{S_{b_j}} = \frac{2,355}{1,335} = 1,764$$

$$t_2 = \frac{|b_2|}{S_{b_j}} = 0,139$$

$$t_3 = \frac{|b_3|}{S_{b_j}} = 1,882$$

$$t_{12} = \frac{|b_{12}|}{S_{b_j}} = 0,032$$

$$t_{13} = \frac{|b_{13}|}{S_{b_j}} = 0,569$$

$$t_{23} = \frac{|b_{23}|}{S_{b_j}} = 0,007$$

$$t_{123} = \frac{|b_{123}|}{S_{b_j}} = 0,006$$

Критерий Стьюдента на уровне значимости $\alpha = 0,1$ и с числом степеней свободы $f = 7$ имеет значение $t_{\alpha,f} = 1,42$, на основании этого и с учётом величины доверительного интервала:

$$\Delta b_j = \pm t_{\alpha,f} S_{b_j} = 1,42 \cdot 1,335 = 1,89$$

исключим незначимые коэффициенты из уравнения регрессии:

$$\hat{y} = 7,108 + 2,355x_1 + 2,513x_3$$

Если получен незначимый линейный коэффициент, то возможны следующие ситуации:

- данный фактор на исследуемый процесс не влияет;
- выбран слишком малый интервал варьирования, в связи с чем изменение выхода процесса, обусловленное изменением фактора, соизмеримо случайным отклонениям, вызываемым влиянием неучтенных факторов и погрешностью измерительных приборов;
- значение данного фактора в центре эксперимента соответствует вершине поверхности отклика, в связи с чем, одинаковое его уменьшение или увеличение понизит выход процесса приблизительно на одну и ту же величину.

§ 2. Оценка адекватности математической модели

Модель, с помощью которой успешно достигается поставленная цель, называется адекватной этой цели [37]. Оценка адекватности производится с помощью специально разработанных критериев.

1) Критерий Фишера.

Для проверки математической модели на адекватность вычисляется опытное значение критерия Фишера, которое сравнивается с теоретическим значением, взятом при заданном уровне значимости α .

Опытное значение критерия Фишера берётся равным отношению остаточной дисперсии (называемой также дисперсией адекватности) к общей дисперсии опытов (или дисперсии воспроизводимости):

$$F_{on} = \frac{S_{ocm}^2}{S^2\{y\}} \quad (1.25)$$

S_{ocm}^2 – остаточная дисперсия (дисперсия адекватности)

$S^2\{y\}$ – дисперсия опытов (дисперсия воспроизводимости):

$$S^2\{y\} = \frac{\sum_{u=1}^N (y_u - \bar{y})^2}{N-1} \quad (1.26)$$

N – число опытов,

\bar{y} – среднее значение функции отклика:

$$\bar{y} = \frac{1}{N} \sum_{u=1}^N y_u \quad (1.27)$$

k – число факторов.

$$S_{ocm}^2 = \frac{\sum_{u=1}^N (y_u - \hat{y}_u)^2}{N - (k+1)}. \quad (1.28)$$

Если теоретическое значение критерия Фишера $F_T > F_{on}$ – модель адекватно описывает реальный процесс, если $F_T < F_{on}$, то уравнение неадекватно описывает экспериментальные данные, точность описания процесса данным уравнением значимо ниже той точности, с которой получены экспериментальные результаты.

Если линейная модель, описываемая полиномом первого порядка, оказывается неадекватной, то необходимо либо менять параметр оптимизации, либо увеличивать степень полинома до тех пор, пока не будет достигнут критерий адекватности.

2) Критерий Кохрена.

Данный критерий оценки адекватности используется, если количество дисперсий больше двух и одна значительно больше остальных [10]:

$$G = \frac{S_{\max}^2}{\sum_{i=1}^N S_i^2}. \quad (1.29)$$

Если параллельные опыты отсутствуют, то оценку адекватности модели (уравнение 1.20) удобно проводить с помощью критерия Фишера:

$$S^2\{y\} = \frac{\sum_{u=1}^N (y_u - \bar{y})^2}{N-1} = \frac{99.78}{7} = 14,254;$$

$$S_{ocm}^2 = \frac{\sum_{u=1}^N (y_u - \hat{y}_u)^2}{N - (k + 1)} = \frac{97.29}{4} = 24,322.$$

Для регрессионной модели (1.20):

$$F_{on} = \frac{S_{ocm}^2}{S^2\{y\}} = 1,706.$$

Для регрессионной модели (1.25):

$$F_{on} = \frac{S_{ocm}^2}{S^2\{y\}} = \frac{24,341}{14,254} = 1,708.$$

Опытное значение критерия Фишера не превышает табличное ($F_T = 4,12$), следовательно, математическая модель адекватно описывает процесс переноса вещества в двухфазной системе ионит – раствор.

§ 3. Методы оптимизации эксперимента

Арсенал математических средств, используемых для решения задач оптимизации эксперимента, обширен так же, как и характер решаемых задач. Одни и те же методы могут быть применены для задач разных классов, одна и та же задача или задачи одного класса могут решаться разными методами. В наиболее простых задачах решение находится с помощью классических математических методов, например, дифференциального и вариационного исчисления. Однако при увеличении числа переменных и появлении дискретных функций [38] эти методы неприменимы. Решением экстремальных задач неклассического типа занимается *оптимальное* (математическое) *программирование*. Суть этих задач сводится к отысканию переменных (x_1, x_2, \dots, x_k) дающих *экстремум целевой функции* при известных условиях (ограничениях), заданных уравнениями или неравенствами.

Выбор метода решения – один из важнейших этапов оптимизации [30]. Можно выделить следующие группы методов:

- *аналитические* методы
- *методы математического программирования*.

Группа аналитических методов оптимизации объединяет аналитический поиск экстремума функции, метод множителей Лагранжа, вариационные методы и принцип максимума. Аналитический поиск экстремума функций, заданных без ограничений на независимые переменные, применяется к задачам, у которых оптимизируемая функция имеет аналитиче-

ское выражение, дифференцируемое во всем диапазоне исследования, а число переменных невелико. Это один из наиболее простых методов.

Аналитические методы основаны на классических методах математического анализа. Задача оптимизации формулируется следующим образом. Существует процесс, известна его математическая модель и установлен критерий оптимизации R в виде функции:

$$R = R(\bar{X}, \bar{U}) \quad (1.30)$$

либо функционала:

$$R = \int_0^t R(\bar{X}, \bar{U}) \quad (1.31)$$

где $\bar{U} = f(t)$

Заданы ограничения:

$$\bar{X} \in \bar{X}_{\text{дон}} \text{ и } \bar{U} \in \bar{U}_{\text{дон}} \quad (1.32)$$

Необходимо при заданных ограничениях найти такие значения \bar{U} , при которых R достигает экстремума. В случае функционала необходимо найти вид функции $\bar{U} = f(t)$, при которой R достигает экстремума.

Аналитические методы поиска экстремума критерия оптимальности применяются к задачам, у которых оптимизируемая функция имеет аналитическое выражение, а число переменных невелико.

В качестве примера рассмотрим определение оптимального времени процесса [28].

Для двух последовательных реакций $A \xrightarrow{k_1} B \xrightarrow{k_2} D$ (где k – константа скорости реакции) необходимо определить оптимальное время t_{opt} , при котором выход целевого продукта B будет достигать максимума.

Пусть a – начальная концентрация компонента A . В начальный момент времени концентрации компонентов B и D равны нулю: $C_B = C_D = 0$ при $t = 0$. Критерий оптимизации – выход целевого продукта $R = \frac{C_B}{a}$. Управляющее воздействие – время t . Характер изменения концентраций компонентов во времени приведен на (рис. 1.4).

Для математического описания нестационарных режимов обычно используются обыкновенные дифференциальные уравнения:

$$\frac{dC_A}{dt} + k_1 C_A = 0.$$

$$\begin{aligned}
 w_1 &= k_1 C_A, \\
 w_2 &= k_2 C_B, \\
 \frac{dC_B}{dt} &= w_1 - w_2.
 \end{aligned}
 \tag{1.33}$$

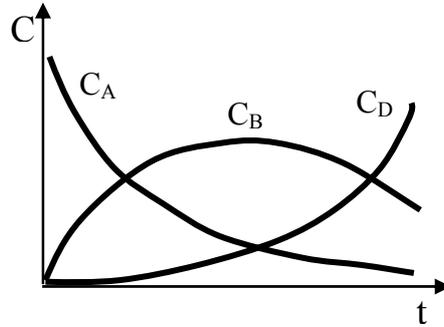


Рис. 1.4. Изменение концентрации компонентов смеси во времени

Изменение концентрации компонента A во времени (рис. 1.4) определится как:

$$-\frac{dC_A}{dt} = k_1 C_A, \quad \frac{dC_A}{C_A} = -k_1 dt.
 \tag{1.34}$$

Проинтегрировав выражение (1.34), получим:

$$\ln \frac{C_A}{a} = -k_1 t.
 \tag{1.35}$$

Тогда изменение концентрации компоненты B во времени определится как:

$$\frac{dC_B}{dt} = k_1 a e^{-k_1 t} - k_2 C_B
 \tag{1.36}$$

или

$$\frac{dC_B}{dt} + k_2 C_B = k_1 a e^{-k_1 t}.$$

Решая полученное уравнение, находим выражение для определения текущей концентрации компонента B :

$$C_B = a \frac{k_1}{k_2 - k_1} (e^{-k_1 t} - e^{-k_2 t}).
 \tag{1.37}$$

Тогда выход целевого продукта определится следующим образом:

$$R = \frac{C_B}{a} = \frac{k_1}{k_2 - k_1} (e^{-k_1 t} - e^{-k_2 t}). \quad (1.38)$$

Исследуем экстремум полученной целевой функции (1.38).

Из условия существования максимума:

$$\frac{dR}{dt} = 0, \quad \frac{d^2 R}{dt^2} < 0$$

найдем первую производную и приравняем ее нулю:

$$\frac{\partial R}{\partial t} = -\frac{k_1}{k_2 - k_1} (-k_1 e^{-k_1 t} - k_2 e^{-k_2 t}) = 0.$$

Решая уравнения (1.40), определим оптимальное время пребывания:

$$t_{opt} = \frac{1}{k_2 - k_1} \ln \frac{k_2}{k_1}. \quad (1.39)$$

Для проверки выполнения достаточного условия существования максимума вычисляем вторую производную:

$$\frac{\partial^2 R}{\partial t^2} = \frac{k_1}{k_2 - k_1} (k_1^2 e^{-k_1 t} - k_2^2 e^{-k_2 t}) < 0. \quad (1.40)$$

Так как вторая производная меньше нуля, то в данной точке существует максимум целевой функции R . Подставив (1.39) в (1.38), получим выражение для определения максимальной концентрации компонента B :

$$(C_B)_{opt} = a^2 \left(\frac{k_1}{k_2} \right)^{\frac{k_1}{k_2 - k_1}}. \quad (1.41)$$

Группа методов математического программирования включает линейное, нелинейное и динамическое программирование.

Математически задачу оптимизации можно записать в следующем виде:

$$f = (x_1, x_2, \dots, x_k) \rightarrow \max. \quad (1.42)$$

при ограничениях

$$\begin{aligned} g_1(x_1, x_2, \dots, x_k) &\leq 0; \\ g_2(x_1, x_2, \dots, x_k) &\leq 0; \\ g_n(x_1, x_2, \dots, x_k) &\leq 0; \\ x_k &\geq 0, (i = 1, 2, \dots, k) \end{aligned} \quad (1.43)$$

В общем случае методы математического программирования классифицируются в зависимости от свойств функции f и ограничений g .

Задачи, в которых функция (1.30) и ограничения (1.31) линейны, решаются методами линейного программирования [36].

Линейное программирование – метод для решения задач оптимизации с линейными выражениями для критерия оптимальности и линейными ограничениями на область изменения переменных. Подобные задачи решаются итерационными способами. Эти методы используются при оптимальном планировании процессов, производств, для транспортных задач и др.

Основным и универсальным методом линейного программирования является симплекс-метод. Некоторые классы линейных задач (например, широко распространенные транспортные задачи) решаются более простыми специальными методами (распределительным, венгерским, методом потенциалов).

В некоторых линейных задачах переменные x_k могут принимать лишь целочисленные значения. Такие задачи решаются методами *целочисленного программирования*.

Задачи, в которых функции или ограничения (полностью или хотя бы частично) нелинейны, изучаются нелинейным программированием. В этот класс входят задачи квадратичного программирования.

Методы нелинейного программирования объединяют различные способы решения оптимальных задач: *градиентные, безградиентные* и *случайного поиска*. Общим для методов нелинейного программирования является то, что их используют при решении задач с нелинейными критериями оптимальности. Все методы нелинейного программирования – это численные методы поискового типа [35]. Суть их в определении набора независимых переменных, дающих наибольшее приращение оптимизируемой функции. Данная группа методов применяется как для детерминированных, так и стохастических процессов.

Когда целевая функция (1.42) и ограничения (1.43) нелинейны и для поиска точки экстремума нельзя или очень сложно использовать аналитические методы решения, тогда для решения задач оптимизации применяются методы нелинейного программирования.

В основном методы нелинейного программирования могут быть охарактеризованы как многошаговые методы или методы последующего улучшения исходного решения. В этих задачах обычно заранее нельзя сказать, какое число шагов гарантирует нахождение оптимального значения с заданной степенью точности. Кроме того, в задачах нелинейного программирования выбор величины шага представляет серьезную проблему, от успешного решения которой во многом зависит эф-

эффективность применения того или иного метода. Разнообразие методов решения задач нелинейного программирования как раз и объясняется стремлением найти оптимальное решение за наименьшее число шагов.

Большинство методов нелинейного программирования используют идею движения в n -мерном пространстве в направлении оптимума. При этом из некоторого исходного или промежуточного состояния x^k осуществляется переход в следующее состояние x^{k+1} изменением вектора x^k на величину называемую шагом, т. е.

$$x^{k+1} = x^k + \Delta x^k. \quad (1.44)$$

В ряде методов шаг, т. е. его величина и направление определяется как некоторая функция состояния x^k

$$\Delta x^k = f(x^k). \quad (1.45)$$

Следовательно, согласно (1.44) новое состояние x^k , получаемое в результате выполнения шага (1.45) может рассматриваться как функция исходного состояния x^k :

$$x^{k+1} = x^k + f(x^k). \quad (1.46)$$

В некоторых методах Δx^k обусловлен не только состоянием x^k , но и рядом предшествующих состояний:

$$\Delta x^k = f(x^k, x^{k-1}, \dots, x^{k-2}). \quad (1.47)$$

$$x^{k+1} = x^k + f(x^k, x^{k-1}, \dots, x^{k-2}). \quad (1.48)$$

Естественно, что алгоритмы поиска типа (1.48) являются более общими и принципиально могут обеспечить более высокую сходимость к оптимуму, т. к. используют больший объем информации о характере поведения оптимальной функции.

В настоящее время для решения подобных задач разработано значительное число методов, однако нельзя отдать предпочтение какому-либо одному. Выбор метода определяется сложностью объекта и решаемой задачей оптимизации.

Методы нелинейного программирования в соответствии со способом определения шага поиска можно назвать *прямыми итеративными методами*.

Операционные задачи, имеющие случайные параметры, решаются методами *стохастического программирования*, а задачи, в которых переменные и ограничения могут меняться в определенных пределах – методами *параметрического программирования*.

В ряде случаев задачи имеют большую размерность, и их решение требует сложных вычислительных действий. В таких случаях прибега-

ют к *блочному программированию*, для чего задачу разбивают на ряд частных задач меньшей размерности. Блочное программирование в ряде случаев позволяет получить оптимальное или близкое к нему решение.

Динамическое программирование – эффективный метод решения задач оптимизации многостадийных процессов. Метод предполагает разбивку анализируемого процесса на стадии (во времени или в пространстве) – например, *колонна в каскаде* или *тарелка в колонне*. При использовании этого метода в отличие от классического подхода и линейного программирования решение многомерной многоэтапной задачи производится не сразу, а последовательно, выбором оптимального решения на каждом этапе. Таким образом, рассмотрение задачи начинается с последней стадии процесса, и оптимальный режим определяется постадийно. Решение на каждом отдельном этапе принимается с учетом всех его последствий в будущем, т. е. чтобы оно вело к максимальному успеху всей операции в целом. Динамическое программирование эффективно используется при оптимизации целевых функций любого типа, при требованиях целочисленности решения, в динамических системах (когда требуется учитывать фактор времени). Этим методом решаются задачи распределения, управления и упорядочения.

§ 3.1. Планирование экстремальных экспериментов

Экстремальным экспериментом называют эксперимент, направленный на отыскание экстремума функции регрессии.

При выборе метода экстремального планирования решающим фактором является стоимость экспериментов. Если стоимость экспериментов высока и их допустимое количество невелико, то наиболее целесообразным может являться статический план, согласно которому измерения проводятся в точках некоторой сетки в заданной подобласти факторного пространства. Часто в качестве планов проведения эксперимента используют последовательные планы. При некоторых предположениях эти планы обладают асимптотическим свойством сходимости получаемой последовательности точек к точке локального экстремума функции регрессии.

При постановке экспериментов с целью определения оптимальных условий проведения процессов на первом этапе исследований необходимо найти направление движения к области оптимума, т. е. необходимо определить экспериментально координаты экстремальной точки $(x_1^{opt}, x_2^{opt}, \dots, x_k^{opt})$ функции $f = (x_1, x_2, \dots, x_k)$.

Для решения этой задачи изучается поверхность отклика (геометрическая интерпретация оптимизируемой функции – рис. 1.4) на небольшом участке и составляется математическая модель процесса в ли-

нейном приближении. Затем осуществляется движение к оптимуму (например, методами крутого восхождения или симплексным). В области оптимума математические модели поверхности отклика представляются в виде полиномов второго, а иногда и третьего порядка.

Допустим, что задача оптимизации решена и экстремум найден. Построим контурные сечения $y = const$ поверхности отклика для $k = 2$ (рис. 1.5), т. е. проведем через точку y_{opt} в n -мерном пространстве двумерную плоскость.

Тогда при удалении от точки y_{opt} в любом направлении $y(x)$ будет изменяться. Если $y(x)$ является непрерывной функцией от x_{opt} , то вокруг точки y_{opt} всегда можно провести в данной плоскости замкнутую линию, на которой значение $y(x)$ будет постоянным. Таких замкнутых линий, называемых линиями постоянного уровня и отвечающих различным значениям $y(x)$, можно провести вокруг точки y_{opt} сколько угодно, причем каждая из этих линий (например, для случая поиска максимума) будет целиком охватывать любую линию постоянного уровня, для которой значение y_{opt} больше. Форма линий постоянного уровня, соответствующих различным значениям коэффициентов b_k , может при этом быть различной.

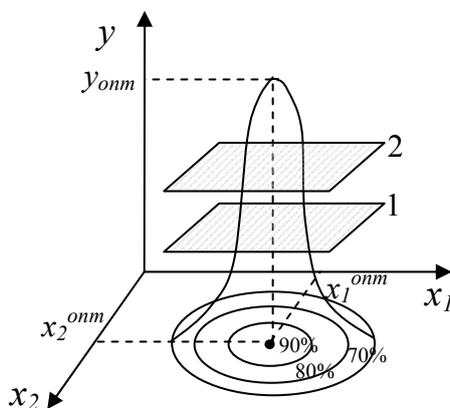


Рис. 1.5. Поверхность отклика.

x_1 и x_2 – факторы, y – величина функции отклика

Уравнение регрессии второго порядка, адекватно описывающее почти стационарную область, исследуют для определения координат оптимума. Кроме того, представляет интерес изучение свойств поверхности отклика в окрестности оптимума. При этом полезно перейти от полинома второго порядка, полученного по результатам опыта, к стандартному каноническому уравнению:

$$\hat{y} - y_s = \lambda_{11} X_1^2 + \lambda_{22} X_2^2 + \dots + \lambda_{kk} X_k^2, \quad (1.49)$$

где y_S – значение выхода в центре поверхности; X_1, X_2, \dots, X_k – канонические переменные, являющиеся линейными функциями факторов x_1, x_2, \dots, x_k ; $\lambda_1, \lambda_2, \dots, \lambda_k$ – коэффициенты канонической формы.

Первый этап канонического преобразования – перенос начала координат в особую точку поверхности отклика – центр поверхности. Координаты центра S определяются решением системы уравнений:

$$\frac{\partial y}{\partial x_1} = 0, \quad \frac{\partial y}{\partial x_2} = 0, \quad \dots, \quad \frac{\partial y}{\partial x_k} = 0. \quad (1.50)$$

При аппроксимации поверхности отклика полиномом второго порядка приходится решать систему k линейных уравнений. Если определитель этой системы равен нулю, то поверхность не имеет центра. В этом случае можно или перенести начало координат в точку с наилучшим значением выхода, или совсем не переносить центр. При этом для нецентральной поверхности:

$$\hat{y} - y_S = \sum_{i=1}^N \lambda_k X_k + \sum_{i=1}^N \lambda_{kk} X_k^2$$

оптимум будет лежать на границе области определения факторов.

Если поверхность имеет центр:

$$\hat{y} - y_S = \sum_{i=1}^N \lambda_{kk} X_k^2,$$

то в него переносят начало координат. При этом в уравнении поверхности исчезают члены, содержащие линейные эффекты, и изменяется свободный член. Коэффициенты при вторых степенях и взаимодействия инвариантны относительно переноса. Второй этап – поворот координатных осей в новом центре таким образом, чтобы исчезли члены с эффектами взаимодействия; свободный член инвариантен относительно поворота. В результате получим уравнение вида (1.49). Поверхности второго порядка классифицируются по их каноническим формам (рис. 1.6).

- Все коэффициенты канонической формы имеют одинаковые знаки. Поверхность – эллиптический параболоид (рис. 1.6, а). В центре поверхности максимум при $\lambda_{ii} < 0$ и минимум – при $\lambda_{ii} > 0$.
- Коэффициенты имеют разные знаки. Поверхность – гиперболический параболоид, «седло» (рис. 1.6, б). В центре поверхности «минимум».
- Один или несколько (но не все) коэффициентов близки к нулю. При этом центр лежит далеко за областью экспериментирования.

Поверхность такого типа называется «возрастающим возвышением» («гребнем»).

- Возможен еще вырожденный случай параллельных плоскостей, который не представляет практического интереса (рис. 1.6, в).

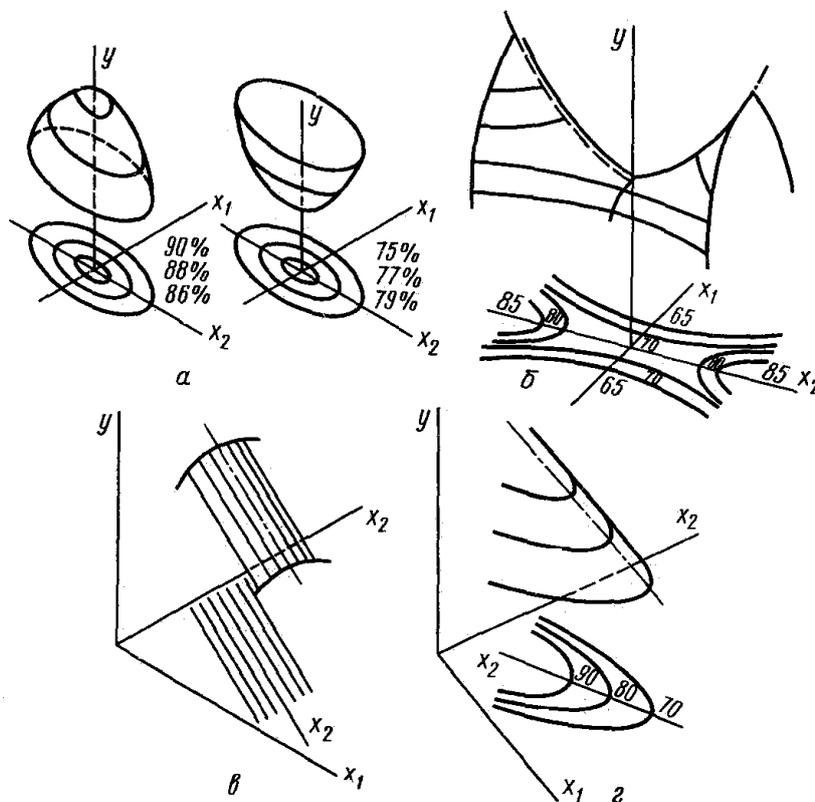


Рис. 1.6. Поверхности отклика и их сечения

При $\lambda_{22} = 0$ (рис. 1.6, г), перенеся начало координат в точку S (обычно вблизи центра плана), получаем уравнение параболы:

$$\hat{y} - y_S = \lambda_{11} X_1^2 + \lambda_{22} X_2. \quad (1.51)$$

Перейдем от уравнения регрессии второго порядка для $k = 2$, полученного по экспериментальным данным

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2 \quad (1.52)$$

к каноническому уравнению (1.49). Определим координаты точки S – центра поверхности. Для этого необходимо решить систему уравнений:

$$\begin{aligned} \frac{\partial y}{\partial x_1} = 0, \quad 2b_{11} X_1 + b_{12} X_2 + b_1 = 0; \\ \frac{\partial y}{\partial x_2} = 0, \quad b_{12} X_1 + 2b_{22} X_2 + b_2 = 0. \end{aligned} \quad (1.53)$$

Решение, системы (1.53) дает координаты центра x_{1S} и x_{2S} .

Подставив их в уравнение регрессии (1.52), получим значение выходной величины в точке $S - y_S$. Перенесем начало координат в точку S (y_S, x_{1S}, x_{2S}). Старые координаты (x_1, x_2, y) связаны с новыми (x_1', x_2', y') соотношениями:

$$\begin{aligned}x_1 &= x_{1S} + x_1', \\x_2 &= x_{2S} + x_2', \\y &= y_S + y' .\end{aligned}\tag{1.54}$$

В новой системе координат уравнение (1.52) примет вид:

$$\hat{y} - y_S = b_{11} (X_1')^2 + b_{22} (X_2')^2 + b_{12} X_1' X_2' .\tag{1.55}$$

Второй этап преобразования предполагает при помощи поворота осей координат исключение эффекта взаимодействия. Для этого необходимо повернуть оси координат на такой угол α , чтобы

$$\operatorname{ctg}(2\alpha) = \frac{b_{11} - b_{22}}{b_{12}} .\tag{1.56}$$

Тогда получим в новой системе координат (X_1, X_2, y) уравнение регрессии в каноническом виде:

$$\hat{y} - y_S = \lambda_{11} X_1^2 + \lambda_{22} X_2^2 .\tag{1.57}$$

Старые координаты (x_1, x_2) связаны с новыми соотношениями:

$$\begin{aligned}x_1 &= (X_1 + x_{1S}) \cos \alpha - (X_2 + x_{2S}) \sin \alpha, \\x_2 &= (X_1 + x_{1S}) \sin \alpha + (X_2 + x_{2S}) \cos \alpha.\end{aligned}\tag{1.58}$$

Для определения коэффициентов канонического уравнения λ_{11} и λ_{22} воспользуемся следующими двумя инвариантами уравнения (функциями коэффициентов, которые не изменяют своего значения при любом преобразовании координат):

$$\begin{aligned}I_1 &= b_{11} + b_{22} = \operatorname{const}, \\I_2 &= \begin{vmatrix} b_{11} & \frac{1}{2} b_{12} \\ \frac{1}{2} b_{12} & b_{22} \end{vmatrix} = \operatorname{const}.\end{aligned}\tag{1.59}$$

С учётом исходного уравнения регрессии (1.57), получаем:

$$\begin{aligned} b_{11} + b_{22} &= \lambda_{11} + \lambda_{22}, \\ b_{11}b_{22} - \frac{1}{4}b_{12}^2 &= \lambda_{11}\lambda_{22} - \frac{1}{4}\lambda_{12}^2. \end{aligned}$$

Определим λ_{11} и λ_{22} как корни квадратного уравнения:

$$\lambda^2 - (b_{11} + b_{22})\lambda + \left(b_{11}b_{22} - \frac{1}{4}b_{12}^2 \right) = 0. \quad (1.60)$$

В многомерных задачах каноническое преобразование осуществляется методами линейной алгебры. Составим из коэффициентов уравнения регрессии второго порядка

$$\hat{y} = b_0 + b_1x_1 + \dots + b_kx_k + b_{12}x_1x_2 + \dots + b_{k-1,k}x_{k-1}x_k + b_{11}x_1^2 + \dots + b_{kk}x_k^2,$$

полученного по эксперименту, квадратную матрицу:

$$B = \begin{bmatrix} b_{11} & \frac{1}{2}b_{12} & \dots & \frac{1}{2}b_{1k} \\ \frac{1}{2}b_{21} & b_{22} & \dots & \frac{1}{2}b_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{2}b_{k1} & \frac{1}{2}b_{k2} & \dots & b_{kk} \end{bmatrix}. \quad (1.61)$$

в которой $b_{ji} = b_{ij}$. Для определения коэффициентов $(\lambda_{11}, \dots, \lambda_{kk})$ уравнения регрессии в каноническом виде (1.49) необходимо найти корни характеристического полинома $P_k(\lambda)$ матрицы B :

$$P_k(\lambda) = |B - \lambda E|, \quad (1.62)$$

здесь E – единичная матрица, тогда

$$P_k(\lambda) = \begin{bmatrix} b_{11} - \lambda & \frac{1}{2}b_{12} & \dots & \frac{1}{2}b_{1k} \\ \frac{1}{2}b_{21} & b_{22} - \lambda & \dots & \frac{1}{2}b_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{2}b_{k1} & \frac{1}{2}b_{k2} & \dots & b_{kk} - \lambda \end{bmatrix}. \quad (1.63)$$

Проведём расчёт на примере определения максимальной степени разложения флотационного концентрата (y , %) в зависимости от температуры и примесей. Математическая модель процесса имеет вид уравнения регрессии второго порядка:

$$\hat{y} = 35,4 + 4,51x_3 - 1,3x_5 - 1,5x_1^2 + 2,66x_2^2 - 1,47x_3^2 + 1,61x_1x_4.$$

В качестве факторов были выбраны z_1 – температура процесса, °С; z_2, z_3, z_4, z_5 – концентрации примесей (мас. %): MgO, SO₃, Al₂O₃ и F соответственно, границы значений которых представлены в табл. 1.4.

Таблица 1.4

Интервалы и уровни варьирования переменных

Величина	z_1	z_2	z_3	z_4	z_5
Основной уровень	50	2,1	2,0	1,33	0,75
Интервал варьирования	20	0,9	1,0	0,37	0,25
+2	90	3,9	4,0	2,07	1,25
-2	10	0,3	0,0	0,59	0,25

Требуется определить условия, соответствующие максимальной степени разложения (y_{max}), при ограничениях, накладываемых сферой с радиусом, равным звездному плечу ($\alpha = \pm 2$).

Для определения условий максимальной степени разложения переменные, характер влияния которых ясен из уравнения регрессии, принимаем равными: $x_2 = +2$; $x_5 = -2$. Влияние концентрации SO₃ представлено в уравнении положительным линейным и отрицательным квадратичным членами. Оптимальное значение x_3 равно 1,533, определяем из условия экстремального значения y по x_3 . При этих значениях факторов x_2, x_3, x_5 уравнение регрессии примет вид:

$$\hat{y} = 52,12 - 1,5x_1^2 + 1,61x_1x_4.$$

Координаты центра поверхности S определяются следующим образом:

$$\frac{\partial \hat{y}}{\partial x_1} = 2 \cdot 1,5 \cdot x_1 + 1,61x_4 = 0;$$

$$\frac{\partial \hat{y}}{\partial x_4} = 1,61x_1 = 0;$$

$$x_{1S} = 0; \quad x_{4S} = 0; \quad y_S = 52,12.$$

Таким образом, центр поверхности совпадает с центром плана. Характеристический полином определится как

$$P_k(\lambda) = \begin{bmatrix} 1,5 - \lambda & 0,805 \\ 0,805 & 0 - \lambda \end{bmatrix};$$

$$P_k(\lambda) = \lambda^2 + 1,5\lambda - 0,64 = 0.$$

Корни полинома $\lambda_1 = +0,35$, $\lambda_2 = -1,85$. Уравнение в канонической форме

$$\hat{y} - 52,12 = 0,35 X_1^2 - 1,85 X_4^2.$$

Поверхность отклика – гиперболический параболоид. В сечениях поверхности отклика плоскостями $y = \text{const}$ – гиперболы. В центре поверхности «минимакс». Линейное преобразование задается системой:

$$X_1 = 0,92 x_1 + 0,39 x_4,$$

$$X_4 = -0,39 x_1 + 0,92 x_4.$$

Для определения максимальной степени разложения выходим из «минимакса» по оси X_1 (коэффициент канонической формы положительный), приравняв X_4 равным нулю:

$$X_1 = \pm \sqrt{\frac{\hat{y} - 52,12}{0,35}}; \quad X_4 = 0.$$

Увеличивая \hat{y} , проверяем при этом выполнение условий $x_1 = x_4 \leq 2$. Максимальная величина степени разложения получилась равной 53,5 % ($x_1 = \pm 1,82$; $x_4 = \pm 0,795$).

При увеличении y до 54 % значение $x_1 > 2$. В полученных оптимальных условиях ($x_1 = +1,82$; $x_2 = +2$; $x_3 = +1,533$; $x_4 = +0,795$; $x_5 = -2$) и ($x_1 = -1,82$; $x_2 = +2$; $x_3 = +1,533$; $x_4 = -0,795$; $x_5 = -2$) были поставлены контрольные опыты. Степень разложения получилась соответственно равной 55,8 и 53,7 %. Таким образом, расхождения с расчётными данными лежат в пределах ошибки эксперимента ($S\{y\} = 2,1$).

Если процесс описывается несколькими уравнениями регрессии, приходится решать компромиссную задачу – определять экстремальное значение одной функции отклика при ограничениях, накладываемых другими функциями отклика и границами области исследования.

§ 3.2. Симплексное планирование экспериментов при определении оптимальных условий разделения веществ

Этот метод многофакторного планирования эксперимента возник в 1962 г. Он предусматривает многошаговый процесс движения к экстремуму поверхности отклика с одновременным, если это необходимо, описанием многофакторной линейной зависимостью соответствующей части поверхности отклика.

Название метода объясняется тем, что после назначения центра эксперимента и интервалов варьирования факторов рассчитывают условия опытов плана, соответствующие вершинам симплекса, расположенного в n -мерном факторном пространстве и имеющего центр тяжести в центре эксперимента ($z_{i0}; x_{i0} = 0$).

Под n -мерным симплексом подразумевают выпуклую геометрическую фигуру в n -мерном факторном пространстве, имеющую $(n + 1)$ вершин [18], соединенных прямыми отрезками, называемыми рёбрами. Если вершины находятся друг от друга на одинаковом расстоянии, то такой симплекс называют правильным. Любые n вершин комплекса лежат в одной гиперплоскости. Часть такой гиперплоскости, ограниченной ребрами, будет называться гранью симплекса, противоположной вершине, не лежащей в этой гиперплоскости.

Одномерным симплексом будет отрезок прямой, двумерным – плоский треугольник, трехмерным – тетраэдр, и т. д.

Число опытов симплексного плана равно числу оценок коэффициентов линейного уравнения:

$$N = n + 1 \quad (1.69)$$

Следовательно, этот план можно назвать насыщенным, позволяющим получить оценки лишь линейного уравнения.

Симплексный план записывают, используя безразмерное выражение величины фактора, а затем переходят к натуральной размерности факторов. Любое ребро, параллельное какой-либо оси z_{i0} , принимается равным соответствующему интервалу варьирования. Тогда одномерный симплекс будет иметь два опыта с координатами:

$$z_1 = z_0 - \frac{\Delta}{2} = +0,5 \text{ и } z_2 = z_0 + \frac{\Delta}{2} = +0,5$$

или в безразмерном выражении факторов

$$x_1 = \frac{z_1 - z_0}{\Delta} = +0,5, \quad x_2 = \frac{z_2 - z_0}{\Delta} = +0,5.$$

Рассмотрим, когда на процесс влияют два фактора, симплексом будет равносторонний треугольник (рис. 1.7). Здесь сплошными линиями изображен исходный симплексный план двухфакторного эксперимента, причем одна из вершин симплекса лежит на одной из осей координат, перпендикулярной противоположной грани (в данном случае – ребру).

Начало координат $X_1O X_2$ соответствует назначенному центру эксперимента z_{i0} или $x_{i0} = 0$. Нумерация опытов произвольная, в данном случае она осуществляется против часовой стрелки.

Координаты вершин правильного симплекса для любого числа факторов можно задать табл. 1.5. Для пользования табл. 1.6 необходимо выделить из нее n столбцов и $(n+1)$ строку.

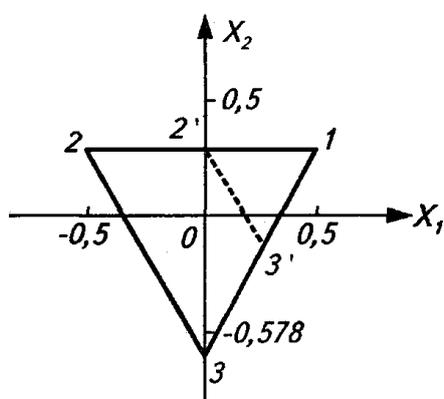


Рис. 1.7. Симплексный двухфакторный план

За основной уровень (центр плана), обозначаемый символом «0» принимаются значения факторов, соответствующих наиболее благоприятным из известных режимов оптимизируемого процесса.

Таблица 1.5

Координаты вершин правильного симплекса для кодированных переменных

№ опыта	x_1	x_2	...	x_{n-1}	x_n	Функция отклика
1	k_1	k_2	...	k_{n-1}	k_n	y_1
2	$-R_1$	k_2	...	k_{n-1}	k_n	y_2
3	0	$-R_2$...	k_{n-1}	k_n	y_3
...
$n-1$	0	0	...	k_{n-1}	k_n	y_{n-1}
n	0	0	...	$-R_{n-1}$	k_n	y_n
$n+1$	0	0	...	0	R_n	y_{n+1}

Центр симплекса, приведенного в табл. 1.5, совпадает с началом координат (в кодированных переменных), а сторона симплекса равна

единице. Результаты расчетов, выполненных на основании табл. 1.5, рис. 1.7 и формул (1.70), (1.71) приведены в табл. 1.6.

Таблица 1.6

Матрица исходного симплекса

№ опыта	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	0,5	0,289	0,204	0,158	0,129	0,109	0,0945
2	-0,5	0,289	0,204	0,158	0,129	0,109	0,0945
3	0	-0,578	0,204	0,158	0,129	0,109	0,0945
4	0	0	-0,612	0,158	0,129	0,109	0,0945
5	0	0	0	-0,632	0,129	0,109	0,0945
6	0	0	0	0	-0,645	0,109	0,0945
7	0	0	0	0	0	-0,655	0,0945
8	0	0	0	0	0	0	-0,661

Здесь первый столбец для 2-х факторов (рис. 1.7) $x_{11} = +0,5$ и $x_{12} = -0,5$. При заполнении второго столбца x_{2u} рассчитывают высоту треугольника:

$$h = \sqrt{1 - 0,5^2} = 0,866$$

Поскольку центр тяжести треугольника отстоит от основания на $\frac{1}{3}h$, то $x_{21} = x_{22} = 0,289$ и $x_{23} = -0,578$.

При исследовании влияния 3-х факторов на процесс опыты следует расположить в 4-х вершинах правильного тетраэдра. Четвертая вершина проецируется в центр эксперимента, а горизонтальная плоскость $X_1O X_2$ будет проходить через центр тяжести тетраэдра. Тогда координаты x_{1u} и x_{2u} первых трех опытов останутся прежними, а координаты $x_{14} = 0$, $x_{24} = 0$. Несложные геометрические расчеты дают возможность заполнить и третий столбец таблицы:

$$x_{31} = x_{32} = x_{33} = 0,204; x_{34} = -0,612.$$

Для четырех и большего числа факторов координаты последнего фактора в опытах плана можно рассчитать по формулам:

Расчет остальных значений k_i и R_i , входящих в табл. 1.5, осуществляется по формулам:

$$k_i = \sqrt{\frac{1}{2i(i+1)}} \quad (1.70)$$

$$R_i = i \cdot k_i, i = 1, 2, 3, \dots, n, \quad (1.71)$$

где i – номер фактора в матрице планирования.

В основе использования симплекса лежит следующее его важное свойство: из любого симплекса можно, отразив одну из вершин и используя оставшуюся грань, получить новый симплекс, добавив всего лишь одну точку, являющуюся зеркальным отображением отбрасываемой вершины (рис. 1.8).

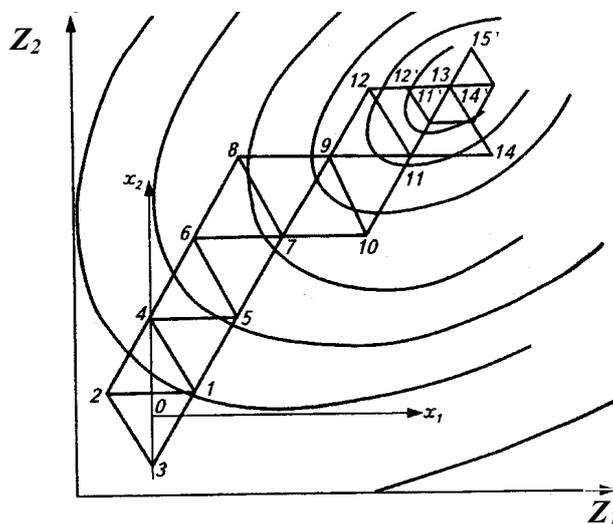


Рис. 1.8. Схема оптимизации эксперимента

Путем последовательного отбрасывания вершин с минимальным значением параметра оптимизации можно перемещать симплекс в факторном пространстве в направлении экстремума функции отклика, причем это перемещение будет происходить с каждым экспериментом. Иногда может случиться, что значение отклика в новой вершине при построении симплекса наименьшее. В этом случае следует вернуться к исходному симплексу, отражая вторую наименьшую вершину (точки 6 ÷ 10 рис. 1.8). Процедура отражения вершин и построения симплексов продолжается до достижения почти стационарной области.

При достижении оптимальной области симплекс начинает вращаться вокруг точки оптимума (точка 13 рис. 1.8). Точку, вокруг которой вращается симплекс, используют как центр при постановке экспериментов по планам второго порядка.

Следует отметить, что при вращении системы симплексов вокруг одной вершины целесообразно повторить опыт в данной вершине. Здесь либо эта вершина лежит недалеко от экстремума и тогда следует:

- 1) перейти к планированию второго порядка,
- 2) уточнить результат.
- 3) продолжить адаптивную оптимизацию, уменьшив размеры симплекса в $\delta' = \frac{1}{\delta}$ раз.

Во втором случае после уточнения значения отклика в этой вершине симплекса следует продолжить процедуру шагового восхождения с последовательным отбрасыванием наихудших точек.

Если уменьшить симплекс 11–12–13, изображённый на рис. 1.7. в 2 раза ($\delta' = 2, \delta = 0,5$) стянув его к лучшей точке $S = 13$, то координаты i -й вершины уменьшенного симплекса рассчитываются по формуле:

$$x'_{iu} = (1 - \delta) x_{is} + \delta x_{iu}, \quad (1.72)$$

x_{is} – i -я координата лучшей вершины, x'_{iu} – i -я координата уменьшенного симплекса n -м опыте.

Условия опыта, давшего самый хороший результат можно считать оптимальным на данном этапе исследований. Важно отметить, что наибольшее количество опытов ($n + 1$) приходится ставить лишь в начале эксперимента. Затем на каждом шаге оптимизации выполняется только один опыт.

С помощью табл. 1.6, приступая к оптимизации, необходимо рассчитать матрицу исходной серии опытов в натуральных (физических) переменных. Переход от кодированных переменных к натуральным производится по формуле:

$$x_i = x_{oi} + \Delta x_i \cdot x_i, \quad (1.73)$$

где x_{oi} – основной уровень, Δx_i – интервал варьирования, x_i – безразмерная (кодированная переменная).

Процедура шагового восхождения производится только с натуральными (физическими) переменными. Условия каждого нового опыта рассчитываются по формулам:

$$x_i^{n+2} = 2x_{ic} - x'_{ij}, \quad i = 1, 2, \dots, n, \quad (1.74)$$

где x_i^{n+2} – i -я координата точки, получаемой в результате отражения наихудшей точки, x_{ic} – i -я координата центра противоположной грани, x_{ij} – i -я координата наихудшей точки, j – номер «самого плохого» опыта.

Координата центра грани, лежащей против наихудшей i -й точки, находится по формуле:

$$x_{ic} = \frac{\sum_{u=1}^{N+1} x_{iu}}{k}, \quad k \neq j, \quad (1.75)$$

где k – число факторов в матрице планирования, i – номер фактора.

В процессе оптимизации, осуществляемом симплексным методом, можно в любое время и на любом шаге добавить в экспериментирова-

ние новый фактор, который до тех пор оставался на постоянном уровне, вводя по существу только одну точку (опыт). При этом значения всех ранее рассматриваемых факторов рассчитываются по формуле:

$$x_i = \frac{1}{n+1} \sum_{j=1}^{u+1} x_{ij}, \quad (1.76)$$

где $i = 1, 2, 3, \dots, n$, т. е. являются средними арифметическими значениями соответствующих координат предыдущего симплекса.

Значение вновь вводимого фактора определяется по формуле:

$$x_{n+1} = x_{0(n+1)} + \Delta x_{n+1} (R_{n+1} + k_{n+1}), \quad (1.77)$$

где $x_{0(n+1)}$ – основной уровень этого фактора,

Δx_{n+1} – интервал варьирования для данного фактора,

R_{n+1}, k_{n+1} – величины, рассчитываемые по формулам (1.71) и (1.70).

Следует отметить, что при введении нового фактора в состав полного факторного эксперимента количество опытов удвоится. В этом состоит одно из преимуществ симплексного метода планирования.

Рассмотрим пример использования симплексного метода для оптимизации процесса, когда на качество продукта (y) влияли 5 факторов, основные уровни и интервалы варьирования, которых представлены в табл. 1.7.

Таблица 1.7
Основные уровни и интервалы варьирования $k = 5$

№ П/П	Факторы	Основной уровень	Интервал варьирования
1	τ	2,0 [время процесса, ч.]	0,2
2	α	0,65 [содержание вещества А, мол. доли]	0,15
3	β	0,10 [содержание вещества В, мол. доли]	0,025
4	γ	0,25 [содержание вещества С, мол. доли]	0,05
5	λ	1,2 [молярное соотношение веществ С и А]	0,2

Для определения оптимальных условий эксперимента составим матрицу планирования в натуральных переменных (табл. 1.8). Чтобы зеркально отобразить точки находится координата центра грани (формула 1.75): $x_{1c} = \frac{4 \times 2 + 2,1}{5} = 2,02$, тогда $x_{2c} = 0,641$; $x_{3c} = 0,099$; $x_{4c} = 0,248$; $x_{5c} = 1,195$.

Отражённая точка 7 с координатами (табл. 1.8):
 $x_1^7 = 2 \cdot 2,02 - 1,9 = 2,14$ и т. д., образует новый 6-мерный симплекс 1–3÷7, после реализации восьмого опыта необходимо провести анализ результатов.

Таблица 1.8

Симплексный план эксперимента

№ опыта	Факторы					у (функция отклика)
	τ	α	β	γ	λ	
1	2,1	0,693	0,105	0,258	1,255	0,760
2	1,9	0,693	0,105	0,258	1,255	0,491
3	2,0	0,564	0,105	0,258	1,255	0,513
4	2,0	0,650	0,085	0,258	1,255	0,675
5	2,0	0,650	0,1	0,218	1,255	0,693
6	2,0	0,650	0,1	0,250	1,075	0,666
7	2,14	0,589	0,093	0,238	1,165	0,810
8	2,04	0,633	0,098	0,147	1,190	

Если линейное уравнение не может с достаточной точностью описать исследуемый участок поверхности отклика, то следует применить более сложный план, который позволит получить более точное уравнение для оптимизации процесса (планы второго порядка, многоуровневые планы), либо сменить параметр оптимизации.

В процессах разделения изотопов и очистки веществ в качестве критериев оптимизации применяют однократный коэффициент разделения (α), коэффициент обогащения ($\varepsilon = \alpha - 1$), ВЭТС, степень разделения (q), относительный выход обогащённого продукта, относительное содержание выделяемого компонента в отборе и отвале. Однако данные критерии могут дать неполную характеристику процесса, поэтому наряду с однозначными критериями используют обобщённые критерии оптимизации.

При разработке обобщённых параметров оптимизации учитывают тот факт, что с увеличением обогащения продукта его себестоимость и ценность возрастают не пропорционально относительной концентрации, а по более сложной зависимости. Относительная концентрация может входить в обобщённый критерий в виде степенной, либо логарифмической зависимости. Например, для описания процесса разделения изотопов в системе ионит–раствор применяют критерий Мажумдара [12]:

$$G_1 = \frac{\alpha c^2}{c_0^2} \quad (1.78)$$

где α – относительный выход обогащённого продукта, c – текущая концентрация, c_0 – начальная концентрация.

§ 3.3. Оптимизация разделительных процессов методом крутого восхождения

Одним из наиболее распространенных методов поиска оптимальной области при изучении химических и технологических процессов методами математической статистики, является метод крутого восхождения или так называемый метод движения по градиенту, предложенный Д. Боксом и К. Уилсоном в 1951 г.

Метод крутого восхождения – это такое движение в факторном пространстве в направлении градиента, при котором переход от точки к точке (в отличие от симплексного) сопровождается одновременным изменением значений всех факторов.

Для начала в интересующей области факторного пространства наилучший результат (полученный согласно априорной информации) принимается за основной уровень [12] и ставится полный факторный эксперимент (ПФЭ) или дробная реплика (ДФЭ) от него (в зависимости от числа факторов), при этом выбирается относительно узкий интервал варьирования переменными с целью получения адекватного линейного уравнения регрессии.

Линейная модель описывает наклонную гиперплоскость в $(n + 1)$ -мерном факторном пространстве. Однако линейное уравнение отражает с достаточной точностью поверхность отклика лишь в некоторой локальной области, соответствующей изученному диапазону изменения факторов. Характер поверхности отклика на некотором удалении от изученного участка всё больше отходит от плоскости, описанной линейным уравнением.

Если найти направление самого крутого подъема (градиента) этой плоскости и двигаться в этом направлении, осуществляя опыты при соответствующих условиях, за пределы изученной области, то можно найти такое сочетание значений факторов, которое будет соответствовать началу снижения поверхности отклика (процесс идет по типу ингибирования) или постепенной стабилизации получаемых результатов (процесс идет по типу насыщения). Определение такого сочетания факторов, оптимального на данном этапе исследования, можно осуществить лишь экспериментальным путем, запланировав *в направлении градиента*, найденного по полученному уравнению, серию опытов.

Градиентом (*grad y*) скалярной функции $y = f(x_1, x_2, \dots)$ называют вектор, перпендикулярный линии (поверхности, гиперповерхности) равного выхода, указывающий направление наибо́льшего изменения (роста или падения) скалярной функции

$$\text{grad } y = \frac{\partial y}{\partial x_1} \vec{i} + \frac{\partial y}{\partial x_2} \vec{j} + \dots + \frac{\partial y}{\partial x_i} \vec{u}, \quad (1.79)$$

где $\frac{\partial y}{\partial x_i}$ – угловой коэффициент b_i наклона поверхности отклика в рассматриваемой точке к соответствующей оси; $\vec{i}, \vec{j}, \vec{u}$ – единичные векторы (орты) в положительном направлении координатных осей (их модуль равен единице).

Это выражение задает кратчайшее направление движения к экстремуму. Следовательно, изменяя факторы пропорционально величинам и знакам коэффициентов регрессии, можно осуществить движение в направлении градиента функции отклика, т. е. по самому крутому пути. Поэтому процедура движения к почти стационарной области и называется крутым восхождением. Крутое восхождение представляют в натуральных переменных.

Изменение факторов при расчете серии опытов программы оптимизации будет пропорционально составляющим вектора градиента, т. е.

$$\Delta x_1 : \Delta x_2 \dots \Delta x_i = \frac{\partial y}{\partial x_1} \vec{i} : \frac{\partial y}{\partial x_2} \vec{j} \dots \frac{\partial y}{\partial x_i} \vec{u}, \quad (1.80)$$

где Δx_i – изменение i -й координаты последующего опыта по сравнению с координатой предыдущего.

Для линейного уравнения соотношение (1.80) запишется следующим образом:

$$\Delta x_1 : \Delta x_2 \dots \Delta x_i = b_1 : b_2 \dots b_i, \quad (1.81)$$

где b_i – оценки линейных эффектов факторов с соответствующими знаками.

Программы оптимизации представляет собой последовательность следующих операций:

- 1) за начало программы оптимизации ($u = 0$) принимают условие опыта в центре эксперимента $x_i = 0$;
- 2) назначают величину Δx_i для так называемого базового фактора, которым может быть любой фактор, в том числе и первый;
- 3) принимают ограничение: $|\Delta x_i| \leq 0,7$, обычно назначают $|\Delta x_i| = 0,5$ с знаком коэффициента b_i , слишком большое изменение какого-либо фактора от опыта к опыту программы оптимизации создают опасность проскочить самую высокую точку на линии пересечения градиента функции с поверхностью отклика, а слишком малое изме-

нение может потребовать большого числа опытов для достижения самой высокой точки;

- 4) изменения других факторов рассчитывают, соизмеряя это изменение с базовым фактором, например, первым:

$$\begin{aligned} \Delta x_2 : \Delta x_1 &= b_2 : b_1; \Delta x_2 = \Delta x_1 \frac{b_2}{b_1}; \\ \Delta x_3 : \Delta x_1 &= b_3 : b_1; \Delta x_3 = \Delta x_1 \frac{b_3}{b_1}. \end{aligned} \quad (1.82)$$

Если по какому-либо другому фактору это изменение $|\Delta x_i| > 0,7$ следует изменить исходную величину $|\Delta x_i| = 0,5$.

Условия опытов программы оптимизации в безразмерном выражении и в натуральной размерности факторов рассчитывают по формулам:

$$x_{iu} = x_{i0} + u \Delta x_i; \quad z_{iu} = z_{i0} + u \delta_i z_i, \quad (1.83)$$

где $u = 0 \div N$ – номер опыта программы оптимизации.

- 5) Составить программу оптимизации из $5 \div 6$ опытов и реализовать её в рамках одной серии опытов при том же числе повторностей m .

Если последние опыты программы дали устойчивое снижение результатов (процесс по типу ингибирования) или их стабилизацию (процесс по типу насыщения), программа оптимизации признается состоявшейся, за оптимальные на данном этапе исследования принимают условия опыта, давшего самый высокий результат y_u .

Метод формализуется в виде таблицы и основывается на тех же свойствах градиента функции [16]. На примере оптимизации двухфакторного процесса приведён расчёт программы в табл. 1.9:

- 1) В таблице записывают известные значения b_i , δ_i и произведения $b_i \delta_i$, при этом учитываются знаки при коэффициентах b_i .
- 2) Выбирают базовый фактор. Пусть базовым фактором будет первый; назначают Δx_1 со знаком b_1 .
- 3) Рассчитывают величину

$$K_i = \frac{|b_i \delta_i|}{|b_i \delta_i|_{\max}}, \quad (1.84)$$

где $b_i \delta_i$ – произведение для базового j -го фактора.

- 4) Рассчитывают шаг изменения других факторов $\Delta z_i = k_i \Delta z_j$, т. е. $\Delta z_2 = k_2 \Delta z_1$.

- 5) Назначают координаты исходной точки программы оптимизации. Она должна быть в центре эксперимента z_{i0} , $u = 0$.
- 6) Рассчитывают значения факторов для каждого опыта программы оптимизации. Координаты каждой новой точки получаются последовательным алгебраическим сложением нулевого уровня с пропорциональным приращением по каждому фактору: $z_{iu} = z_{i0} + u \delta_i z_i$. Чтобы движение осуществлялось по градиенту, необходимо сохранять соотношение произведений коэффициентов уравнения регрессии на соответствующие интервалы, т. е. знак при $u \delta_i z_i$ должен соответствовать знаку при коэффициенте b_i в уравнении оптимизируемого процесса.

Таблица 1.9

Процедура Бокса-Уилсона при оптимизации эксперимента

Параметры	Факторы	
	I	II
b_i	b_1	b_2
δ_i	δ_1	δ_2
$b_i \delta_i$	$b_1 \delta_1$	$b_2 \delta_2$
k_i	1	$k_2 = b_2 \delta_2 / b_1 \delta_1 $
Δz_i	Δz_1	$\Delta z_2 = k_2 \Delta z_1$
Δz_{i0}	Δz_{10}	Δz_{20}
Δz_{i1}	$\Delta z_{10} + \Delta z_1$	$\Delta z_{20} + \Delta z_2$
Δz_{i2}	$\Delta z_{10} + 2\Delta z_1$	$\Delta z_{20} + 2\Delta z_2$
Δz_{i3}	$\Delta z_{10} + 3\Delta z_1$	$\Delta z_{20} + 3\Delta z_2$
...

Крутое восхождение считается эффективным, если хотя бы один из реализованных опытов даст лучший результат по сравнению с наилучшим результатом исходной серии опытов.

§ 3.4. Ортогональное планирование 2-го порядка при поиске оптимальных условий

Вблизи экстремума, т. е. в почти стационарной области, поверхность отклика имеет значительную кривизну. Эта область с существенной нелинейностью функции отклика, для адекватного описания которой необходимо использовать модель в виде полного квадратного уравнения (полинома второй степени):

$$y = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i,j=1}^k b_{ij} x_i x_j + \sum_{i=1}^k b_{ii} x_i^2. \quad (1.85)$$

Планы второго порядка являются многоуровневыми, т. к. для описания части поверхности отклика вблизи стационарной (околооптимальной) области полиномами второго порядка надо факторы (переменные) варьировать уже на трех уровнях. Наиболее часто на практике используют последовательно строящиеся или композиционные планы. Их получают путем добавления $2k$ звездных точек типа $(\pm\alpha, 0, 0, \dots)$, \dots , $(\dots, 0, 0, \pm\alpha)$ и некоторого числа нулевых точек к ядру, образованному полным факторным экспериментом (ПФЭ) типа 2^k .

Чаще всего I-ой частью является план ПФЭ 2^k (реже ДФЭ 2^k), опыты которого уже осуществлены, по их результатам получено уравнение или хотя бы оценка b_0 , что позволяет провести анализ работоспособности уравнения по опыту в центре эксперимента (негативный результат этого анализа и некоторые другие соображения показывают необходимость перехода к плану второго порядка).

Второй частью является совокупность так называемых «звездных» точек с плечом $\pm\alpha$ по одному из факторов и нулевым уровнем для остальных факторов.

Третьей частью является опыт (или несколько опытов) в центре эксперимента.

Такие планы называются композиционными, т. к. они строятся последовательно; а также центральными, поскольку все опыты симметрично располагаются вокруг центра эксперимента.

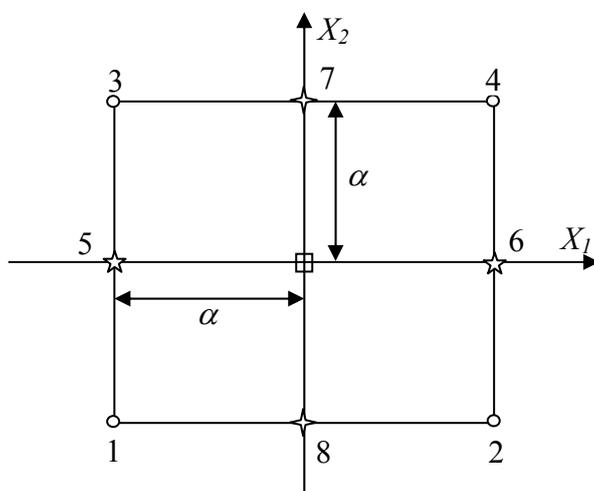


Рис. 1.9. Композиционный план второго порядка $k = 2$:
 о полный факторный эксперимент 2^k ;

☆ опыты в звездных точках; □ опыты в центре плана

Таким образом, общее число точек композиционного плана при k факторах определяется исходя из количества уровней плана:

$$N = 2^k + 2k + n_0, \quad (1.86)$$

- факторный план типа 2^k составляет «ядро» центрального композиционного плана,
- звездные (нулевые) точки строятся на осях факторного пространства и достраиваются к «ядру» плана,
- опыты в центре плана (n_0) дополняют план.

Различают два типа композиционных планов второго порядка – ортогональный и равномер-ротатабельный. Первый из них позволяет путем некоторых преобразований получить по результатам его реализации квадратное уравнение с независимыми, ортогональными друг другу оценками всех коэффициентов (табл. 1.10), второй – уравнение, в котором некоторые коэффициенты связаны (коррелированы) с другими коэффициентами, но оценка дисперсии предсказания практически одинакова для всех опытов плана.

Таблица 1.10

*Центральное композиционное ортогональное планирование
второго порядка для $k = 2$*

Система опытов	N	X_0	X_1	X_2	X_1X_2	$X_1^2 - \bar{X}_1^2$	$X_2^2 - \bar{X}_2^2$	y
Полный факторный эксперимент	1	+1	-1	-1	+1	+0,33	+0,33	y_1
	2	+1	+1	-1	-1	+0,33	+0,33	y_2
	3	+1	-1	+1	-1	+0,33	+0,33	y_3
	4	+1	+1	+1	+1	+0,33	+0,33	y_4
Опыты в звездных точках	5	+1	-1	0	0	+0,33	-0,67	y_5
	6	+1	+1	0	0	+0,33	-0,67	y_6
	7	+1	0	+1	0	-0,67	+0,33	y_7
	8	+1	0	-1	0	-0,67	+0,33	y_8
Опыты в центре плана	9	+1	0	0	0	-0,67	-0,67	y_9

Выбор расстояния от нулевой точки до звездной, определяемой плечом α , зависит от критерия оптимальности плана. Однако, в общем случае в матрице для центрального композиционного планирования второго порядка не все вектор-столбцы ортогональны.

Скалярные произведения

$$\left. \begin{aligned} \sum_{u=1}^N x_{0u} x_{iu}^2 &\neq 0 \\ \sum_{u=1}^N x_{iu}^2 x_{ju}^2 &\neq 0 \end{aligned} \right\} \quad (1.87)$$

т. к. и x_{0u} и x_{iu}^2 всегда положительны (x_{0u} – фиктивная переменная, которая показывает выход в центре плана).

Ортогонализация оценок всех входящих в уравнение (1.85) оценок коэффициентов достигается переносом точки отсчёта x_{iu}^2 на величину среднеарифметического значения суммы квадратов безразмерного выражения i -го фактора, т. е. на

$$\bar{x}_i^{-2} = \frac{1}{N} \sum_{u=1}^N x_{iu}^2. \quad (1.88)$$

Тогда преобразованный расчётный столбец квадратичного эффекта (табл. 1.10) будет иметь элементы не x_{iu}^2 , а

$$x'_{iu} = x_{iu}^2 - \bar{x}_i^{-2} \quad (1.89)$$

x_{iu}^2 может быть больше, меньше или равным \bar{x}_i^{-2} , таким образом для нового расчётного столбца сохранится свойство симметричности:

$$\sum_{u=1}^N x'_{iu} = x_{iu}^2 - \frac{\sum_{u=1}^N x_{iu}^2}{N} = x_{iu}^2 - \bar{x}_i^{-2} = 0, \quad (1.90)$$

а оценки квадратичных эффектов будут ортогональны свободному члену b_0 , т. к.

$$\sum_{u=1}^N x'_{iu} x_{0u} = \sum_{u=1}^N x_{iu}^2 - N \bar{x}_i^{-2} = 0. \quad (1.90)$$

При ортогонализации 2-го из соотношений (1.87) должно подтвердиться равенство:

$$\sum_{u=1}^N x'_{iu} x'_{ju} = 0 \quad (1.90)$$

или

$$\sum_{u=1}^N (x_{iu}^2 - \bar{x}_i^{-2})(x_{ju}^2 - \bar{x}_j^{-2}) = 0, \quad \bar{x}_i^{-2} = \bar{x}_j^{-2}, \quad (1.91)$$

тогда расчёт оценок квадратичных эффектов b_{ii} , b_{jj} будет проводиться независимо друг от друга.

Величина \bar{x}_i^{-2} зависит от числа факторов k и плеча α :

$$\bar{x}_i^{-2} = \frac{2^k + 2\alpha^2}{2^k + 2k + 1} = \frac{2^k + 2\alpha^2}{N}. \quad (1.92)$$

Таблица 1.10

Параметры ортогональных планов второго порядка

k	Ядро плана	α	\bar{x}_{iu}^2	N
2	ПФЭ 2^2	1,000	0,667	9
3	ПФЭ 2^3	1,215	0,730	15
4	ПФЭ 2^4	1,414	0,800	25
5	ДФЭ 2^{5-1}	1,547	0,770	27
5	ПФЭ 2^5	1,596	0,863	43
6	ДФЭ 2^{6-1}	1,722	0,843	45
6	ПФЭ 2^6	1,761	0,912	77
7	ДФЭ 2^{7-1}	1,885	0,900	79
7	ПФЭ 2^7	1,910	0,946	143

Поскольку на число нулевых точек не накладывается какого-либо ограничения, принимаем их минимум (1).

$$\text{Для 3-х факторного плана: } \bar{x}_i^{-2} = \frac{1}{N} \sum_{u=1}^N x_{iu}^2 = \frac{8 + 2\alpha^2}{15}.$$

Здесь ортогонализация достигается выбором величины звёздного плеча, для определения α необходимо решить уравнение (1.91):

$$\sum_{u=1}^N \left(x_{1u}^2 - \frac{8 + 2\alpha^2}{15} \right) \left(x_{2u}^2 - \frac{8 + 2\alpha^2}{15} \right) = 0.$$

После несложных преобразований для 3-х факторного плана получаем:

$$25R^4 + 84R^2 - 178 = 0,$$

$$R^2 = 1,476 \text{ и } R = \pm 1,215.$$

В качестве примера построим ортогональные центральные композиционные планы второго порядка для трех факторов (рис. 1.10, табл. 1.11).

Благодаря ортогональности матрицы планирования, коэффициенты регрессии вычисляются независимо друг от друга по следующим простым формулам:

$$b_0 = \frac{1}{N} \sum_{u=1}^N x_{0u} y_u, \quad b_i = \frac{\sum_{u=1}^N x_{iu} y_u}{\sum_{u=1}^N (x_{iu})^2}, \quad b_{ij} = \frac{\sum_{u=1}^N x_{iu} x_{ju} y_u}{\sum_{u=1}^N (x_{iu} x_{ju})^2}, \quad b_u = \frac{\sum_{u=1}^N x'_{iu} y_u}{\sum_{u=1}^N (x'_{iu})^2} \quad (1.93)$$

где $i \neq 0$, $i \neq j$, i – номер столбца в матрице планирования; x_{ij} – элементы i -го столбца.

Таблица 1.11

Центральное композиционное ортогональное планирование
второго порядка для $k=3$

Система опытов	u	X_0	X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	$X_1^2 - 0,73$	$X_2^2 - 0,73$	$X_3^2 - 0,73$	y_u
ПФЭ 2^3	1	+1	-1	-1	-1	+1	+1	+1	0,27	0,27	0,27	y_1
	2	+1	+1	-1	-1	-1	-1	+1	0,27	0,27	0,27	y_2
	3	+1	-1	+1	-1	-1	+1	-1	0,27	0,27	0,27	y_3
	4	+1	+1	+1	-1	+1	-1	-1	0,27	0,27	0,27	y_4
	5	+1	-1	-1	+1	+1	-1	-1	0,27	0,27	0,27	y_5
	6	+1	+1	-1	+1	-1	+1	-1	0,27	0,27	0,27	y_6
	7	+1	-1	+1	+1	-1	-1	+1	0,27	0,27	0,27	y_7
	8	+1	+1	+1	+1	+1	+1	+1	0,27	0,27	0,27	y_8
Опыты в звездных точках	9	+1	1,2	0	0	0	0	0	0,746	-0,73	-0,73	y_9
	10	+1	-1,2	0	0	0	0	0	0,746	-0,73	-0,73	y_{10}
	11	+1	0	1,2	0	0	0	0	-0,73	0,746	-0,73	y_{11}
	12	+1	0	-1,2	0	0	0	0	-0,73	0,746	-0,73	y_{12}
	13	+1	0	0	1,2	0	0	0	-0,73	-0,73	0,746	y_{13}
	14	+1	0	0	-1,2	0	0	0	-0,73	-0,73	0,746	y_{14}
Опыты в центре плана	15	+1	0	0	0	0	0	0	-0,73	-0,73	-0,73	y_{15}

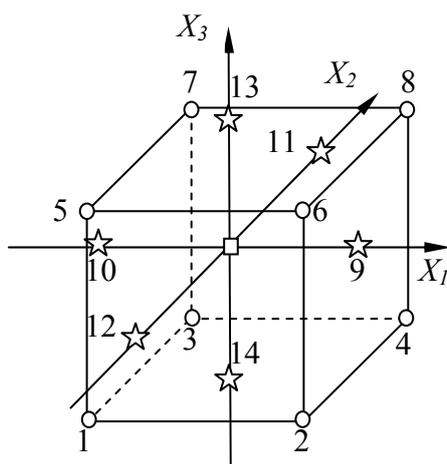


Рис. 1.10. Схема построения ортогонального ЦКП
второго порядка для трех факторов:
○ полный факторный эксперимент 2^k ;
☆ опыты в звездных точках; □ опыты в центре плана

Таблица. 1.12

Значения $\sum_{u=1}^N x_{iu}^2$ для ортогональных планов второго порядка.

Столбец в матрице планирования

Переменная	Значение $\sum_{u=1}^N x_{iu}^2$ при ядре типа								
	ПФЭ						ДФЭ		
	2^2	2^3	2^4	2^5	2^6	2^7	2^{5-1}	2^{6-1}	2^{7-1}
x_0	9	15	25	43	77	143	27	45	79
x_i	6	10.95	20	37.09	70.20	135.29	20.78	37.95	71.10
x_{ij}	4	8	16	32	64	128	16	32	64
$x_i^2 - \bar{x}_i^2$	2	4.36	8	16.77	19.22	26.59	8.24	17	25.24

Знаменатели в формулах (1.93) не зависят от результатов опытов, поэтому их можно вычислить заранее, что упрощает обработку результатов экспериментов.

§ 4. Моделирование и оптимизация процессов изотопного разделения

Рассмотрим процесс разделения изотопов осуществляемый в обменной противоточной колонне. В колонне создаётся противоток двух фаз: катионита (твёрдофазного ионообменника, способного поглощать катионы раствора в обмен на катионы своих ионогенных групп [1]) и раствора, содержащего обменные катионы.

Для определения коэффициентов уравнения регрессии составляется матрица планирования. Рассмотрим основные уровни и интервалы варьирования, а также составим матрицу планирования для двух факторов (x_1 – отбор, x_2 – поток). Пример расчётов приведён в табл. 1.13 ÷ 1.15.

Таблица 1.13

Интервалы и уровни варьирования переменных

Величина	P , мкг/с	J , мкг/с
Основной уровень	0,15	406
Интервал варьирования	0,05	30
Верхний уровень (+1)	0,2	436
Нижний уровень (-1)	0,1	376

Математическая модель процесса разделения примет вид уравнения регрессии первого порядка:

$$y = 1,2711013 - 0,1431955x_1 + 0,0308327x_2 \quad (1.94)$$

Таблица 1.14

Матрица планирования двухфакторного эксперимента

Опыты	Значения факторов				
	Переменные	x_0	x_1	x_2	q
1		+	+	-	1.08958
2		+	-	+	1.43764
3		+	+	+	1.16623
4		+	-	-	1.39096
Коэффициенты уравнения регрессии		1.271101	-0.143196	0.030833	

Таблица 1.15

Значение функции отклика в зависимости от уровней факторов и определяемых величин

u	P , мг/с	J , мг/с	Δc	c_2	q	\hat{y}
1	0,0002	0,37600	0,00055	0,04705	1,08958	1,28267
2	0,0001	0,43600	0,00235	0,04885	1,43764	1,28453
3	0,0002	0,43600	0,00098	0,04748	1,16623	1,28452
4	0,0001	0,37600	0,00214	0,04864	1,39096	1,28268

Таблица 1.16

Оценка адекватности

Дисперсия опытов	Остаточная дисперсия	Значение критерия Фишера	
		F_{on}	F_T
0,0286823	0,0864387	3,01	10,1

Теоретическое значение критерия Фишера меньше опытного (табл. 1.16): $F_T > F_{on}$ – линейная модель адекватна.

Движение к области оптимума будем проводить методом крутого восхождения или по градиенту, когда переход от точки к точке сопровождается одновременным изменением значений факторов. Для этого необходимо рассчитать направление градиента, определить величину шага, найти координату оптимума и исследовать форму поверхности отклика.

Оптимальной в данном случае является степень разделения:

$$q = \alpha^n = 1,087085211,$$

при этом число ступеней разделительной установки – $n = 7$, коэффициент обогащения $\varepsilon = \alpha - 1 = 0,012$. На основе полученного уравнения (1.94) произведём движение по градиенту для нахождения оптимума.

Координаты новой точки находятся пропорционально приращению по каждому фактору. Движение считается эффективным, если последующий опыт показывает лучший результат по сравнению с предыдущим. Пример расчётов приводится в табл. 1.17 ÷ 1.19, здесь изменяя факторы пропорционально величинам и знакам коэффициентов регрессии, осуществляется движение в направлении градиента оптимизируемой величины. Затем, выбирается опыт с наилучшим значением параметра оптимизации, и его условия принимаются за основной уровень факторов.

Таблица 1.17

Определение диапазона изменения величин

№	Переменные	x_0	x_1	x_2	y
1.	Основной уровень	1	0,15	406	1,266
2.	Интервал варьирования Δ_i		0,05	30	
3.	Коэффициент регрессии	1,271	0,143	0,038	
4.	Произведение $b_i \cdot \Delta_i$		-0,007	1,140	
5.	$k_i = \frac{ b_i \cdot \Delta_i }{ b_i \cdot \Delta_i _{\max}}$		0,006	1	
6.	Δ_i'		0,006	1	
7.	Округление Δ_i'		0,01	1	

Таблица 1.18

Процедура Бокса-Уилсона при оптимизации эксперимента

№	Опыты	x_1	x_2	y
1.	(реализованный)	0,16	405	1,237
2.	(реализованный)	0,17	404	1,197
3.	(мысленный)	0,18	403	
4.	(реализованный)	0,19	402	1,151
5.	(мысленный)	0,20	401	
6.	(реализованный)	0,21	400	1,097
7.	(реализованный)	0,22	399	1,070
	Δ_i''	0,005	0,5	
8.	(реализованный)	0,215	399,5	1,083
9.	(реализованный)	0,220	399	1,070
	Δ_i'''	0,0036	0,359	
10.	(реализованный)	0,214	399,6	1,087

Таблица 1.19

Центральное композиционное планирование второго порядка.

$$(\alpha=1, \bar{x}_{iu}^2=0,667)$$

Система опытов	u	x_0	x_1	x_2	x_1x_2	$x_1^2 - 0,667$	$x_2^2 - 0,667$	y
ПФЭ	1	+1	+1	-1	-1	0,33	0,33	1,061
	2	+1	-1	+1	-1	0,33	0,33	1,114
	3	+1	+1	+1	+1	0,33	0,33	1,063
	4	+1	-1	-1	+1	0,33	0,33	1,111
в звездных точках	5	+1	+1	0	0	0,33	-0,67	1,062
	6	+1	-1	0	0	0,33	-0,67	1,113
	7	+1	0	+1	0	-0,67	0,33	1,088
	8	+1	0	-1	0	-0,67	0,33	1,086
в центре плана	9	+1	0	0	0	-0,67	-0,67	1,087

Математическая модель, учитывающая взаимосвязь факторов, имеет вид полинома второго порядка:

$$y = 1,087252 - 0,02538x_1 + 0,001356x_2 + 0,000036 x_1 x_2 - 0,01605x_1^2 - 0,01631x_2^2.$$

Проверка на адекватность данной модели показала, что $F_{on} < F_{т}$, следовательно, полученная математическая модель адекватно описывает процесс разделения в двухфазной системе.

Контрольные вопросы и задания

1. Назовите известные вам методы оптимизации эксперимента, объясните их суть.
2. Запишите формулы, определяющие величину дисперсии адекватности и дисперсии воспроизводимости опытов.
3. Какие критерии адекватности Вам известны, чем они отличаются?
4. Составьте линейную математическую модель обменного разделительного процесса методом планирования экспериментов, определите величину степени разделения ($J = 0,224$ мг/с), проведите анализ влияния каждого фактора на функцию отклика.
5. Какими свойствами должны обладать планы, чтобы оценки коэффициентов полученного по ним линейного уравнения не зависели друг от друга?
6. Постройте план ПФЭ типа 2^3 в безразмерном выражении и в натуральной размерности факторов по следующим параметрам плана: $z_{10} = 0,07$ см/с, $z_{20} = 0,025$ см/с, $z_{30} = 0,025$ см, $\delta_1 = 0,01$, $\delta_2 = 0,015$,

$\delta_3=0,005$, $y_i = (3,35; 4,47; 2,60; 3,69; 6,36; 8,47; 5,61; 7,69)$ [см] – в соответствии с планом эксперимента табл. 1.2.

7. Рассчитайте коэффициенты уравнения регрессии по плану эксперимента, построенному на условиях п.6, в безразмерном выражении и натуральной размерности факторов.
8. Каким образом проводится статистическая оценка значимости коэффициентов уравнения регрессии (покажите на примере математических моделей, полученных по условиям п.4 и п.6)? Результаты ранжировки факторов представьте графически.
9. Как можно определить насколько полученная математическая модель соответствует реально протекающим процессам? Покажите на примере математических моделей, полученных в п.4 и п.6.
10. Когда переход от планов первого порядка к планам второго порядка является целесообразным?
11. Структура композиционных планов второго порядка.
12. Чем отличаются D- и G-оптимальные планы?
13. Что собой представляют ротатабельные и униформ-ротатабельные планы эксперимента?
14. Каким образом достигается ортогонализация центральных композиционных планов?
15. Определите оптимальные условия протекания процесса изотопного обмена (по условиям п. 4) и составьте математическую модель для стационарной области. Постройте поверхность отклика $q = f(P, J)$.
16. Каким образом можно привести к канонической форме уравнение регрессии, если оно описывает: а) центральную поверхность, б) нецентральную поверхность?

ГЛАВА 2

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ С ПРИМЕНЕНИЕМ МЕТОДОВ ТЕОРИИ ПОДОБИЯ

При решении технических задач широко применяют технику моделирования как математического (математические модели процессов), так и физического (физическая модель процесса). Если явления, подлежащие исследованию, протекают на производственном объекте, где непосредственный эксперимент затруднителен (например, в процессах разделения ионных или изотопных смесей), то рационально проводить его или на уменьшенном объекте, копирующем оригинал, или на установке, воспроизводящей физически иное явление, но описываемое такими же уравнениями, что и явление в оригинале.

В первом случае используется *физическое моделирование*, при котором сравниваемые явления имеют одинаковую природу и отличаются только количественно. Такое моделирование позволяет углубить знания о комплексе происходящих явлений и уточнить математическое описание отдельных процессов. Физическая модель может быть широко использована для различных демонстраций в учебных целях (учебные стенды, установки) и научных целях.

Во втором случае осуществляется *математическое моделирование* или моделирование аналогиями. Этот вид моделирования имеет своей целью только решение уравнений, что может быть выполнено самыми различными путями.

При этом возникает вопрос, будут ли воспроизводиться полученные результаты на промышленных установках. Если рассматриваемые системы подобны, то результаты, полученные на лабораторных установках, математических моделях, можно воспроизвести и на укрупнённых производственных объектах.

§ 1. Теория подобия

Теория подобия – учение о подобных процессах (явлениях) в природе и технике.

Теория подобия [17, 27, 29] основывается на том очевидном факте, что явления в природе и технике не могут зависеть от выбора единиц измерения и их размерности. Поэтому первым этапом использования теории подобия является описание изучаемых явлений в безразмерной форме путем перехода от размерных переменных (T , w , λ , c , v и т. д.) к безразмерным величинам.

π -теорема размерности устанавливает число безразмерных переменных (формула 3.1.) при общем числе размерных переменных n и числе первичных (основных) размерных переменных k .

$$\pi = n - k. \quad (2.1)$$

Первичные размерные переменные в системе СИ – это масса [кг], время [с], длина [м], температура [К], сила тока [А] и т. п.

Первичные размерные величины определяют путем прямого измерения. Другие размерные переменные, выражаемые через основные, называют *вторичными*, и их размерность выражается через размерность первичных величин: скорость [м/с], сила [Н = кг · м/с²], работа (теплота) [Дж = Н · м] и др.

Безразмерные переменные представляют собой определенную комбинацию размерных переменных, например, в виде отношения двух одномерных величин $X = x/L$, $T = T_1/T_0$ и др.

Вместе с тем существуют более сложные безразмерные комплексы, включающие три и более размерные величины. Эти безразмерные комплексы не выбирают произвольно, а получают непосредственно из уравнений, описывающих рассматриваемые процессы.

§ 2. Константы подобия

Подобные процессы должны происходить в подобных геометрических и временных условиях, иметь одинаковую физическую природу и описываться одинаковыми по форме и по существу сравнениями.

Необходимым и достаточным условием подобия двух процессов разделения является то, что процессы удовлетворяют требованиям условия однозначности.

Под подобием условий однозначности следует понимать:

1. *Геометрическое подобие* (геометрические размеры аппарата по разделению изотопов и веществ с близкими свойствами).

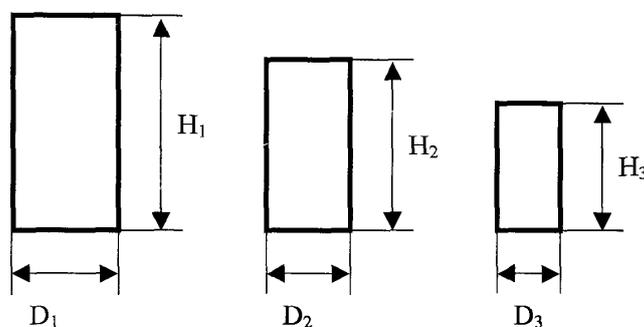


Рис. 2.1. Масштабируемость подобных объектов

У подобных тел отношение сходных размеров постоянно, а сами измеряемые размеры параллельны в пространстве (рис. 2.1):

$$\frac{D_1}{D_2} = \frac{H_1}{H_2} = \dots = K_L, \quad (2.2)$$

K_L – константа подобия

Если выбирается относительная величина для геометрического подобия можно записать, следующее выражение:

$$\frac{H_1}{D_1} = \frac{H_2}{D_2} = \frac{H_3}{D_3} = iH, \quad (2.3)$$

iH – инварианта подобия.

2. *Временное подобие* (постоянные времени в системе).

Отношение между сходными стадиями двух подобных процессов есть величина постоянная:

$$\frac{\tau_1}{\tau_2} = \frac{\tau'_1}{\tau'_2} = \frac{\tau''_1}{\tau''_2} = K_\tau, \quad (2.4)$$

K_τ – масштабный множитель временного подобия,

' , '' – обозначают сходные стадии подобных процессов.

Процессы, для которых выполняется временное подобие – **гомохронные**. Если. $K_\tau = 1$ – процессы **синхронные**.

3. *Подобие физических величин* (физические константы среды в системе).

При соблюдении подобия физических величин отношение этих величин в сходных точках, в сходные моменты времени есть величины постоянные:

$$\frac{U_1}{U'_1} = \frac{U_2}{U'_2} = \dots = \frac{U_n}{U'_n} = K_U, \quad (2.5)$$

где U_1, U_2, \dots, U_n – физические величины; U'_1, U'_2, \dots, U'_n – подобные величины; K_U – константа подобия (масштабный множитель).

4. *Подобие начальных и граничных условий* (это такие параметры процесса как начальная скорость, концентрация, давление и т. п., а также переменные, характеризующие процесс, например, на границах фаз)

Если подобные физические величины являются векторными, то их направления совпадают. Например, в подобных системах в сходные моменты времени при подобном движении ионов их скорости имеют одно направление в сходных точках: $w_1 \uparrow \uparrow w_2$

Рассмотрим подобное преобразование II-го закона Ньютона:

$$f_1 = m_1 \frac{dw_1}{d\tau_1}, \quad f_2 = m_2 \frac{dw_2}{d\tau_2}$$

$$\underbrace{\frac{f_1 \cdot d\tau_1}{m_1 dw_1}} = 1 \quad (2.6)$$

$$\underbrace{\frac{f_2 \cdot d\tau_2}{m_2 \cdot dw_2}} = 1 \quad (2.7)$$

1-й процесс 2-й процесс

Будем считать, что эти два процесса подобные, введем масштабные множители:

$$\frac{f_1}{f_2} = K_f; \quad \frac{m_1}{m_2} = K_m; \quad \frac{w_1}{w_2} = K_w; \quad \frac{\tau_1}{\tau_2} = K_\tau.$$

$$f_2 = \frac{f_1}{K_f}; \quad m_2 = \frac{m_1}{K_m}; \quad w_2 = \frac{w_1}{K_w}; \quad \tau_2 = \frac{\tau_1}{K_\tau}.$$
(2.8)

Подставим в (2.6–2.7):

$$\frac{f_1}{K_f} = \frac{m_1}{K_m} \cdot \frac{dw_1 \cdot K_\tau}{K_w \cdot d\tau_1}$$

$$\frac{f_1 \cdot d\tau_1}{m_1 \cdot dw_1} \cdot \frac{K_w \cdot K_m}{K_f \cdot K_\tau} = 1$$

$$\frac{f_1 \cdot d\tau_1}{m_1 \cdot dw_1} = 1, \quad \frac{K_w \cdot K_m}{K_f \cdot K_\tau} = 1, \quad \frac{K_f \cdot K_\tau}{K_w \cdot K_m} = 1$$

$$\frac{K_f \cdot K_\tau}{K_w \cdot K_m} = const - \text{индикатор подобия}$$
(2.9)

Индикатор подобия составляется из масштабных множителей.

§ 3. Критерии подобия

В теории и практике моделирования при реализации подобия имеют большое значение *критерии подобия*, т. е. условия, указывающие на возможность осуществления или наличие подобия. Понятие «критерий подобия» является фундаментальным понятием, существенным для всех видов моделирования во всех научных дисциплинах.

Комплексы размерных величин, сочетание констант подобия которых равно единице, называют критериями подобия. Критерии подобия сохраняют для всех подобных явлений одинаковые (но не обязательно постоянные) значения. Существенным признаком критериев подобия является нулевая размерность.

Инварианты подобия, составленные из разнородных величин, называются критериями подобия.

Критерии подобия, получаемые как комбинации физических величин, характеризующих изучаемые процессы, могут быть получены и как комбинации (произведения, отношения) ранее определенных критериев подобия. Критериями подобия можно характеризовать любое конкретное явление. Установить функциональную зависимость между критериями подобия обычно легче, чем между физическими величинами.

Условие подобия: у подобных процессов индикаторы подобия = 1. Заменим в индикаторе подобия (3.9) масштабные множители:

$$\frac{f_1 \tau_1}{w_1 m_1} = 1, \quad \underbrace{\frac{f_1 \tau_1}{m_1 w_1} = \frac{f_2 \tau_2}{m_2 w_2} = \frac{f \tau}{m w}}_{idem} = idem . \quad (2.10)$$

Критерий Ньютона

Где *idem* означает одинаковое (одно и то же) значение для всех подобных процессов.

Установив соотношения между критериями подобия (критериальные уравнения), легко перейти и к соотношениям непосредственно между самими физическими величинами. Критерии подобия являются инвариантами, характеризуя не только какое-либо одно конкретное явление, но целую группу подобных явлений.

Первая теорема подобия: *подобные между собой процессы имеют равные критерии подобия.*

Критерии подобия, составленные из условия однозначности, называются *определяющими критериями*. Критерии, которые в своём составе имеют величины, не относящиеся к условиям однозначности, назы-

ваются *определяемыми критериями*. В подобных процессах все критерии равны (определяющие равны определяемым).

Вторая теорема подобия: *подобны те процессы, условия однозначности которых подобны, а определяющие критерии равны.*

Третья теорема подобия: *любая зависимость между переменными, характеризующими процесс, может быть представлена в виде зависимости между критериями или в виде обобщающего критериального уравнения.*

Таким образом, результаты опытов, проведённых в лабораторных условиях можно переносить на подобные процессы, у которых определяющие критерии равны, а условия однозначности подобны.

Практические применения теории подобия весьма обширны. Она даёт возможность предварительного качественно-теоретического анализа и выбора системы определяющих безразмерных параметров сложных физических явлений. Теория подобия является основой для правильной постановки и обработки результатов экспериментов.

§ 3.1. Критерии подобия в тепломассообменных процессах разделения веществ

Движение жидкости. Если движение идеальной (невязкой) жидкости описывают уравнения Эйлера или Бернулли, то для вязкой жидкости используют уравнение Навье – Стокса. Рассмотрим уравнение движения вязкой изотермически несжимаемой жидкости (уравнения Навье – Стокса):

$$\rho \left(\frac{\partial w_z}{\partial \tau} + w_x \frac{\partial w_z}{\partial x} + w_y \frac{\partial w_z}{\partial y} + w_z \frac{\partial w_z}{\partial z} \right) = -\rho g - \frac{\partial p}{\partial z} + \mu \left(\frac{\partial^2 w_z}{\partial x^2} + \frac{\partial^2 w_z}{\partial y^2} + \frac{\partial^2 w_z}{\partial z^2} \right), \quad (2.11)$$

где ρ – плотность среды, w – скорость потока, g – ускорение в поле силы тяжести, p – давление, μ – динамическая вязкость, τ – время.

Коэффициент динамической вязкости (внутреннего трения)

$$\mu = \rho \cdot \nu, \quad (2.12)$$

представляющий собой импульс силы, отнесенный к единице поверхности, и характеризующий вязкое трение в жидкости имеет размерность $[\text{Н} \cdot \text{с}/\text{м}^2]$. Здесь ν – кинематическая вязкость, $[\text{м}^2/\text{с}]$.

Для газов коэффициент динамической вязкости связан с их теплопроводностью (λ), $\mu \approx \lambda/c_v$, где c_v – теплоёмкость при постоянном объёме.

ме. Для жидкостей и газов $\mu = \rho \cdot \nu$ практически не зависит от давления, так как для газов ν обратно, а ρ прямо пропорционально давлению, а для жидкостей ρ и ν практически не зависят от давления. Вязкость жидкостей и газов зависит от температуры. Для капельных жидкостей вязкость убывает с повышением температуры, а для газов возрастает.

Левая часть уравнения (2.12) представляет силы инерции элемента объема жидкости, записанные в общем виде. Правая часть есть сумма сил, действующих на элемент объема жидкости со стороны окружающих ее частей: сила ($P = -grad p$), сила вязкого трения ($\rho \nu \nabla^2 \bar{w}$) и сила веса (ρg).

Данное уравнение описывает целый класс процессов, обусловленных неустановившимся движением вязкой несжимаемой жидкости. В теории разделения данное уравнение применяется для описания процесса массообмена.

Рассмотрим два подобных процесса, описываемых данным дифференциальным уравнением и введём масштабные множители:

$$\begin{aligned} \frac{x_1}{x_2} = \frac{y_1}{y_2} = \frac{z_1}{z_2} = k_L \quad \frac{\mu_1}{\mu_2} = k_\mu \quad \frac{\tau_1}{\tau_2} = k_\tau \quad \frac{w_1}{w_2} = k_w \\ \frac{\rho_1}{\rho_2} = k_\rho \quad \frac{p_1}{p_2} = k_p \quad \frac{g_1}{g_2} = k_g \end{aligned} \quad (2.13)$$

Преобразуем $x_1 = x_2 k_L \quad y_1 = y_2 k_L \quad z_1 = z_2 k_L \quad \mu_1 = \mu_2 k_\mu$
 $w_1 = w_2 k_w \quad \tau_1 = \tau_2 k_\tau \quad \rho_1 = \rho_2 k_\rho \quad g_1 = g_2 k_g \quad p_1 = p_2 k_p$

Будем считать, что исходное дифференциальное уравнение характеризует второй подобный процесс из двух рассматриваемых, тогда $\rho = \rho_2$

Для первого подобного процесса можно с учётом приведённых соотношений записать:

$$\begin{aligned} k_\rho \rho \left(\frac{k_w}{k_\tau} \frac{\partial w_z}{\partial \tau} + \frac{k_w^2}{k_L} w_x \frac{\partial w_z}{\partial x} + \frac{k_w^2}{k_L} w_y \frac{\partial w_z}{\partial y} + \frac{k_w^2}{k_L} W_z \frac{\partial w_z}{\partial z} \right) = \\ = k_\rho k_g (-\rho g) - \frac{k_\rho}{k_L} \frac{\partial p}{\partial z} + \frac{k_\mu k_w}{k_L^2} \mu \left(\frac{\partial^2 w_z}{\partial x^2} + \frac{\partial^2 w_z}{\partial y^2} + \frac{\partial^2 w_z}{\partial z^2} \right). \end{aligned} \quad (2.14)$$

Разделим левую и правую часть на $\frac{k_\rho k_w^2}{k_L}$

$$\begin{aligned} \rho \left(\frac{k_L}{k_\tau k_w} \frac{\partial w_z}{\partial \tau} + w_x \frac{\partial w_z}{\partial x} + w_y \frac{\partial w_z}{\partial y} + W_z \frac{\partial w_z}{\partial z} \right) = \\ = \frac{k_L k_g}{k_w^2} (-\rho g) - \frac{k_\rho}{k_\rho k_w^2} \frac{\partial p}{\partial z} + \frac{k_\mu}{k_\rho k_L k_w} \mu \left(\frac{\partial^2 w_z}{\partial x^2} + \frac{\partial^2 w_z}{\partial y^2} + \frac{\partial^2 w_z}{\partial z^2} \right). \end{aligned} \quad (2.15)$$

В случае подобных процессов индикаторы подобия должны быть равны (табл. 2.1.)

Таблица 2.1

Индикаторы подобия физических величин

Индикатор подобия	Физический смысл
$\frac{k_L}{k_\tau k_w} = 1$	Учитывает нестационарность процесса (присутствует временной фактор k_τ).
$\frac{k_L k_g}{k_w^2} = 1$	Учитывает влияние силы тяжести – k_g .
$\frac{k_p}{k_\rho k_w^2} = 1$	Учитывает давление в потоке.
$\frac{k_\mu}{k_\rho k_L k_w} = 1$	Учитывает влияние сил внутреннего трения – $\rho\mu w$.

На основании полученных индикаторов подобия можно получить критерии подобия. Распишем первый индикатор подобия. Инварианта подобия определится как:

$$\frac{\ell_1/\ell_2}{(\tau_1/\tau_2)(w_1/w_2)} = 1. \quad (2.16)$$

Преобразуем (3.16):

$$\frac{w_1 \tau_1}{\ell_1} = \frac{w_2 \tau_2}{\ell_2} = idem. \quad (2.17)$$

Для второго индикатора:

$$\frac{k_L k_g}{k_w^2} = \frac{g_1 \ell_1 / w_1^2}{g_2 \ell_2 / w_2^2} = 1 \Rightarrow \frac{w_1^2 g_1}{\ell_1} = \frac{w_2^2 g_2}{\ell_2} = idem. \quad (2.18)$$

Третий индикатор:

$$\frac{k_p}{k_\rho k_w^2} = \frac{p_1 / \rho_1 w_1^2}{p_2 / \rho_2 w_2^2} = 1 \Rightarrow \frac{p_1}{\rho_1 w_1^2} = \frac{p_2}{\rho_2 w_2^2} = idem. \quad (2.19)$$

Иногда при оценке критерия Эйлера учитывается разница давлений.
Четвёртый индикатор:

$$\frac{k_\mu}{k_\rho k_L k_w} = \frac{\mu_1/\mu_2}{(\rho_1/\rho_2)(\ell_1/\ell_2)(w_1/w_2)} = 1 \Rightarrow \frac{w_1 \ell_1 \rho_1}{\mu_1} = \frac{w_2 \ell_2 \rho_2}{\mu_2} = idem. \quad (2.20)$$

С учётом полученных критериев можно составить *обобщающее критериальное уравнение* (математическую модель процесса)

$$\Phi(H_0, Fr, Eu, Re) = 0. \quad (2.21)$$

Если процесс установившийся, тогда преобразуем (2.21)

$$\Phi(H_0, Eu, Re) = 0.$$

Рассмотрим *дифференциальное уравнение диффузии в движущейся среде*:

$$\frac{w \partial c}{\partial x} - D \frac{\partial c}{\partial x} \delta + \varepsilon_0 \left(\frac{\partial c}{\partial \tau} \right) = 0, \quad (2.22)$$

ε_0 – константа, c – концентрация, D – коэффициент диффузии, δ – площадь единичного объёма:

$$\delta = \frac{\partial F}{\partial V} = \left[\frac{1}{cm} \right], \quad k_\delta = \frac{1}{k_L}, \quad (2.23)$$

D – Коэффициент молекулярной диффузии [m^2/c], c – концентрация

Запишем (2.22) с помощью критериев:

$$k_w \frac{k_c}{k_L} - k_D \frac{k_c}{k_L} k_\delta + \varepsilon_0 \frac{k_c}{k_\tau} = 0. \quad (2.24)$$

1 $k_w \frac{k_c}{k_L} = 1$ $1 \text{ и } 3$ $k_w \frac{k_c}{k_L} = \frac{k_c}{k_\tau}$ $\frac{k_w}{k_L} k_\tau = 1$ $H_0 = \frac{w \tau}{\ell}$	2 $k_D \frac{k_c}{k_L} k_\delta = 1$ $2 \text{ и } 3$ $k_D \frac{k_c}{k_L} k_\delta = \frac{k_c}{k_\tau}$ $\frac{D \tau}{\ell^2} = 1$ $F'_0 = \frac{D \tau}{\ell^2}$	3 $\frac{k_c}{k_\tau} = 1$ $1 \text{ и } 2$ $k_w \frac{k_c}{k_L} = k_D \frac{k_c}{k_L} k_\delta$ $\frac{w \ell}{D} = 1$ $Pe' = \frac{w \ell}{D}$
критерий гомохронности	Число Фурье	Число Пекле

Критерий Пекле является произведением двух критериев – Рейнольдса и Шмидта:

$$Pe' = \frac{w\ell}{D} = \frac{w\ell}{\nu} \frac{\nu}{D} = Re Sc. \quad (2.26)$$

Здесь $\nu = \frac{\mu}{\rho}$ – кинематическая вязкость, μ – динамическая вязкость.

Из (2.26) можно выделить число Шмидта [9]:

$$Sc = \frac{\nu}{D} = \frac{\mu}{D\rho}. \quad (2.27)$$

Рассмотрим уравнение массопередачи:

$$\beta \Delta c = -D \frac{\partial c}{\partial x} \delta \quad (2.28)$$

здесь β – коэффициент массопередачи, Δc – градиент концентраций.

$$k_{\beta} k_c = \frac{k_D k_c}{k_L^2} \Rightarrow \frac{k_{\beta} k_L^2}{k_D} = 1 \Rightarrow \frac{\beta_1 \ell_1^2}{\frac{D_1}{D_2}} = 1 \Rightarrow \frac{\beta_1 \ell_1^2}{D_1} = \frac{\beta_2 \ell_2^2}{D_2} = idem. \quad (2.29)$$

Отсюда $Ki = \frac{\beta \ell^2}{D}$ – критерий Кирпичева

Обобщающее критериальное уравнение: $\Phi(Re, Pr', F_0', H_0, Ki) = 0$

Основные критерии, которые применяют для расчёта тепломассообменных процессов, приведены в табл. 2.2.

Таблица 2.2

Основные критерии подобия физических величин

№	Критерий	Наименование	Физический смысл
1	$H_0 = \frac{w\tau}{\ell}$	критерий гомохронности	учитывает нестационарность процесса.
2	$Fr = \frac{w^2}{g\ell}$	число Фруда – критерий гравитационного подобия	характеризует соотношение сил тяжести и сил инерции в потоке.
3	$Eu = \frac{\Delta p}{\rho w^2}$	число Эйлера (критерий подобия полей давления)	характеризует соотношение сил давления и сил инерции в потоке.

№	Критерий	Наименование	Физический смысл
4	$Re = \frac{wl\rho}{\mu}$	число Рейнольдса (критерий режима движения)	характеризует соотношение сил инерции и сил вязкости.
5	$F_0 = \frac{a\tau}{\ell^2}$ $F_0' = \frac{D\tau}{\ell^2}$	число Фурье (критерий тепловой или массовой гомохронности)	характеризует связь между физическими свойствами и размерами тела и скоростью изменения в нём полей температуры (концентрации).
6	$Pe = \frac{wl}{a}$ $Pe' = \frac{wl}{D}$	Критерий теплового (массового) подобия. Тепловое (диффузионное) число Пекле.	Характеризует соотношение конвективного и молекулярного переноса теплоты (вещества) в потоке.
7	$Pr = \frac{\nu}{a_p} = \frac{\mu c_p}{\lambda_f}$, $Sc = \frac{\nu}{D} = Pr'$	тепловое (диффузионное) число Прандтля (Шмидта), λ_f – теплопроводность, c_p – теплоёмкость, a_p – коэффициент температуропроводности среды [м ² /с] при $p = \text{const}$, $a_p = \frac{\lambda_f}{c_p \rho_f}$.	Характеризует подобие скоростных и температурных (массовых) полей. При $\nu = a_p = D$ поля скоростей, температур, концентраций подобны.
8	$Ki = \frac{\beta \ell^2}{D}$	критерий Кирпичева	характеризует размер частицы взвеси через её равновесную скорость и коэффициент сопротивления.
9	$Gr = \frac{\beta \Delta T \ell^3 g}{\nu^2}$	критерий свободного движения среды. Число Грасгофа	характеризует режим движения при свободной конвекции, являясь отношением подъемной силы, возникающей вследствие разности плотностей жидкости, и сил вязкости в неизотермическом потоке.
10	$Ga = \frac{g \ell^3}{\nu^2}$	критерий полей свободного течения (число Галилея)	характеризует соотношение сил вязкости и сил тяжести в потоке.
11	$Ar = \frac{g \ell^3 \Delta \rho}{\nu^2 \rho}$	число Архимеда	характеризует отношение подъемных сил к силам инерции.
12	$Nu = \frac{\alpha \ell}{\lambda_f}$ $Nu' = \frac{\alpha \ell}{D}$	безразмерный коэффициент теплоотдачи (массоотдачи). Тепловое (диффузионное) число Нуссельта, α – коэффициент теплообмена.	Характеризует увеличение теплообмена (массообмена) за счет конвекции по сравнению с чисто молекулярным переносом.

№	Критерий	Наименование	Физический смысл
13	$We = \frac{\rho w \ell}{\sigma}$	число Вебера.	Соотношение сил инерции и поверхностного натяжения в двухфазном потоке.
14	$P_0 = \frac{Q_v \ell^2}{\Delta T \lambda_w}$	критерий обобщенной интенсивности внутренних источников теплоты (число Померанцева), где Q_v – объёмная плотность источников теплоты.	Характеризует отношение мощности источника теплоты в единице объема к количеству теплоты, переносимому в этом объеме через единицу поверхности с характерным размером ℓ .
15	$Le = \frac{D}{a}$	число Льюиса – Семёнова ($Le = \frac{Pr}{Pr'}$)	характеризует подобие безразмерных полей температур и концентраций.
16	$St = \frac{\alpha}{c_p \rho w}$ $St' = \frac{\beta}{w}$	Критерий конвективного переноса теплоты (вещества). Тепловое (диффузионное) число Стентона c_p – теплоёмкость среды.	Характеризует соотношение скорости переноса теплоты (вещества) и линейной скорости потока.
17	$Bi = \frac{\alpha \ell}{\lambda_w}$ $Bi' = \frac{\beta \ell}{D}$	критерий краевого подобию (число Био)	характеризует связь между полем температур в твердом теле и условиями теплоотдачи на его поверхности.
18	$Ku = \frac{r}{c_p \Delta T_s}$	критерий фазового перехода, ΔT_s – перегрев (переохлаждение) относительно температуры фазового перехода, r – теплота фазового перехода	Характеризует соотношение теплоты фазового перехода и теплоты перегрева (переохлаждения) вновь образующейся фазы.

В гидродинамике вязкой жидкости различают два основных режима движения: ламинарное и турбулентное.

При *ламинарном режиме* отдельные слои (струйки) движутся строго в одном направлении, параллельно и не взаимодействуя между собой (силы вязкости между слоями намного меньше сил инерции каждого слоя). При *турбулентном режиме*, возникающем при увеличении скорости движения, взаимодействие слоев (струек) жидкости приводит к их беспорядочному (хаотичному) перемещению и смешению. Каждая частица жидкости одновременно совершает неупорядоченные колебания

относительно некоторого среднего положения. Поэтому турбулентное движение имеет вероятностный (статистический) характер.

Критическое число Рейнольдса характеризует переход от ламинарного режима движения к турбулентному. Число Рейнольдса $Re = \frac{w\ell}{\nu}$ (ℓ – характерный размер) представляет отношение сил инерции потока к силам вязкости и названо в честь О. Рейнольдса, открывшего явление такого изменения режимов движения.

Число Re , при котором осуществляется изменение режима движения, называют критическим, $Re_{кр}$. Величина $Re_{кр}$ зависит от условий движения, скорости, турбулентности потока и др.

Число Re при расчётах ионообменных процессов используется в качестве показателя гидродинамики движения жидкости и частиц ионообменника. Например, рассмотрим движение одиночной частицы сорбента в неподвижном растворе [19] или витание этой частицы. С учётом сил, действующих на частицу, равномерно движущуюся в жидкости, её скорость определяется по известному уравнению:

$$v_0 = \sqrt{4gd_c (\rho_c - \rho_{ж}) 3 \zeta \rho_{ж}} , \quad (2.30)$$

Здесь d_c – диаметр частицы, ρ_c и $\rho_{ж}$ – плотность частицы и жидкости, ζ – коэффициент сопротивления частиц.

В литературе приведено много уравнений для расчета коэффициента сопротивления частиц, однако почти все они применяются в узких пределах изменения скоростей движения частиц. Для *ламинарного режима* обтекания сферических частиц ($Re < 2$) коэффициент сопротивления рассчитывается по закону Стокса:

$$\zeta = \frac{24}{Re} \quad (2.31)$$

В *турбулентном режиме* ($Re > 500$) действует закон сопротивления Ньютона, и коэффициент сопротивления стремится к постоянному значению: $\zeta = 0,44$. В *переходном режиме* ($2 < Re < 500$) наиболее часто используется эмпирическая формула Алена:

$$\zeta = 18,5 Re^{-0.6} \quad (2.32)$$

Для *ламинарной и переходной области* теоретически получена формула Озеена:

$$\zeta = \frac{24 \left(1 + \frac{3}{16} Re \right)}{Re} \quad (2.33)$$

Наиболее точным уравнением для расчета коэффициента сопротивления, учитывающим несферичность частиц и справедливым во всех областях обтекания частиц потоком, является уравнение И. А. Вахрушева:

$$\zeta = \left[\frac{28,47}{\text{Re} \lg \left(\frac{15,38}{f_u} \right)} + \frac{4,565 f_u}{\sqrt[3]{\text{Re}}} - \frac{0,491 f_u}{\sqrt{\text{Re}}} \right] \cdot \left\{ 1 - \text{th} \left[0,0128 \text{Re} (f_u - 0,9805) \right] \right\} + 2,86 \cdot (f_u - 0,8531) \text{th} \left[0,0128 \text{Re} (f_u - 0,9805) \right] + \left(7,76 - 2,86 f_u - \frac{4,88}{f_u} \right) \text{th} \left[0,0104 \text{Re} (f_u - 0,9038) \right]. \quad (2.34)$$

Недостатками этого уравнения являются его громоздкость и необходимость использования табличных функций. Более простая формула предложена в работе:

$$\zeta = 5,31 - \frac{4,88}{f_u} - \left\{ 5,31 - \frac{4,88}{f_u} - \frac{28,47}{\text{Re} \lg \left(\frac{15,38}{f_u} \right)} \sqrt{ \left[1 - 0,142 (2,052 - f_u)^4 \text{Re}^{2,658 f_u - 2,218} \right] } \right\}. \quad (2.35)$$

П. В. Лященко, преобразовав уравнение (2.35) путем подстановки в него скорости осаждения, выраженной через Re , и возведения обеих частей уравнения в квадрат, получил обобщенное критериальное уравнение:

$$\zeta \text{Re} = \frac{4}{3} Ar. \quad (2.36)$$

В критерий Ar скорость осаждения не входит, он состоит из величин, которые либо заданы, либо могут быть заранее определены. Подставив в уравнение (2.36) значения ζ , соответствующие различным режимам осаждения, можно получить аналитические выражения функции $Re = f(Ar)$ для этих режимов.

Для ламинарного течения

$$\text{Re} = \frac{Ar}{18}. \quad (2.37)$$

$$\text{В турбулентном режиме} \quad \text{Re} = \frac{\sqrt{Ar}}{0,6}. \quad (2.38)$$

$$\text{В переходной области} \quad \text{Re}^{1,4} = \frac{4Ar}{55,5}. \quad (2.39)$$

Подставляя в уравнения (2.37) и (2.38) критические значения Re для ламинарного и турбулентного режима, находят критические значения Ar для этих режимов. Переходная область осаждения соответствует изменению критерия Ar в пределах $36 < Ar < 83000$. Таким образом, рассчитав критерий Ar , определяют по его значению область осаждения, рассчитывают критерий Re по соответствующему уравнению (2.37), (2.38) или (2.39) и затем находят скорость осаждения:

$$g_0 = \frac{\text{Re} \mu}{d_{\text{ч}} \rho_{\text{жс}}}$$

Формула для расчета скорости витания одиночной сферической частицы, охватывающая все режимы обтекания, предложена О. М. Годесом

$$\text{Re} = \frac{Ar}{18 + 0.61\sqrt{Ar}}. \quad (2.40)$$

Анализ зависимости коэффициента сопротивления от критерия Re для исследованных сорбентов показывает, что средние скорости оседания, как узких фракций, так и товарных сорбентов лежат в переходной области ($2 < Re < 500$) и достаточно точно описываются уравнением (2.32) с учетом коэффициента формы для цилиндрических частиц.

Критерии Re и Ar в процессах псевдооживления и расширения псевдооживленного слоя в ионообменных процессах широко используются. Слой твердых частиц при малых скоростях восходящего потока жидкости остается неподвижным, но при увеличении скорости до некоторого значения, называемого критической скоростью псевдооживления, твердые частицы свободно взвешиваются в жидкости и образуют псевдооживленный слой. При дальнейшем увеличении скорости слой продолжает расширяться вследствие увеличения среднего расстояния между твердыми частицами.

При очень высокой скорости жидкостного потока твердые частицы уносятся им из системы. Псевдооживленный слой образуется при увеличении скорости восходящего потока оживляющего агента через неподвижный слой. Следовательно, можно предположить, что при скорости начала псевдооживления к псевдооживленному слою применимы закономерности, справедливые для неподвижного слоя. Если же слой расши-

рился до порозности, близкой к единице, и состоит преимущественно из одиночных изолированных частиц, взвешенных в потоке жидкости, то любая зависимость для псевдооживленного слоя при экстраполяции должна оказаться применимой к одиночной частице. В промежуточных условиях однородный псевдооживленный слой по своим гидродинамическим свойствам подобен отстаивающейся суспензии. При этом в псевдооживленном слое частицы в целом не перемещаются относительно стенок аппарата, они поддерживаются восходящим потоком жидкости. В оседающей суспензии твердые частицы непрерывно движутся вниз, а движение жидкости обусловлено ее вытеснением оседающими частицами. Очевидно, что зависимости скорость – порозность для оседающей суспензии и однородного псевдооживленного слоя окажутся сходными.

К основным параметрам, характеризующим гидродинамические закономерности псевдооживленного слоя, относятся значения скорости начала псевдооживления ($W_{кр}$), гидравлического сопротивления (Δp) и степени расширения псевдооживленного слоя (H/H_0 – отношение высоты псевдооживленного слоя частиц H к высоте стационарного слоя частиц H_0). Условием перехода неподвижного слоя в псевдооживленный является равенство перепада давления в слое массе содержащихся в нем твердых частиц с учетом силы Архимеда:

$$\Delta p = (\rho_{\text{ч}} - \rho_{\text{ж}})(1 - \varepsilon_0)H_0, \quad (2.41)$$

ε_0 – порозность стационарного слоя частиц.

Скорость начала псевдооживления вычисляют, пользуясь зависимостью между скоростью потока жидкости и перепадом давления в неподвижном слое по уравнению:

$$\frac{\Delta p}{H} = 150(1 - \varepsilon)^2 \frac{\mu \mathcal{G}_p}{\varepsilon^3 d_c^2} + 1,75(1 - \varepsilon) \frac{\rho_{\text{ж}} \mathcal{G}_p^2}{\varepsilon^3 d_c^2}, \quad (2.42)$$

ε – порозность псевдооживленного слоя частиц; \mathcal{G}_p – скорость движения раствора.

С учётом выражения (2.41) после несложных преобразований, принимая $\varepsilon_0 = 0,4$, получают следующее выражение:

$$Re_{кр} = 25,7 \left(\sqrt{1 + 5,53 \cdot 10^{-5} Ar} - 1 \right). \quad (2.43)$$

Однако более удачная форма расчетного уравнения предложена О. М. Тодесом [19]:

$$Re_{кр} = \frac{Ar}{150(1 - \varepsilon_0) \varepsilon_0^3 + \sqrt{\frac{1,75 Ar}{\varepsilon_0^3}}}, \quad (2.44)$$

при $\varepsilon_0 = 0,4$

$$Re_{кр} = \frac{Ar}{1400 + 5,22\sqrt{Ar}}. \quad (2.45)$$

Величиной, наиболее полно отражающей свойства системы жидкость – твердые частицы, является характеристическая скорость \mathcal{G}_0 . Поэтому представляет особый интерес связь скорости начала псевдооживления с этой величиной. Если принять порозность в точке начала псевдооживления равной 0,4, то отношение

$$\frac{\mathcal{G}_0}{W_{кр}} = \frac{Re_0}{Re_{кр}}$$

можно определить из сопоставления уравнений для расчета скорости витания и начала псевдооживления, уравнений (2.41) и (2.45) или уравнений (2.43) и (2.37), (2.38), (2.39).

Учитывая, что большинства систем ионит – раствор значение критерия Архимеда лежит в переходной области ($10^2 < Ar < 10^5$), можно вывод, что отношение $\mathcal{G}_0/W_{кр}$ для обменных материалов существенно зависит порозности в начала псевдооживления и критерия Архимеда.

В табл. 2.2 приведены основные критерии. Здесь числа H_0 , Re , Gr , Eu , Er – критерии гидродинамического подобия. Критерии Pr , Pe , St , Fo , Ro , Nu – критерии теплового подобия, а Re' , Nu' , Bi' , Ki , Pe' , St' , Fo' – критерии диффузионного подобия.

Тепловое число Прандтля $Pr = \frac{\nu}{a_p}$ для газов практически не зави-

сит от температуры [29], давления и определяется только атомностью газов. Для одноатомных газов $Pr \approx 0,67$; двухатомных $Pr \approx 0,7$, многоатомных $Pr = 1$. Для капельных жидкостей $Pr \approx 10^2 \div 10^3$ и уменьшается с ростом температуры. Для расплавленных металлов, используемых в качестве теплоносителей (литий, натрий и т. д.), $Pr \approx 0,005 \div 0,05$. Диффузионное число Прандтля (Шмидта) $Sc = \nu/D > Pr$ и для жидкостей его значения могут достигать 10^3 и более. Числа Шмидта Sc для газов в смеси с воздухом с погрешностью до 30 % могут быть подсчитаны по формуле $Sc = 0,145\mu^{0,556}$. Для газов коэффициенты диффузии, температуропроводности и кинематической вязкости примерно равны между собой, а безразмерные поля скоростей, температур и концентраций подобны, т. е. имеет место тройная аналогия процессов переноса теплоты, вещества и количества движения (без учета сжимаемости и при безградиентном течении, $grad p = 0$).

В критерии подобия входит характерный линейный размер ℓ . При этом для течения в трубах и каналах принимают $\ell = d_r$, где $d_r = 4A_f / \Pi$ – гидравлический диаметр, A_f – поверхность поперечного сечения, Π – смоченный периметр. Для внешней задачи (при обтекании тела) ℓ есть размер обтекаемого тела, и выбирается в зависимости от конкретных условий. Так, при поперечном обтекании цилиндра и сферы $\ell = d$. При обтекании плоской пластины (стенки) $\ell = x$ – расстояние от кромки поверхности до данной точки.

Критерий Био используется при оценке механизма ионного обмена: так при $Bi' \leq 0,1$ (здесь в качестве линейного размера выбирается радиус зерна сорбента) наблюдается внешняя диффузия, внешнедиффузионная кинетика проявляется в очень разбавленных растворах. Если $Bi' > 30$ определяющее влияние оказывает внутренняя диффузия, внутридиффузионная кинетика преобладает в концентрированных растворах. Диапазон критерия от 0,1 до 30, когда $30 > Bi' \geq 0,1$ указывает на смешанную диффузию [19].

При разделении изотопов в колоннах необходимо обеспечить большое число ступеней и перерабатывать большие материальные потоки с минимальной затратой энергии и задержкой продукта. Сравнительно высокой эффективностью обладают колонны с насадкой из колец Рашига. При теоретическом расчёте высоты единицы переноса (ВЕП) [39] для системы газ–жидкость необходимо знание коэффициентов массоотдачи и поверхности контакта фаз. При этом коэффициент массоотдачи определяется соотношением

$$\alpha = \frac{D}{\delta_{эфф}}, \quad (2.46)$$

где $\delta_{эфф}$ – эффективная толщина диффузионного пограничного слоя, являющаяся обычно функцией критериев подобия – числа Рейнольдса и диффузионного критерия Прандтля $\frac{V}{D} = Pr'$.

$$\delta_{эфф} = \frac{\ell}{Nu}, \quad (2.47)$$

где ℓ – определяющий размер, Nu – диффузионный критерий Нуссельта.

При ламинарном движении в трубах и каналах (внутренняя задача), когда скорости w малы, $Nu' = const$ и $\delta_{эфф}$ пропорциональна размеру ℓ . Однако это простое решение реализуется, как правило, только для газов, для жидкостей же в ряде случаев участок стабилизации превышает длину канала и $Nu' \sim \sqrt{Pe'}$ (формула (2.49)).

При ламинарном движении с внешним обтеканием (насадки)

$$Nu' \approx c \sqrt{Pe'} \quad (2.48)$$

Для газов $\nu = a_p = D$, диффузионный пограничный слой имеет примерно такую же толщину, как гидродинамический и тепловой пограничные слои. Это приводит к аналогии между теплопередачей и массоотдачей от газа, позволяет применить известные зависимости $Nu' = f(RePr')$ для определения коэффициента массоотдачи от газа, заменяя тепловой критерий Нуссельта Nu на диффузионный $Nu' = \frac{\alpha \ell}{D}$, Pr на $Pr' = \frac{\nu}{D}$. Однако возможности использования этой аналогии ограничены случаями движения газа в трубах, в упорядоченной и хаотической насадке.

Особенности массопередачи от жидкостей связаны с тем, что коэффициенты диффузии в жидкостях $D_{жс}$ и молекулярный перенос вещества $j \sim D_{жс}$ — очень малы ($D_{жс} \sim 10^{-5}$ см²/с, тогда как $D_{газ} \sim 10^{-1}$ см²/с), а значение Pr' — очень велико ($\sim 10^3$).

В результате конвективный перенос вещества даже при очень малых скоростях движения жидкости будет больше обусловленного молекулярной диффузией, а диффузионный пограничный слой будет тоньше гидродинамического. На учете этих обстоятельств основана разработанная Левичем теория конвективной диффузии в жидкостях. В частности, было установлено, что при турбулентном режиме турбулентные пульсации, проникающие в вязкий погранслоя, будут переносить значительно больше вещества, чем молекулярная диффузия, тогда как на переносе количества движения и гидродинамике это не скажется, поскольку кинематическая вязкость ν сравнительно велика. Лишь около границы раздела фаз, т. е. в очень тонком слое, пульсации затухают настолько, что молекулярный перенос становится существенным и приводит к резкому изменению концентрации вещества. Толщина диффузионного погранслоя при $Pr' = 10^3$ составляет лишь около 0,1 толщины гидродинамического.

При массообмене с пленкой жидкости толщиной Δ и длиной менее $50 Re \Delta$ и ламинарном режиме движения ($Re < 100$)

$$\begin{aligned} Nu'_{жс} &= c Re^{1/2} (Pr')^{1/2} = c \sqrt{Pe'} \\ Nu'_{жс} &= \frac{\alpha \Delta}{D}, \quad Pe' = \frac{w \Delta}{D} \end{aligned} \quad (2.49)$$

w — средняя скорость стекающей пленки, $c = \sqrt{\frac{6}{\pi}}$

Аналогичное выражение, но с несколько измененным коэффициентом ($c = 1/2$, если за определяющий размер принять диаметр трубы) теория дает и для случая ламинарного движения жидкости в трубах; в обоих случаях диффузионный погранслои не успевают «прорасти» через всю толщину пленки или гидродинамического погранслоя, и осуществляется лишь «входной» участок массообмена (в отличие от гидродинамики и теплопередачи, когда «входной» участок имеет незначительную длину и основная часть трубки работает в стабилизированных условиях).

В случае *турбулентного движения* пленки теория дает соотношение

$$Nu' = c Re^{3/2} (Pr')^{1/2}. \quad (2.50)$$

В ряде случаев полезна аналогия между конвективной диффузией и поверхностным трением (гидродинамическая аналогия), которая приводит к выводу, что критерий поперечного переноса вещества (или фактор массопередачи):

$$\tilde{j}_D = \frac{Nu'}{Pe'} = \frac{\alpha}{w} = \frac{f}{2(Pr')^{2/3}} = \frac{\lambda}{8(Pr')^{2/3}}, \quad (2.51)$$

где f и $\lambda = 4f$ – коэффициенты трения.

Гидродинамическая аналогия и соотношение (2.51) выполняются при турбулентном движении газов в трубе, ламинарном обтекании пластинки жидкостями и газами. В случае тел необтекаемой формы аналогия нарушается $\tilde{j}_D < \frac{f}{2}$.

Более общий характер имеет полуэмпирическое соотношение

$$\tilde{j}_D = \frac{\alpha}{w} = c \frac{\lambda}{8(Pr')^{2/3}}, \quad (2.52)$$

которое выполняется при ламинарном движении газа в трубе ($c = \frac{1}{2}$) и приближенно при вязком и турбулентном движении газа в насадке ($c = 0,06 \div 0,1$)

Широкое применение гидродинамической аналогии для решения задач массопередачи и теплопередачи связано с работами Кольборна, который записывает соотношение (2.52) в форме:

$$\tilde{j}_D = \tilde{j}_{\text{менл}} = \tilde{j}_{\text{мп}}, \quad (2.53)$$

$$\tilde{j}_D = \frac{Nu'}{Re \sqrt[3]{Pr'}} = \tilde{j}_D (Pr')^{2/3}, \quad (2.54)$$

где \tilde{j}_D – критерий Кольборна, аналогично:

$$\tilde{j}_{менл} = \frac{Nu}{Re\sqrt{Pr}}, \quad \tilde{j}_{mp} = \frac{1}{2}. \quad (2.55)$$

Аналогия между теплопередачей и массопередачей, гидродинамическая аналогия и, наконец, теоретические соотношения (2. 49) – (2. 51) могут быть использованы только для расчета аппаратов с пленочным течением жидкости и фиксированной поверхностью контакта фаз, например трубок с орошаемыми стенками, пакетной насадки. В случае насадочных колонн возникает дополнительное затруднение – необходимость определения поверхности контакта фаз. Самым же сложным случаем являются барботажные тарелки, когда поверхность контакта фаз и коэффициенты массопередачи наиболее сильно зависят от гидродинамической обстановки.

Рассмотрим особенности массопередачи и возможности расчета ВЕП с использованием критериев подобия для нескольких наиболее важных типов насадок и тарелок.

Как показали многочисленные исследования, при пленочной ректификации в трубках массопередача, как правило, лимитируется диффузионным сопротивлением со стороны газа. Уравнение конвективной диффузии в случае вязкого движения газа в трубке имеет вид

$$\frac{D}{r} \frac{d}{dr} r \frac{dc}{dr} = v \frac{dc}{dz}, \quad (2.56)$$

где D – коэффициент диффузии;

$v = v(r)$ – скорость движения газа;

$c = c(r, z)$ – концентрация извлекаемого компонента в газе.

Уравнение (2.56) аналогично уравнению распространения тепла в вязком потоке жидкости, решение которого имеет вид $Nu' = const$ (в случае постоянной температуры стенки $Nu = 3,66$, при постоянном тепловом потоке $Nu = 48/11$). Полагая

$$Nu = Nu' = \beta \frac{2r_0}{D}, \quad (2.57)$$

(где r_0 – радиус трубки), с помощью соотношения для ВЕП:

$$h = \frac{G}{k S} = \frac{w_y F_y}{\beta_c S} = \frac{w_y}{\beta_c} \frac{F_y}{S F} = \frac{w}{\beta_c S} \quad (2.58)$$

w_y – средняя скорость газа, G – поток газовой фазы, w – фиктивная скорость газа, отнесённая к полному сечению колонны, F_y, F – сечение, занятое потоком газовой фазы, и полное сечение колонны, S – удельная поверхность колонны, $\text{м}^2/\text{м}^3$, $\bar{S} = S F$, $G = w_y F_y n_G$, $\beta = n \beta_c$ – коэффициент массопередачи, n – плотность или суммарная концентрация вещества.

При $F_y = \pi r_0^2$, $\bar{S} = 2 \pi r_0$ находится составляющая ВЕП, определяемую истинным коэффициентом массопередачи:

$$h_0 = \frac{\bar{v} r}{2\alpha} = \frac{w r_0^2}{D Nu'}, \quad (2.59)$$

где $\bar{v} = w$ – средняя скорость газа. Принимая значение Nu' для случая постоянного теплового потока и учитывая продольную диффузию в газовой фазе:

$$h = h_0 + h_\circ = \frac{11}{48} \frac{w r_0^2}{D} + \frac{D}{w}. \quad (2.60)$$

Вводя диффузионное число Пекле $Pe' = \frac{2 w r_0}{D} = \frac{w d_0}{D}$, можно формулу (2.60) записать в виде:

$$\frac{h}{d} = 0.058 Pe' + \frac{1}{Pe'}. \quad (2.61)$$

Подобный вид имеют выражения для ВЕП в случае всех процессов разделения, сопровождающихся вязким движением газов, в том числе, при термодиффузии и масс-диффузий. Основная часть ВЕП пропорциональна числу Пекле, диффузионная добавка h_\circ обратно пропорциональна Pe' .

Нетрудно видеть также, что h_\circ играет существенную роль лишь при малых скоростях движения газа, когда ВЕП мала; как следует из соотношения (2.61):

$$\frac{h_\circ}{h_0} = \frac{17}{(Pe')^2}. \quad (2.62)$$

Таким образом, при $Pe' > 10$, когда $h > 6d$, ею можно пренебречь. По мере уменьшения скорости газа или Pe' снижается и ВЕП, однако продольная диффузия ставит предел уменьшению ВЕП (рис. 2.2).

Дифференцируя по Pe' и полагая $\frac{dh}{dPe'} = 0$, можно найти, что h имеет минимум при $Pe' \approx 4$, когда $h_0 = h_\circ$; при этом

$$h_{\min} = \sqrt{\frac{11}{12}} r_0 = 0,96 r_0. \quad (2.63)$$

Таким образом, при ректификации в трубках можно получить очень низкие ВЕП, например, в трубке диаметром 10 мм $h_{\min} = 5$ мм, что позволяет при высоте 1 м иметь 200 ступеней. Однако производительность колонны на таком режиме крайне мала (при $d = 1$ см и $D \sim 0,3$ см²/с значению $Pe' = 4$ соответствует скорость $w = \frac{4D}{d} \sim 1$ см/с), что осложняет практическое использование данного эффекта; в результате Кун, исследовавший работу таких колонн, в дальнейшем применял колонны с насадкой.

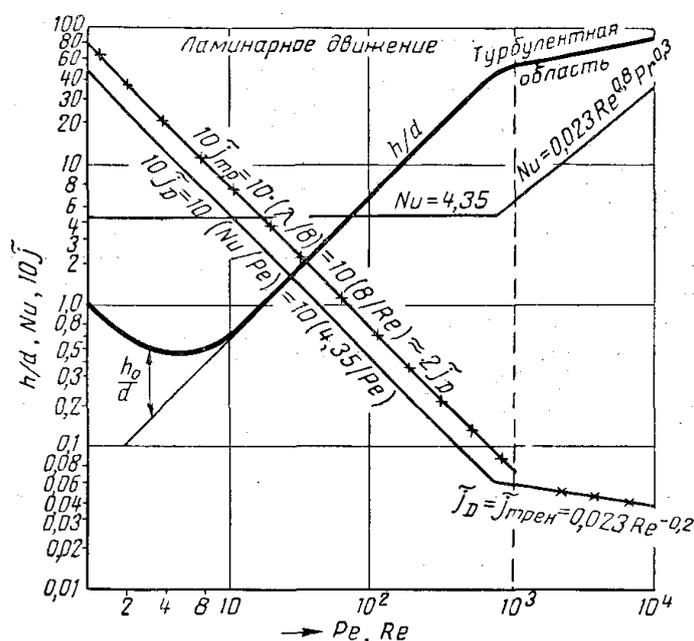


Рис. 2.2. Основные характеристики дистилляции в трубках при $Pr' \approx 1$ [39]

При изменении формы канала ВЕП определяется соотношениями вида (2.60) – (2.61) с другими численными коэффициентами. Так, основная составляющая ВЕП, в случае плоского канала (с зазором между стенками δ) с одной (h_1) и двумя (h_2) орошаемыми стенками будет

$$h_1 = \frac{w\delta}{2\alpha} = \frac{13}{35} \frac{w\delta^2}{D} = \frac{13}{35} \delta Pe'. \quad (2.64)$$

Найдем с помощью уравнения (2.64) эффективную толщину пленки $\delta_{эфф}$ для случая одной орошаемой стенки. Полагая $\alpha = \frac{D}{\delta_{эфф}}$, найдем,

что $\delta_{эфф} = \frac{13}{35} \delta$.

При турбулентном движении газа в трубке ($Re > 1000$) можно применить аналогию с теплопередачей (принимая $Nu' = 0,023 Re^{0,8} (Pr')^{1/3}$) или гидродинамическую аналогию (принимая $\lambda = 0,18 Re^{-0,2}$), что дает

$$\tilde{j}_D = 0,023 Re^{-0,2} (Pr')^{-2/3}, \quad h_{0\Gamma} = \frac{w d}{4 \beta_{\text{газ}}} \approx \frac{d}{4 j_D} = 10,9 d Re_{\text{газ}}^{0,2} Pr_{\text{газ}}^{2/3}. \quad (2.65)$$

При $1000 < Re < 15000$ формула (2.65) будет иметь вид:

$$h_{0\Gamma} = 11,1 d^{0,64} Re_{\text{газ}}^{0,23} Pr_{\text{газ}}^{2/3}, \quad (2.66)$$

где $h_{0\Gamma}$ и d выражаются в сантиметрах.

Сравнение выражений (2.61) и (2.66) показывает, что переход к турбулентному режиму ослабляет зависимость ВЕП от нагрузки при ламинарном режиме (и $Pe' > 10$) ВЕП растет пропорционально скорости газа, при турбулентном – пропорционально $w^{0,2}$; в случае каналов рост ВЕП вообще прекращается, а в насадочных колоннах, когда с ростом нагрузки возрастает смоченная поверхность, ВЕП при турбулентном режиме движения газа уменьшается с ростом w , эта особенность турбулентного режима приводит к возрастанию разделительной мощности колонн с ростом нагрузки, тогда как при ламинарном режиме, при $Pe' > 10$, разделительная мощность от нагрузки не зависит.

Эмпирическая формула, аналогичная уравнению (2.61), была получена Малюсовым, Умником и Жаворонковым для колонн с орошаемыми стенками и вращающимся ротором, когда число Рейнольдса для вертикального потока газа $Re_{\text{газ}} < 800 \div 1000$ (ламинарный режим):

$$h_{0\Gamma} = 0,236 d_{\text{экв}} Pe' Re_{\text{вр}}^{-1/2} + 1,6 d_{\text{экв}}, \quad (2.67)$$

где $Re_{\text{вр}}$ – число Рейнольдса для вращения ротора,

$$Re_{\text{вр}} = \frac{\pi d_{\text{ср}} n 2 \Delta \rho}{60 \mu}, \quad (2.68)$$

n – число оборотов; Δ – зазор между ротором и внутренними стенками колонны; $d_{\text{ср}} = \frac{D_e + d}{2}$, D_e – внутренний диаметр колонны; d – внешний диаметр ротора, $d_{\text{экв}} = \frac{4 F}{\pi d}$, F – сечение для прохода паров.

В случае применения листовой пакетной насадки при ректификации лимитирует массопередача со стороны газа и при увеличении нагрузки с высокими числами Пекле ($Pe \approx 1000$ и более), которые име-

ют место в производственных колоннах, ВЕП значительно возрастает, и диффузионная добавка $h_{\delta} \sim \frac{1}{Pe}$ не существенна. При ламинарном пото-

ке паров (до $Re = \frac{w d_{\text{экв}}}{\nu} = 1000$) ВЕП может быть определена по формуле Жаворонкова и Малюсова, отличающейся от выражения (2.61) численным коэффициентом и отсутствием диффузионной добавки:

$$h = 0,069 d_{\text{экв}} Re Pr', \quad (2.69)$$

где $d_{\text{экв}} = \frac{4F}{\tilde{S}}$, \tilde{S} – смоченный периметр (для пакетов с расстоянием δ между листами $d_{\text{экв}} = 2\delta$).

При турбулентном движении таза ($Re_{\text{газ}} > 1000$) справедливо соотношение (2.66). Таким образом, $h_{0Г}$ возрастает с нагрузкой, как $Re_{\text{газ}}^{0,23}$; при $Re_{\text{газ}} = 1000 \div 3000$ и расстоянии между листами 10 мм $h_{0Г} = 0,6 \div 0,9$ м.

Исследуя абсорбцию на пакетной насадке, Малюсов и Жаворонков нашли значение коэффициентов массоотдачи от обеих фаз

$$Nu'_{\text{ж}} = \frac{\alpha_{\text{ж}} \Delta}{D_{\text{ж}}} = 0,011 Re_{\text{ж}} (Pr'_{\text{ж}})^{0,5}; \quad (2.70)$$

$$Nu'_{\text{газ}} = \frac{\alpha_{\text{газ}} d_{\text{экв}}}{D_{\text{газ}}} = 0,0013 \left(\frac{a}{10} \right)^{0,45} Re_{\text{отн}}^{0,19} (Pr'_{\text{газ}})^{0,5},$$

где Δ – толщина пленки; a – ширина канала в мм; $Re_{\text{ж}} = \frac{4 w_{\text{ж}} \Delta}{\nu_{\text{ж}}}$; $Re_{\text{отн}}$ –

число Рейнольдса, рассчитанное по относительной скорости, т. е. принимая $w = w_{Г} + w_{\text{ж}}$, где $w_{\text{ж}}$ – скорость стекания пленки.

Если массопередача лимитируется газовой пленкой, то, как показали Семенов и Шварцштейн, приближенно справедлива гидродинамическая аналогия, когда отклонения при высоких скоростях газов связаны с диффузионным сопротивлением жидкой фазы, которое не учитывалось ранее.

Уравнение (2.70) для Nu' , справедливое при $4 < Re_{\text{ж}} < 300$, отличается от теоретических соотношений (2.49) и (2.51), что может объясняться как приближенным характером теории, так и сравнительно небольшим интервалом охваченных экспериментом значений $Re_{\text{ж}}$.

На основе критериев подобия определяются предельные нагрузки и гидравлическое сопротивление. Если по какой-либо причине процесс должен вестись только при вязком режиме движения газа или жидкости,

нагрузки ограничиваются критическим значением числа Рейнольдса (для газов в трубах $Re_{кр} = 1000 \div 2000$; для пленки жидкости ламинарный режим имеет место при $Re_{ж} < 50$, волновой – при $50 < Re_{ж} < 1500$). Если ограничения, связанные с турбулентностью, отсутствуют, то предельные нагрузки определяются условиями захлебывания колонн [39]; эти условия исследовались теоретически и экспериментально; формулы, определяющие предельные нагрузки, а также гидравлическое сопротивление некоторых контактных устройств приведены в табл. 2.3.

Таблица 2.3

Гидравлические характеристики насадок и тарелок

Тип контактного устройства	Коэффициент сопротивления	
	ламинарный	турбулентный
	режимы	
Трубки	$\lambda = \frac{64}{Re_{газ}} \quad (Re < 2000)$	$\lambda = \frac{0,023}{Re_{газ}^{0,2}}$
Плоские каналы параллельные листы	$\lambda = \frac{96}{Re} \quad (Re_{газ} < 560)$ $\lambda = \frac{16,5}{Re_{газ}^{0,65}} \quad (Re_{газ} < 2500)$	$\lambda = \frac{0,5}{Re_{газ}^{0,26}}$
Пакетная насадка	$\lambda = \frac{6,6}{Re_{газ}^{0,55}} \quad (Re_{газ} = 400 \div 6000)$	
Хаотическая насадка	$\lambda = \frac{140}{Re_{газ}} \quad (Re < 40)$	$\lambda = \frac{16}{Re_{газ}^{0,2}}$
Ситчатые тарелки	$\lambda = \left[(1 - F_c)^2 + \zeta_p + 4000 \frac{F_c d \delta}{d_{экв}} Re_{газ}^{-0,2} \right]$ <p>d – ширина щели, δ – толщина листа, ζ_p – коэффициент потерь за счёт сужения потока в $(1 - F_c)$ раз, F_c – свободное сечение тарелки (отношение площади отверстий к площади тарелки)</p>	

§ 4. Критериальные уравнения

Согласно второй теореме подобия, решение дифференциального уравнения может быть дано в виде соотношения между критериями подобия, следующими из этого уравнения. Для установившихся наиболее характерных процессов переноса теплоты [29] *функциональное уравнение теплообмена* в критериальной форме имеет вид:

$$Nu = f(Re, Gr, Pr), \quad (2.71)$$

для вынужденной конвекции

$$Nu = f(Re, Pr), \quad (2.72)$$

для свободной конвекции

$$Nu = f(Gr, Pr), \quad (2.73)$$

для процессов переноса вещества

$$Nu' = f(Re, Sc), \quad (2.74)$$

$$Nu' = f(Gr, Sc). \quad (2.75)$$

Если функциональные зависимости являются степенными, то для свободной конвекции

$$Nu = C_1 Pr^{n_1} Gr^{m_1}, \quad (2.76)$$

$$Nu' = C_2 Pr^{n_2} Gr^{m_2}, \quad (2.77)$$

для вынужденной конвекции

$$Nu = C_3 Pr^{n_3} Gr^{m_3}, \quad (2.78)$$

$$Nu' = C_4 Pr^{n_4} Gr^{m_4}, \quad (2.79)$$

где C , n , m – постоянные, определяемые аналитическим (для отдельных случаев) или опытным путем и зависящие от конкретных условий.

Необходимо учитывать, что постоянные C , m , n в критериальных уравнениях получены для строго определенного диапазона критериев Re , Gr , Pr , который указывается для каждого типа уравнения. Например, если теплообмен при обтекании шара описывается уравнением

$$Nu = 0,54 Re^{0,5} \text{ при } Pr \approx 1, Re > 200, \quad (2.80)$$

это означает, что данное уравнение пригодно лишь для газов ($Pr \approx 1$) и не годится для капельных жидкостей ($Pr > 1$) и расплавов металлов ($Pr \ll 1$). Так как $Re = wL/\nu = wd/\nu > 200$, то возможны различные комбинации значений w , d , и $\nu = w/\rho$ при условии, что величина wd/ρ всегда будет больше 200.

Достоинством критериальных уравнений является возможность их использования для всех подобных процессов, т. е. процессов, у которых подобны условия однозначности и равны все определяющие критерии (правило Кирпичева – Гухмана). Применительно к уравнениям (2.71 ÷ 2.79) это означает возможность их использования (при выполнении всех условий подобия) для любой комбинации размерных переменных (ско-

рость, температура, физические свойства, определяющий размер), входящих в определяющие критерии, если численные значения этих критериев будут находиться в диапазоне, в котором получено и может быть использовано данное критериальное уравнение.

Контрольные вопросы и задания

1. Какие критерии подобия применяются в расчётах тепломассообменных процессов?
2. Что характеризует критическое число Рейнольдса? Приведите расчётные формулы, от чего зависит величина $Re_{кр}$.
3. Покажите взаимосвязь числа Архимеда и порозности слоя частиц.
4. Определите скорость частицы, равномерно движущейся в жидкости при $d_ч = 0,2$ см, $\rho_ч = 1,95$ г/см³, $\rho_ж = 1,02$ г/см³ и $Re = 150$.
5. Назовите критерии гидродинамического подобия.
6. Охарактеризуйте критерии теплового подобия.
7. Запишите выражения, определяющие критерии диффузионного подобия.
8. Чем отличаются тепловой и диффузионный критерий Нуссельта? Приведите пример использования данных величин.
9. Какие значения может принимать тепловое число Прандтля для одноатомных, двухатомных, многоатомных газов? В каком диапазоне изменяется величина Pr для жидких лития, натрия?
10. Объясните, с помощью какого критерия проводится оценка механизма ионного обмена, приведите диапазоны изменения критерия и вид кинетики при этом.
11. Может ли критерий подобия быть представлен в виде произведения критериев? Приведите примеры.
12. С помощью какого критерия определяется ВЕП?
13. Функцией, каких критериев подобия, является величина эффективной толщины диффузионного пограничного слоя?
14. Какие обобщающие критериальные уравнения Вы знаете? Приведите примеры.

ГЛАВА 3

ИНФОРМАЦИОННОЕ МОДЕЛИРОВАНИЕ С ПРИМЕНЕНИЕМ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

Новые способы представления информации, ее передачи и использования появляются ежегодно. Одними из самых свежих и самых динамичных направлений развития является информационное моделирование с использованием методик искусственных нейронных сетей. Совсем недавно известные только узкому кругу специалистов нейронные сети (НС), нечеткая логика, генетические алгоритмы и ряд других информационных технологий получили широкое применение за последнее десятилетие. Появился термин *интеллектуальные системы* – это системы, способные к «пониманию» и обучению в отношении возмущений, внешней среды и условий работы. Основное отличие интеллектуальных систем – наличие механизма системной обработки знаний.

В основе создания интеллектуальных систем лежат два принципа: ситуационное управление (управление на основе анализа внешних ситуаций или событий) и использование современных информационных технологий обработки знаний.

Существует несколько современных информационных технологий, позволяющих создавать данные системы: экспертные системы, искусственные нейронные сети, нечеткая логика, генетические алгоритмы и ряд других. Интеллектуальные технологии между собой различает прежде всего то, что именно положено в основу концепции интеллектуальности – либо умение работать с формализованными знаниями человека (экспертные системы, нечеткая логика), либо свойственные человеку приемы обучения и мышления (искусственные нейронные сети и генетические алгоритмы).

Под *искусственными нейронными сетями* (далее – просто нейронными сетями, НС) подразумевают вычислительные структуры, состоящие из большого количества однотипных элементов, каждый из которых выполняет относительно простые функции. Процессы в искусственных НС иногда ассоциируют с процессами, происходящими в нервной системе живых организмов. НС сходна с человеческим мозгом с двух точек зрения:

- Знания поступают в нейронную сеть из окружающей среды и используются в процессе обучения.
- Для накопления знаний используются связи между нейронами, называемые синаптическими весами.

Искусственные нейронные сети (ИНС) строятся по принципу организации и функционирования их биологических аналогов. Они способны решать широкий круг задач распознавания образов, идентификации, прогнозирования, оптимизации. Дальнейшее повышение производительности компьютеров связывают с ИНС, в частности с нейрокомпьютерами (НК), основу которых составляет искусственная нейронная сеть.

Бурно развивающийся в последние годы аппарат НС предназначался в начале в основном для решения задач классификации, кластеризации и распознавания образов, но дальнейшее развитие данного направления значительно расширило сферу применения нейросетевого подхода.

Термин «нейронные сети» сформировался к середине 50-х годов XX века. Основные результаты в этой области связаны с именами У. Мак-Каллока, Д. Хебба, Ф. Розенблатта, М. Минского, Дж. Хопфилда. Приведем краткую историческую справку [25].

1943 г. У. Мак-Каллок (W. McCulloch) и У. Питтс (W. Pitts) предложили модель нейрона и сформулировали основные положения теории функционирования головного мозга.

1949 г. Д. Хебб (D. Hebb) высказал идеи о характере соединений нейронов мозга и их взаимодействии (клеточные ансамбли, синаптическая пластичность). Впервые предложил правила обучения нейронной сети

1957 г. Ф. Розенблатт (F. Rosenblatt) разработал принципы организации и функционирования перцептронов, предложил вариант технической реализации первого в мире нейрокомпьютера Mark.

1959 г. Д. Хьюбел (D. Hubel) и Т. Визель (T. Wiesel) показали распределенный и параллельный характер хранения и обработки информации в биологических нейронных сетях.

1960–1968 гг. Активные исследования в области искусственных нейронных сетей, например, Адалина и Мадалина, В. Уидроу (1960–1962 гг.), ассоциативные матрицы К. Штайнбухг (1961 г.).

1969 г. Публикация книги М. Минского (M. Minsky) и С. Пейперта (S. Papert) «Перцептроны», в которой доказывается принципиальная ограниченность возможностей перцептронов. Угасание интереса к искусственным нейронным сетям.

1970–1976 гг. Активные разработки в области перцептронов в СССР (основные заказчики – военные ведомства).

Конец 1970-х гг. Возобновление интереса к искусственным нейронным сетям как следствие накопления новых знаний о деятельности мозга, а также значительного прогресса в области микроэлектроники и компьютерной техники.

1982–1985 гг. Дж. Хопфилд (J. Hopfield) предложил семейство оптимизирующих нейронных сетей, моделирующих ассоциативную память.

1985 г. Появление первых коммерческих нейрокомпьютеров, например, Mark III фирмы TRW (США).

1987 г. Начало широкомасштабного финансирования разработок в области ИНС и НК в США, Японии и Западной Европе (японская программа «Human Frontiers» и европейская программа «Basic Research in Adaptive Intelligence and Neurocomputing»).

1989 г. Разработки и исследования в области ИНС и НК ведутся практически всеми крупными электротехническими фирмами. Нейрокомпьютеры становятся одним из самых динамичных секторов рынка (за два года объем продаж вырос в пять раз). Агентством DARPA (Defence Advanced Research Projects Agency) министерства обороны США начато финансирование программы по созданию сверхбыстродействующих образцов НК для разнообразных применений.

1990 г. Активизация советских исследовательских организаций в области ИНС и НК (Институт кибернетики им. Глушкова в Киеве, Институт многопроцессорных вычислительных систем в Таганроге, Институт нейрокибернетики в Ростове-на-Дону). Общее число фирм, специализирующихся в области ИНС и НК, достигает трехсот.

1991 г. Годовой объем продаж на рынке ИНС и НК приблизился к 140 млн долларам. Создаются центры нейрокомпьютеров в Москве, Киеве, Минске, Новосибирске, С. Петербурге.

1992 г. Работы в области ИНС находятся стадии интенсивного развития. Ежегодно проводится десятки международных конференций и форумов по нейронным сетям, число специализированных периодических научных изданий по указанной тематике достигло двух десятков наименований.

1996 г. Число международных конференций по ИНС и НК достигло ста.

1997 г. Годовой объем продаж на рынке ИНС и НК превысил 2 млрд долларов, а ежегодный прирост составил 50 %.

2000 г. Переход на субмикронные и нанотехнологии, а также успехи молекулярной и биомолекулярной технологии приводят к принципиально новым архитектурным и технологическим решениям по созданию нейрокомпьютеров.

Глубокое изучение ИНС требует знания нейрофизиологии, науки о познании, психологии, физики (статистической механики), теории управления, теории вычислений, проблем искусственного интеллекта, статистики/математики, распознавания образов, компьютерного зрения, параллельных вычислений и аппаратных средств (цифровых и аналоговых). С другой стороны, ИНС также стимулируют эти дисциплины, обеспечивая их новыми инструментами и представлениями. Этот симбиоз жизненно

необходим для исследования нейронных сетей. Представим некоторые проблемы, решаемые искусственными нейронными сетями.

Задачи, решаемые информационными системами, в большинстве случаев можно свести к ряду типовых [42], среди которых можно выделить следующие:

1. *Классификация образов* – определение принадлежности образа к одному или нескольким предварительно определенным классам.
2. *Кластеризация/категоризация* – разделение образов на заранее не определенные классы по каким-либо признакам.
3. *Аппроксимация функций* – оценку неизвестной зависимости по экспериментальным данным.

Предположим, что имеется обучающаяся выборка $((x_1, y_1), (x_2, y_2), \dots, (x_N, y_N))$, которая генерируется неизвестной функцией, искаженной шумом. Задача аппроксимации состоит в нахождении оценки этой функции.

4. *Прогноз (предсказание)* – определение будущего процесса по его прошлому и настоящему.

Пусть заданы N дискретных отсчетов $\{y(t_1), y(t_2), \dots, y(t_N)\}$ в последовательные моменты времени t_1, t_2, \dots, t_N . Задача состоит в предсказании значения $y(t_{N+1})$ в момент t_{N+1} . Прогноз имеет значительное влияние на принятие решений в бизнесе, науке и технике.

5. *Оптимизация* – нахождение решений, которые максимизируют или минимизируют определенный критерий качества при заданных ограничениях.

Многочисленные проблемы в математике, статистике, технике, науке, медицине и экономике могут рассматриваться как проблемы оптимизации. Задачей оптимизации является нахождение решения, которое удовлетворяет системе ограничений и максимизирует или минимизирует целевую функцию

6. *Память*, адресуемая по содержанию (ассоциативную память), а также доступная по указанному содержанию.

В модели вычислений фон Неймана обращение к памяти доступно только посредством адреса, который не зависит от содержания памяти. Более того, если допущена ошибка в вычислении адреса, то может быть найдена совершенно иная информация. Память, адресуемая по содержанию, или ассоциативная память, доступна по указанию заданного содержания. Содержимое памяти может быть вызвано даже по частичному или искаженному содержанию. Ассоциативная память чрезвычайно желательна при создании перспективных информационно-вычислительных систем.

7. *Управление* – перевод и поддержание системы в требуемом состоянии.

Рассмотрим динамическую систему, заданную совокупностью $\{u(t), y(t)\}$, где $u(t)$ является входным управляющим воздействием,

а $y(t)$ – выходом системы в момент времени t . В системах управления с эталонной моделью целью управления является расчет такого входного воздействия $u(t)$, при котором система следует по желаемой траектории, диктуемой эталонной моделью.

§ 1. Основы теории нейронных сетей

Нейрон представляет собой единицу обработки информации в нейронной сети. На блок-схеме рис. 3.2 показана модель (model) нейрона, лежащего в основе искусственных нейронных сетей. В этой модели можно выделить три основных элемента.

1. Набор *синапсов* (synapse) или *связей* (connecting link), каждый из которых характеризуется своим весом (weight) или силой (strength). В частности, сигнал x_j на входе синапса j , связанного с нейроном k , умножается на вес w_{kj} . Важно обратить внимание на то, в каком порядке указаны индексы синаптического веса w_{kj} . Первый индекс относится к рассматриваемому нейрону, а второй – к входному окончанию синапса, с которым связан данный вес. В отличие от синапсов мозга синаптический вес искусственного нейрона может иметь как положительные, так и отрицательные значения.
2. *Сумматор* (adder) складывает входные сигналы, взвешенные относительно соответствующих синапсов нейрона. Эту операцию можно описать как *линейную комбинацию*.
3. *Функция активации* (activation function) ограничивает амплитуду выходного сигнала нейрона. Эта функция также называется функцией сжатия (squashing function). Обычно нормализованный диапазон амплитуд выхода нейрона лежит в интервале $[0,1]$ или $[-1,1]$.

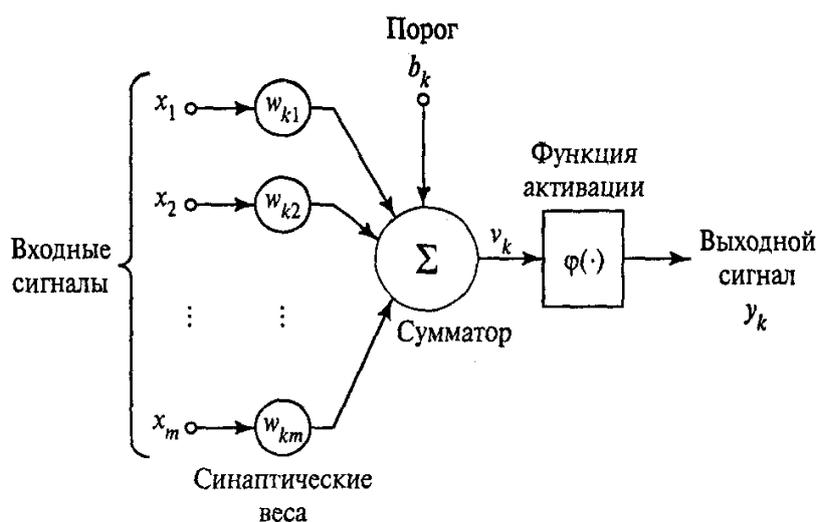


Рис. 3.1. Нелинейная модель нейрона [43]

§ 1.1. Модели нейрона

В модель нейрона, показанную на рис. 3.1, включен пороговый элемент (bias), который обозначен символом b_k . Эта величина отражает увеличение или уменьшение входного сигнала, подаваемого на функцию активации.

В математическом представлении функционирование нейрона k можно описать следующей парой уравнений:

$$u_k = \sum_{j=1}^m w_{kj} x_j, \quad (3.1)$$
$$y_k = \phi(u_k + b_k),$$

где x_1, x_2, \dots, x_m – входные сигналы; w_1, w_2, \dots, w_{km} – синаптические веса нейрона k ; u_k – линейная комбинация входных воздействий (linear combiner output); b_k – порог; $\phi(\cdot)$ – функция активации (activation function); y_k – выходной сигнал нейрона.

Использование порога b_k обеспечивает эффект аффинного преобразования (affine transformation) выхода линейного сумматора u_k . В модели, показанной на рис. 3.2, постсинаптический потенциал вычисляется следующим образом:

$$\mathcal{G}_k = u_k + b_k. \quad (3.2)$$

В частности, в зависимости от того, какое значение принимает порог b_k , положительное или отрицательное, индуцированное локальное поле (induced local field) или потенциал активации (activation potential) \mathcal{G}_k нейрона k изменяется так, как показано на рис. 3.2.

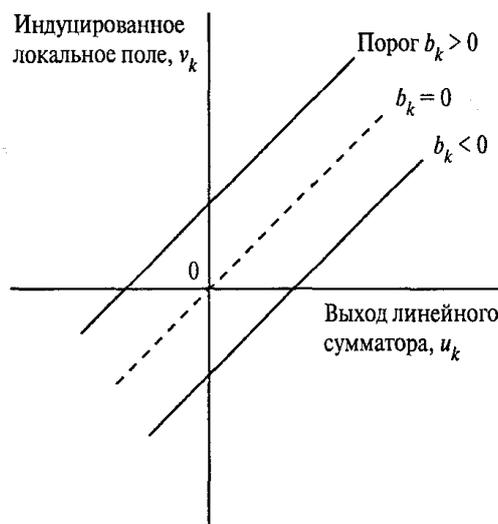


Рис. 3.2. Аффинное преобразование, вызванное наличием порога

Порог b_k является внешним параметром искусственного нейрона k . Его присутствие мы видим в выражении (3.1), которое преобразуем с учётом (3.4.)

$$\mathcal{G}_k = \sum_{j=1}^m w_{kj} x_j, \quad (3.3)$$

$$y_k = \phi(\mathcal{G}_k).$$

В выражении (3.3) добавился новый синапс. Его входной сигнал равен:

$$x_0 = +1, \quad (3.4)$$

а его вес:

$$w_{k0} = b_k. \quad (3.5)$$

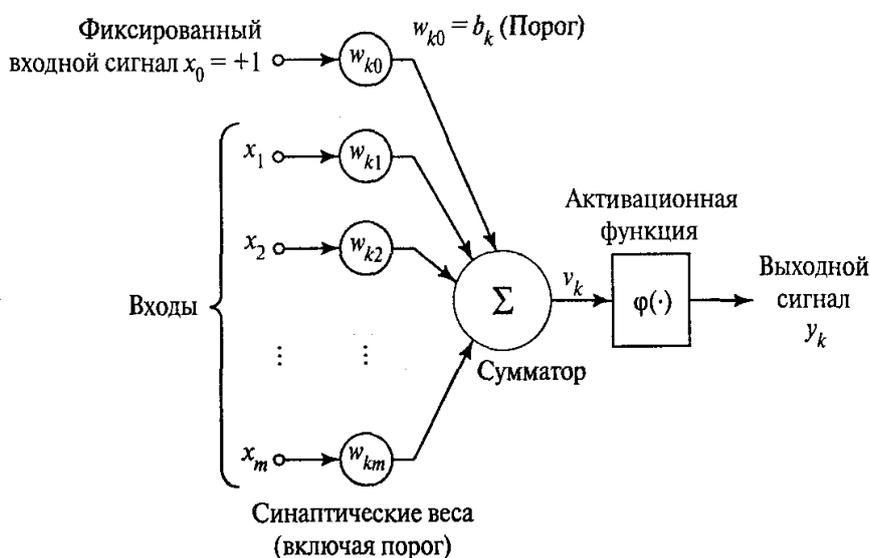


Рис. 3.3. Нелинейная модель нейрона

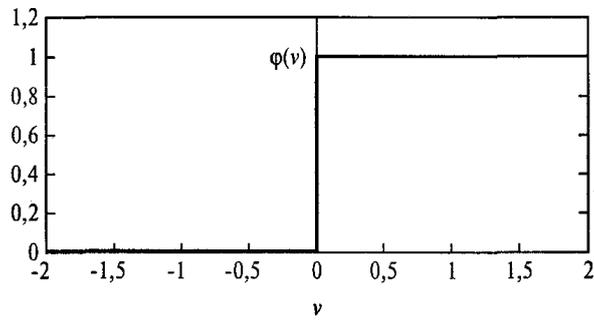
Это позволило трансформировать модель нейрона к виду, показанному на рис. 3.3. На этом рисунке видно, что в результате введения порога добавляется новый входной сигнал фиксированной величины +1, а также появляется новый синаптический вес, равный пороговому значению b_k .

Хотя модели, показанные на рис. 3.2 и 3.3, внешне совершенно не схожи, математически они эквивалентны [43].

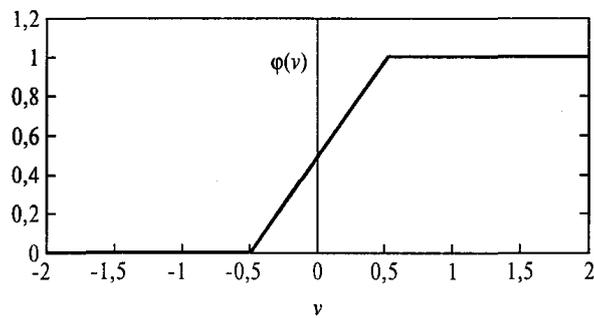
Функции активации, представленные в формулах как $\phi(\mathcal{G})$, определяют выходной сигнал нейрона в зависимости от индуцированного локального поля \mathcal{G} . Можно выделить три основных типа функций активации.

1. *Функция единичного скачка*, или пороговая функция (threshold function). Этот тип функции показан на рис. 3.4, а и описывается следующим образом:

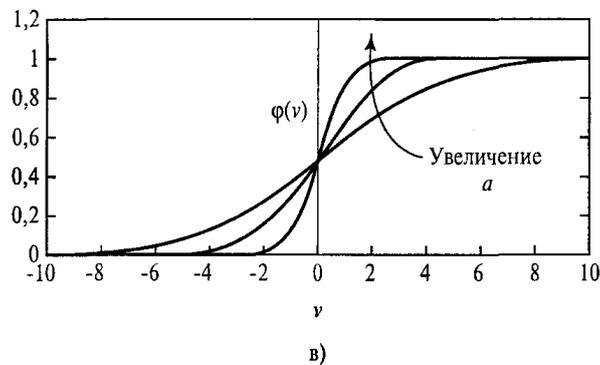
$$\varphi(\mathcal{G}) = \begin{cases} 1, & \text{если } \mathcal{G} \geq 0 \\ 0, & \text{если } \mathcal{G} < 0 \end{cases} \quad (3.6)$$



а)



б)



в)

Рис. 3.4. Виды активационных функций [43]:
 функция единичного скачка (а); кусочно-линейная функция (б)
 и сигмоидальная функция для различных значений параметра a (в)

В технической литературе эта форма функции единичного скачка обычно называется функцией Хэвисаида (Heaviside function). Соответственно выходной сигнал нейрона к такой функции можно представить как

$$y_k = \begin{cases} 1, & \text{если } \mathcal{G}_k \geq 0 \\ 0, & \text{если } \mathcal{G}_k < 0 \end{cases} \quad (3.7)$$

где \mathcal{G}_k — это индуцированное локальное поле нейрона, т. е.

$$\mathcal{G}_k = \sum_{j=1}^m w_{kj} x_j + b_k \quad (3.8)$$

Эту модель в литературе называют моделью МакКаллока-Питца (McCulloch-Pitts model), отдавая дань пионерской работе [8]. В этой модели выходной сигнал нейрона принимает значение 1, если индуцированное локальное поле этого нейрона не отрицательно, и 0 – в противном случае. Это выражение описывает свойство «все или ничего» модели Мак-Каллока – Питца.

2. *Кусочно-линейная функция* (piecewise-linear function), показанная на рис. 3.4, б, описывается следующим выражением:

$$\varphi(\mathcal{G}) = \begin{cases} 1, & \mathcal{G} \geq +\frac{1}{2}; \\ |\mathcal{G}|, & +\frac{1}{2} > \mathcal{G} > -\frac{1}{2}; \\ 0, & \mathcal{G} \leq -\frac{1}{2}, \end{cases} \quad (3.9)$$

где коэффициент усиления в линейной области оператора предполагается равным единице. Эту функцию активации можно рассматривать как аппроксимацию (approximation) нелинейного усилителя. Следующие два варианта можно считать особой формой кусочно-линейной функции.

Выходное значение нейрона лежит в диапазоне [0, 1]. Ценные свойства сигмоидальной функции – дифференцируемость на всей оси абсцисс и простое выражение для ее производной, что используется в некоторых алгоритмах обучения. Кроме того, она обладает свойством усиливать слабые сигналы лучше, чем большие, что предотвращает насыщение от больших сигналов, так как они соответствуют областям аргументов, где сигмоид имеет пологий наклон. Известно также большое количество других разновидностей активационных функций.

- Если линейная область оператора не достигает порога насыщения, он превращается в линейный сумматор (linear combiner).
- Если коэффициент усиления линейной области принять бесконечно большим, то кусочно-линейная функция вырождается в пороговую (threshold function).

3. *Сигмоидальная функция* (sigmoid function). Сигмоидальная функция, график которой напоминает букву S, является, пожалуй, самой распространенной функцией, используемой для создания искусственных нейронных сетей. Это быстро возрастающая функция, которая поддерживает баланс между линейным и нелинейным поведением. Примером сигмоидальной функции может служить ло-

гистическая функция (logistic function), задаваемая следующим выражением:

$$\varphi(\mathcal{G}) = \frac{1}{1 + \exp(-a \cdot \mathcal{G})} \quad (3.10)$$

где a – параметр наклона (slope parameter) сигмоидальной функции. Изменяя этот параметр, можно построить функции с различной крутизной (см. рис. 3.4, в). Первый график соответствует величине параметра, равной $a/4$. В пределе, когда параметр наклона достигает бесконечности, сигмоидальная функция вырождается в пороговую. Если пороговая функция может принимать только значения 0 и 1, то сигмоидальная функция принимает бесконечное множество значений в диапазоне от 0 до 1. При этом следует заметить, что сигмоидальная функция является дифференцируемой (дифференцируемость активационной функции играет важную роль в теории нейронных сетей), в то время как пороговая – нет.

Область значений функций активации, определенных формулами (3.6), (3.9) и (3.10), представляет собой отрезок от 0 до +1. Однако иногда требуется функция активации, имеющая область значений от -1 до $+1$. В этом случае функция активации должна быть симметричной относительно начала координат. Это значит, что функция активации является нечетной функцией индуцированного локального поля. В частности, пороговую функцию в данном случае можно определить следующим образом:

$$\varphi(\mathcal{G}) = \begin{cases} 1, & \text{если } \mathcal{G} \geq +\frac{1}{2}; \\ 0, & \text{если } \mathcal{G} = 0; \\ -1, & \mathcal{G} < -0, \end{cases} \quad (3.11)$$

Эта функция обычно называется сигнум. В данном случае сигмоидальная функция будет иметь форму гиперболического тангенса:

$$\varphi(\mathcal{G}) = \tanh(\mathcal{G}) \quad (3.12)$$

Такой вид сигмоидальной функции обеспечивает ряд преимуществ. Ценные свойства сигмоидальной функции – дифференцируемость на всей оси абсцисс и простое выражение для ее производной, что используется в некоторых алгоритмах обучения. Кроме того, она обладает свойством усиливать слабые сигналы лучше, чем большие, что предотвращает насыщение от больших сигналов, так как они соответствуют областям аргументов, где сигмоид имеет пологий наклон. Известно также большое количество других разновидностей активационных функций.

Модель нейрона, показанная на рис. 3.3, является *детерминистской*. Это значит, что преобразование входного сигнала в выходной точно определено для всех значений входного сигнала. Однако в некоторых приложениях лучше использовать *стохастические нейросетевые модели*, в которых функция активации имеет вероятностную интерпретацию. В подобных моделях нейрон может находиться в одном из двух состояний: +1 или -1. Решение о переключении состояния нейрона принимается с учетом вероятности этого события. Обозначим состояние нейрона символом x , а *вероятность активации нейрона* (probability of firing) – функцией $P(\vartheta)$, где ϑ – индуцированное локальное поле нейрона. Тогда

$$x = \begin{cases} +1, & \text{с вероятностью } P(\vartheta); \\ -1, & \text{с вероятностью } 1 - P(\vartheta). \end{cases} \quad (3.13)$$

Вероятность описывается сигмоидальной функцией следующего вида:

$$P(\vartheta) = \frac{1}{1 + \exp\left(-\frac{\vartheta}{T}\right)}, \quad (3.14)$$

где T – это аналог температуры (temperature), используемый для управления уровнем шума, и, таким образом, степенью неопределенности переключения. При этом важно заметить, что T не описывает физическую температуру нейронной сети, будь то биологической или искусственной. Параметр T управляет термальными флуктуациями, представляющими эффект синаптического шума. Заметим, что если параметр T стремится к нулю, то стохастический нейрон, описанный выражением (3.14), принимает детерминированную форму (без включения шума) нейрона Мак-Каллока – Питца.

§ 1.2. Нейронные сети

НС – это громадный распределённый параллельный процессор, состоящий из элементарных единиц обработки информации, накапливающих экспериментальные знания и представляющих их для последующей обработки. Структура нейронных сетей тесно связана с используемыми алгоритмами обучения. В общем случае можно выделить три фундаментальных класса нейросетевых архитектур.

Однослойные сети прямого распространения

В многослойной нейронной сети нейроны располагаются по слоям. В простейшем случае в такой сети существует входной слой (input layer) узлов источника, информация от которого передается на выходной слой

(output layer) нейронов (вычислительные узлы), но не наоборот. Такая сеть называется сетью прямого распространения (feed-forward) или ациклической сетью (acyclic). На рис. 3.5 показана структура такой сети для случая четырех узлов в каждом из слоев (входном и выходном). Такая нейронная сеть называется однослойной (single-layer network), при этом под единственным слоем подразумевается слой вычислительных элементов (нейронов). При подсчете числа слоев мы не принимаем во внимание узлы источника, так как они не выполняют никаких вычислений.

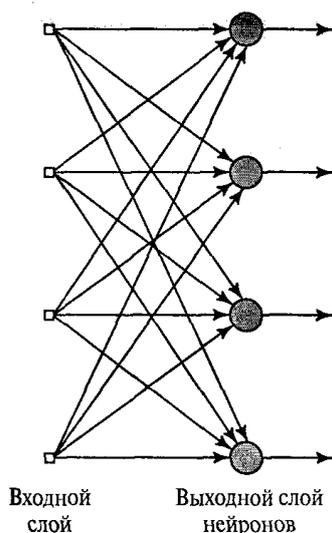


Рис. 3.5. Сеть прямого распространения с одним слоем

Многослойные сети прямого распространения

Другой класс нейронных сетей прямого распространения характеризуется наличием одного или нескольких скрытых слоев (hidden layer), узлы которых называются скрытыми нейронами (hidden neuron), или скрытыми элементами (hidden unit). Функция последних заключается в посредничестве между внешним входным сигналом и выходом нейронной сети. Добавляя один или несколько скрытых слоев, мы можем выделить статистики высокого порядка. Такая сеть позволяет выделять глобальные свойства данных с помощью локальных соединений за счет наличия дополнительных синаптических связей и повышения уровня взаимодействия нейронов [1].

Способность скрытых нейронов выделять статистические зависимости высокого порядка, особенно существенна, когда размер входного слоя достаточно велик. Узлы источника входного слоя сети формируют соответствующие элементы шаблона активации (входной вектор), которые составляют входной сигнал, поступающий на нейроны (вычислительные элементы) второго слоя (т. е. первого скрытого слоя).

Выходные сигналы второго слоя используются в качестве входных для третьего слоя и т. д. Обычно нейроны каждого из слоев сети используют в качестве входных сигналов выходные сигналы нейронов только предыдущего слоя. Набор выходных сигналов нейронов выходного (последнего) слоя сети определяет общий отклик сети на данный входной образ, сформированный узлами источника входного (первого) слоя.

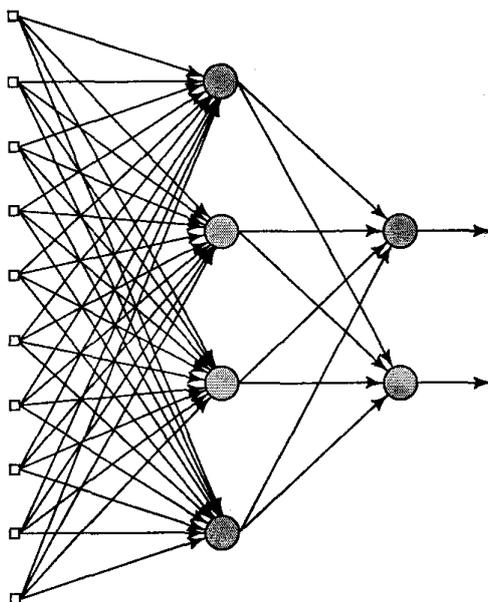


Рис. 3.6. Полносвязная сеть прямого распространения с одним скрытым и одним выходным слоем нейрона

Сеть, показанная на рис. 3.6, называется сетью 10–4–2, так как она имеет 10 входных, 4 скрытых и 2 выходных нейрона. В общем случае сеть прямого распространения с m входами, h_1 нейронами первого скрытого слоя, h_2 нейронами второго скрытого слоя и q нейронами выходного слоя называется сетью $m - h_1 - h_2 - q$.

Нейронная сеть, показанная на рис. 3.6, считается полносвязной (fully connected), в том смысле, что все узлы каждого конкретного слоя соединены со всеми узлами смежных слоев. Если некоторые из синаптических связей отсутствуют, такая сеть называется неполносвязной (partially connected).

Рекуррентные сети

Рекуррентная нейронная сеть (recurrent network) отличается от сети прямого распространения наличием, по крайней мере, одной обратной связи (feedback loop). Например, рекуррентная сеть [43] может состоять из единственного слоя нейронов, каждый из которых направляет свой выходной сигнал на входы всех остальных нейронов слоя.

Архитектура такой нейронной сети показана на рис. 3.7. В приведенной структуре отсутствуют обратные связи нейронов с самими собой. Рекуррентная сеть, показанная на рис. 3.7, не имеет скрытых нейронов. На рис. 3.8 показан другой класс рекуррентных сетей – со скрытыми нейронами. Здесь обратные связи исходят как из скрытых, так и из выходных нейронов.

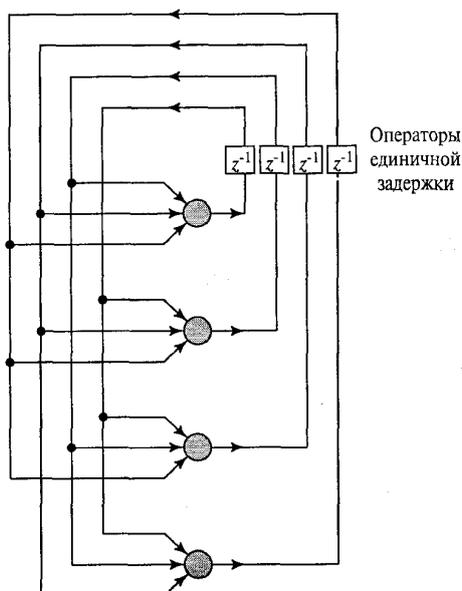


Рис. 3.7. Рекуррентная сеть без скрытых нейронов и обратных связей нейронов с самими собой

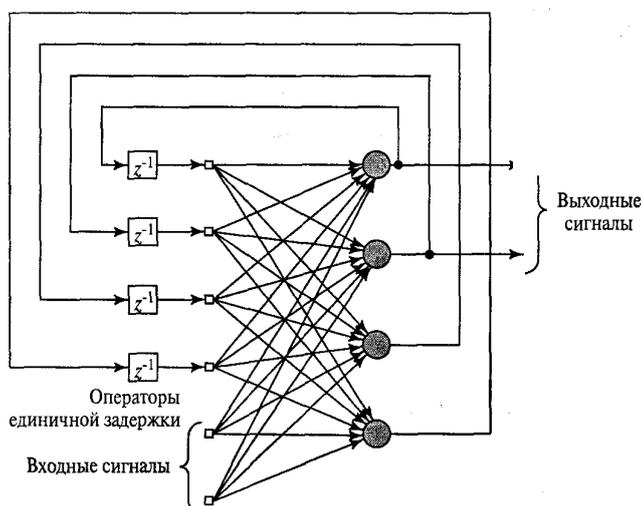


Рис. 3.8. Рекуррентная сеть со скрытыми нейронами

Наличие обратных связей в сетях, показанных на рис. 3.7 и 3.8, оказывает непосредственное влияние на способность таких сетей к обучению и на их производительность. Более того, обратная связь подразумевает ис-

пользование элементов единичной задержки (unit-delay element) (они обозначены как z^{-1}), что приводит к нелинейному динамическому поведению, если, конечно, в сети содержатся нелинейные нейроны.

К настоящему времени разработано большое количество различных типов НС, имеющих свои отличительные особенности. Наиболее широкое распространение получили многослойные НС прямого распространения, или многослойные персептроны, сокращенно – MLP (от Multi Layer Perceptron) [26, 31, 34].

В качестве примера рассмотрим простейшую двухслойную НС (рис. 3.9).

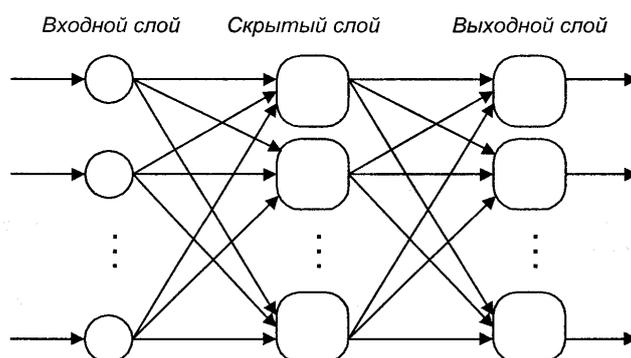


Рис. 3.9. Двухслойная нейронная сеть

Отметим, что нейроны первого слева (входного) слоя сети на рис. 3.9 математических операций не выполняют, а служат лишь для размножения сигналов и при определении числа слоев не учитываются. Чтобы НС могла решить поставленную задачу, её предварительно необходимо обучить. Сущность обучения состоит в подстройке весов нейронов по примерам обучающей выборки [14, 15, 21, 24, 32, 33, 40, 41, 44]. Эффективность использования НС устанавливается рядом так называемых теорем о полноте [2, 5, 6, 31]. Смысл данных теорем сводится к тому, что любая непрерывная функция на замкнутом ограниченном множестве может быть равномерно приближена функциями, вычисленными нейронными сетями, при выполнении некоторых достаточно легко реализуемых условий; таким образом, НС являются универсальными аппроксиматорами.

Основным алгоритмом обучения является алгоритм обратного распространения. В данном случае обучается воспроизводить зависимость, заданную набором из N пар точек $\langle \vec{x}^i, y^i \rangle, i = 1, 2, \dots, N$, с минимизацией суммарной квадратичной ошибки

$$E = \sum_{j=1}^n E_j, \quad (3.15)$$

$E_i = (y^i - o_i)^2$, o_i – выход сети при поступлении на вход \vec{x}^i

Алгоритм состоит в последовательном выполнении следующих шагов:

1. Задают параметр $\eta \in [0, 1]$ и некоторые малые случайные веса сети: $w_j^{(l, k)}$ – j -й вес k -го нейрона в l -м слое НС, а также E_{\max} – максимальное значение суммарной функции ошибок сети.
2. Устанавливается $i = 1$ – номер текущей обучающей точки и $E = 0$ – текущее значение суммарной функции ошибок сети.
3. Вводится обучающая пара $\vec{x} = \vec{x}^i$ и $\vec{y} = \vec{y}^i$. Вычисляется выходной сигнал сети o_i .
4. Производится корректировка весов по формуле:

$$W^{(i)} = W^{(i)} + \eta \frac{\partial E_i}{\partial W^{(i)}}, \quad (3.16)$$

$W^{(i)}$ – матрица весов l -го слоя нейронов, причем коррекция весов происходит в направлении от последнего слоя к первому, т. е. l последовательно меняется от M (число слоев в сети) до 1.

5. Корректируются (наращиваются) значения функции ошибки:

$$E = E + \frac{1}{2} (y^i - o_i)^2 \quad (3.17)$$

6. Если $i < N$, то $i = i + 1$ и переход к п. 3.
7. Если $E > E_{\max}$, то переход к п. 2.
8. Конец.

В формуле (3.19) под $\frac{\partial E_i}{\partial W^{(i)}}$ понимается матрица, элементами которой являются частные производные $\frac{\partial E_i}{\partial W^{(i, k)}}$. Элементы данной матрицы в ряде случаев могут быть определены в аналитическом виде. Например, если нейроны сети, описываются сигмоидальными функциями вида

$$o_i = \frac{1}{1 + \exp(-\vec{w}^T \cdot \vec{x})} \quad (3.18)$$

o – выходной сигнал нейрона; w – вектор весов нейрона; x – вектор входных сигналов нейрона, то для выходного нейрона сети

$$\frac{\partial E_i}{\partial W^{(M)}} = -(y^i - o_i) o_i (1 - o_i) \vec{o}_i^{(M-1)} \quad (3.19)$$

o_i – выход сети; $\vec{o}_i^{(M-1)}$ – выходной сигнал $(M - 1)$ слоя.

Для предыдущего слоя НС:

$$\frac{\partial E_i}{\partial W^{(M)}} = -(y^i - o_i) o_i (1 - o_i) W^{(M-1)} o_i^{(M-1,k)} (1 - o_i^{(M-1,k)}) \bar{o}_i^{(M-2)} \quad (3.20)$$

$\bar{o}_i^{(M-1,k)}$ – выходной сигнал k -го нейрона $M - 1$ слоя; $\bar{o}_i^{(M-2)}$ – выходной сигнал $(M - 2)$ слоя.

Аналогично для всех остальных слоев сети.

Если вид частных производных аналитически определить не удастся, то пользуются приближенными формулами для численной оценки производных:

$$\frac{\partial E_i}{\partial W^{(i)}} = \frac{\Delta E_i}{\Delta W^{(i)}} \quad (3.21)$$

Рассмотренный алгоритм реализует процедуру градиентного метода наискорейшего спуска. Данный метод имеет линейную скорость сходимости, а также резкое замедление оптимизационного процесса в окрестности точки оптимального решения, что делает рассматриваемый алгоритм на практике малоэффективным. Тем не менее, благодаря своей простоте, он остается одним из наиболее распространенных.

Более эффективными методами настройки весов многослойной сети являются квазиньютоновские алгоритмы, например методы переменной метрики (Бройдена-Флетчера-Гольдфарба-Шенно, Девидона-Флетчера-Пауэлла) или Левенберга-Марквардта. При большом числе настраиваемых весов (десятки тысяч и более) очень хорошо зарекомендовал себя алгоритм сопряженных градиентов, хотя при меньшей размерности он уступает квазиньютоновским алгоритмам. Разработано также большое число алгоритмов эвристического типа, не имеющих строгого теоретического обоснования, но показавших свою эффективность на практике; в качестве примеров таких алгоритмов можно привести: симплекс-метод, Quickprop, RPROP и ряд других.

Как известно, перечисленные выше алгоритмы является алгоритмами локальной оптимизации и для увеличения вероятности нахождения глобального экстремума необходимо проводить обучение несколько раз с разными начальными весами нейронов. Для надежного нахождения глобального решения разработан ряд алгоритмов глобальной оптимизации, наиболее известными из которых являются метод имитации отжига и генетические алгоритмы.

Важным вопросом при применении многослойных персептронов является выбор структуры сети – числа слоев сети и числа нейронов в каждом слое. Сеть, содержащая недостаточное количество нейронов,

может оказаться недостаточно гибкой для аппроксимации данных с требуемой точностью. В то же время сеть слишком большого объема будет точно следовать данным из обучающей выборки, но плохо обобщать данные, не входящие в обучающую выборку, такое явление называется переобучением. Для решения данной задачи разработаны многочисленные эмпирические формулы, а также алгоритмы адаптации, настраивающие в процессе обучения не только значения весов нейронов, но и структуру сети. Наиболее простым эмпирическим правилом является следующее: число скрытых слоев многослойного персептрона выбирается $1 \div 2$, число нейронов в слое – от n до $3n$, где n – число входов сети.

Кроме многослойных персептронов, находят применение радиальные НС, или RBF-сети (от Radial Basis Function Network). Сеть RBF – это двух-слойная НС. Первый слой данной сети состоит из так называемых радиальных нейронов. Нейроны данного слоя описываются соотношением

$$o_r = \varphi\left(\frac{\|\vec{x} - \vec{c}_r\|}{\lambda_r}\right), \quad (3.22)$$

o_r – выходной сигнал r -го нейрона; \vec{x} – входной сигнал сети (данный сигнал подается на каждый нейрон рассматриваемого слоя); c_r, λ_r – постоянные параметры, которые могут настраиваться в процессе обучения.

Часто в качестве $\varphi(\cdot)$ используется функция Гаусса:

$$\varphi(s) = \exp\left(-\frac{s^2}{2}\right). \quad (3.23)$$

Второй слой RBF-сети осуществляет линейное или нелинейное преобразование выходных сигналов первого слоя (в частности, если выходной сигнал сети скаляр, данный слой состоит из одного нейрона).

В обобщенно-регрессионных нейронных сетях, или GRNN (от Generalized Regression Neural Network), являющихся разновидностью RBF-сетей, второй слой осуществляет взвешенное суммирование выходных сигналов первого слоя:

$$y = \frac{\sum_{r=1}^m w_r \cdot o_r}{\sum_{r=1}^m o_r}, \quad (3.24)$$

где m – число нейронов первого слоя, w_r – веса, настраиваемые при обучении.

Допустим, что обучающая выборка состоит из N пар значений $\langle \vec{x}^i, y^i \rangle, i = 1, 2, \dots, N$. В простейшем случае при обучении формирует-

ся радиальный слой из $m = N$ нейронов с параметрами $\vec{c}_r = \vec{x}^i$, а параметры второго слоя выбираются из условия $w_r = y^i$. В описанном алгоритме обучение GRNN происходит практически мгновенно. Однако сеть получается достаточно громоздкой, поэтому разработаны алгоритмы обучения GRNN, в которых число радиальных нейронов меньше числа элементов обучающей выборки ($m < N$).

В общем случае процесс обучения сети RBF сводится к определению ряда параметров:

- 1) числа нейронов радиального слоя m ;
- 2) координат центров \vec{c}_r и отклонений λ_r радиальных базисных функций (см. формулу (3.22));
- 3) значений весов нейронов выходного слоя w_r (см. формулу (3.24)).

Настройка параметров сети \vec{c}_r , λ_r , w_r может осуществляться методом обратного распространения ошибки, аналогично тому, как это происходит в сигмоидальных НС, например, градиентным методом.

Однако особенности сетей RBF позволяют выработать более эффективные алгоритмы настройки данных параметров.

Веса нейронов выходного слоя w_r обычно входят линейно в выражение для выходного сигнала сети (см. формулу (2.12)), и их настройку можно осуществить с помощью формул для определения коэффициентов линейной регрессии по методу наименьших квадратов:

$$\vec{w} = (O^T \cdot O) \cdot O^T \cdot \vec{y}, \quad (3.25)$$

где $\vec{w} = [w_1, w_2, \dots, w_m]^T$ – вектор весов выходного слоя; $\vec{o} = [o_r^1, o_r^2, \dots, o_r^n]^T$ – вектор выходов r -го радиального нейрона в обучающих точках; $O = [\vec{o}_1^T, \vec{o}_2^T, \dots, \vec{o}_m^T]^T$ – матрица выходов радиальных нейронов в обучающих точках; $\vec{y} = [y_r^1, y_r^2, \dots, y_r^N]^T$ – вектор обучающих значений.

Для определения координат центров \vec{c}_r могут использоваться, следующие методы:

- а) разместить центры радиальных функций \vec{c}_r в узлах равномерной сетки или в случайных точках;
- б) поместить центры радиальных функций \vec{c}_r в обучающие точки \vec{x}^i ;
- в) поместить центры радиальных функций \vec{c}_r в центры кластеров обучающих данных.

Для определения отклонений λ_r также существуют различные эмпирические методы, использующие в качестве исходной информации расстояния между центрами радиальных функций. Например, можно выбрать все отклонения одинаковыми: $\lambda_r = \lambda = \rho$ – среднее расстояние

между центрами базисных функций \vec{c}_r . Или выбрать в качестве λ_r среднее расстояние между r -м центром базисной функции \vec{c}_r и K_c ближайшими к нему центрами.

Наиболее сложным является выбор числа радиальных базисных функций m . Существует несколько методов и алгоритмов выбора, однако данную задачу пока нельзя считать решенной. Для радиальных сетей доказан ряд теорем, согласно которым данные сети при выполнении определенных условий могут аппроксимировать произвольную гладкую функцию.

Обычно RBF-сети обладают худшими обобщающими свойствами по сравнению с сигмоидальными НС при больших объемах обучающей выборки. Особенно сильно это проявляется в задачах экстраполяции. Однако сети RBF показали очень хорошие результаты в случае, если объем обучающей выборки мал. В сетях RBF нет проблемы выбора числа слоев сети. Кроме того, большинство алгоритмов обучения RBF-сетей работают значительно быстрее по сравнению с алгоритмами обучения многослойных персептронов.

Рассмотрим реализацию динамических операторов нейронными сетями. Приведенные выше НС могут аппроксимировать произвольную гладкую функцию. На рис. 3.10 показана одна из возможных структур реализации произвольного дискретного динамического оператора с помощью НС.

Под z^{-k} на рис. 3.10 понимается элемент задержки на k шагов. Дискретный динамический оператор, реализуемый рассматриваемой структурой, имеет вид:

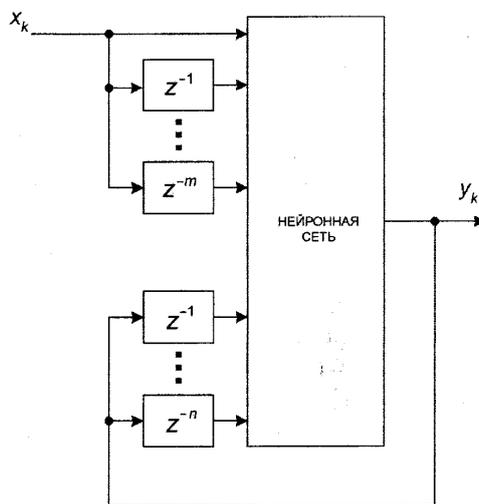


Рис. 3.10. Реализация дискретного динамического оператора с помощью многослойного персептрона

$$y_k = \varphi(x_k, x_{k-1}, \dots, x_{k-m}, y_{k-1}, y_{k-2}, \dots, y_{k-n}), \quad (3.26)$$

где $\varphi(\cdot)$ – функция, реализуемая НС, $k = 0, 1, 2, \dots$

Модель (3.26) в литературе обычно называется авторегрессионной. Параметр n , характеризующий порядок модели, обычно называют глубиной регрессии или глубиной погружения. Параметр m выбирают из условия $m \leq n$. Отсчеты x_k и y_k при отрицательных k описывают начальные условия модели.

§ 2. Преимущества нейронных сетей

Совершенно очевидно, что свою силу нейронные сети черпают, во-первых, из распараллеливания обработки информации и, во-вторых, из способности самообучаться, т. е. создавать обобщения. Под термином обобщение (generalization) понимается способность получать обоснованный результат на основании данных, которые не встречались в процессе обучения. Эти свойства позволяют нейронным сетям решать сложные (масштабные) задачи, которые на сегодняшний день считаются трудноразрешимыми. Однако на практике при автономной работе нейронные сети не могут обеспечить готовые решения. Их необходимо интегрировать в сложные системы. В частности, комплексную задачу можно разбить на последовательность относительно простых, часть из которых может решаться нейронными сетями. Очень важно уяснить, что для создания компьютерной архитектуры, которая будет способна имитировать человеческий мозг (если такое окажется возможным вообще), придется пройти долгий и трудный путь.

Использование нейронных сетей обеспечивает следующие полезные свойства систем [43].

1. *Нелинейность* (nonlinearity). Искусственные нейроны могут быть линейными и нелинейными. Нейронные сети, построенные из соединений нелинейных нейронов, сами являются нелинейными. Более того, эта нелинейность особого сорта, так как она распределена (distributed) по сети. Нелинейность является чрезвычайно важным свойством, особенно если сам физический механизм, отвечающий за формирование входного сигнала, тоже является нелинейным (например, человеческая речь).

2. *Отображение входной информации в выходную* (input-output mapping). Одной из популярных парадигм обучения является *обучение с учителем* (supervised learning). Это подразумевает изменение синаптических весов на основе набора маркированных учебных примеров (training sample). Каждый пример состоит из входного сигнала и соответствующего ему желаемого отклика (desired response). Из этого множества случайным образом выбирается пример, а нейронная сеть модифицирует синаптические веса для минимизации расхождений желаемого выходного сигнала и формируемого сетью согласно выбранному статистическому критерию. При этом собственно модифицируются свободные параметры (free parameters) сети. Ранее использованные примеры могут впослед-

ствии быть применены снова, но уже в другом порядке. Это обучение проводится до тех пор, пока изменения синаптических весов не станут незначительными. Таким образом, нейронная сеть обучается на примерах, составляя таблицу соответствий вход-выход для конкретной задачи. Такой подход заставляет вспомнить непараметрическое статистическое обучение (nonparametric statistical inference). Это направление статистики имеет дело с оценками, не связанными с какой-либо конкретной моделью, или, с биологической точки зрения, с обучением с нуля. Здесь термин «непараметрический» используется для акцентирования того, что изначально не существует никакой предопределенной статистической модели входных данных. Для примера рассмотрим задачу классификации образов (pattern classification). В ней требуется соотнести входной сигнал, представляющий физический объект, или событие, с некоторой предопределенной категорией (классом). При непараметрическом подходе к этой задаче требуется «оценить» рамки решения в пространстве входного сигнала на основе набора примеров. При этом не используется никакая вероятностная модель распределения. Аналогичный подход применяется и в парадигме обучения с учителем. Это еще раз подчеркивает параллель между отображением входных сигналов в выходные, осуществляемым нейронной сетью, и непараметрическим статистическим обучением.

3. *Адаптивность* (adaptivity). Нейронные сети обладают способностью адаптировать свои синаптические веса к изменениям окружающей среды. В частности, нейронные сети, обученные действовать в определенной среде, могут быть легко переучены для работы в условиях незначительных колебаний параметров среды. Более того, для работы в нестационарной (nonstationary) среде (где статистика изменяется с течением времени) могут быть созданы нейронные сети, изменяющие синаптические веса в реальном времени. Естественная для классификации образов, обработки сигналов и задач управления архитектура нейронных сетей может быть объединена с их способностью к адаптации, что приведет к созданию моделей адаптивной классификации образов, адаптивной обработки сигналов и адаптивного управления. Известно, что чем выше адаптивные способности системы, тем более устойчивой будет ее работа в нестационарной среде. При этом хотелось бы заметить, что адаптивность не всегда ведет к устойчивости; иногда она приводит к совершенно противоположному результату. Например, адаптивная система с параметрами, быстро изменяющимися во времени, может также быстро реагировать и на посторонние возбуждения, что вызовет потерю производительности. Для того чтобы использовать все достоинства адаптивности, основные параметры системы должны быть достаточно стабильными, чтобы можно было не учитывать внешние помехи, и достаточно гибкими, чтобы обеспечить реакцию на

существенные изменения среды. Эта задача обычно называется дилеммой стабильности-пластичности (stability-plasticity dilemma).

4. *Очевидность ответа* (evidential response). В контексте задачи классификации образов можно разработать нейронную сеть, собирающую информацию не только для определения конкретного класса, но и для увеличения достоверности (confidence) принимаемого решения. Впоследствии эта информация может использоваться для исключения сомнительных решений, что повысит продуктивность нейронной сети.

5. *Контекстная информация* (contextual information). Знания представляются в самой структуре нейронной сети с помощью ее состояния активации. Каждый нейрон сети потенциально может быть подвержен влиянию всех остальных ее нейронов. Как следствие, существование нейронной сети непосредственно связано с контекстной информацией.

6. *Отказоустойчивость* (fault tolerance). Нейронные сети, облаченные в форму электроники, потенциально отказоустойчивы. Это значит, что при неблагоприятных условиях их производительность падает незначительно. Например, если поврежден какой-то нейрон или его связи, извлечение запомненной информации затрудняется. Однако, принимая в расчет распределенный характер хранения информации в нейронной сети, можно утверждать, что только серьезные повреждения структуры нейронной сети существенно повлияют на ее работоспособность. Поэтому снижение качества работы нейронной сети происходит медленно. Незначительное повреждение структуры никогда не вызывает катастрофических последствий. Это очевидное преимущество робастных вычислений, однако, его часто не принимают в расчет. Чтобы гарантировать отказоустойчивость работы нейронной сети, в алгоритмы обучения нужно закладывать соответствующие поправки [7].

7. *Масштабируемость* (VLSI Implementability). Параллельная структура нейронных сетей потенциально ускоряет решение некоторых задач и обеспечивает масштабируемость нейронных сетей в рамках технологии VLSI (very large scale integrated). Одним из преимуществ технологий VLSI является возможность представить достаточно сложное поведение с помощью иерархической структуры.

8. *Единообразие анализа и проектирования* (Uniformity of analysis and design). Нейронные сети являются универсальным механизмом обработки информации. Это означает, что одно и то же проектное решение нейронной сети может использоваться во многих предметных областях. Это свойство проявляется несколькими способами.

- Нейроны в той или иной форме являются стандартными составными частями любой нейронной сети.
- Эта общность позволяет использовать одни и те же теории и алгоритмы обучения в различных нейросетевых приложениях.

- Модульные сети могут быть построены на основе интеграции целых модулей.

9. *Аналогия с нейробиологией* (Neurobiological analogy). Строение нейронных сетей определяется аналогией с человеческим мозгом, который является живым доказательством того, что отказоустойчивые параллельные вычисления не только физически реализуемы, но и являются быстрым и мощным инструментом решения задач. Нейробиологи рассматривают искусственные нейронные сети как средство моделирования физических явлений. С другой стороны, инженеры постоянно пытаются почерпнуть у нейробиологов новые идеи, выходящие за рамки традиционных электросхем. Эти две точки зрения можно продемонстрировать на следующих примерах. В работе модели линейных систем вестибуло-окулярного рефлекса сравнивались с моделями *рекуррентных нейронных сетей*.

Вестибуло-окулярный рефлекс, или рефлекс VOR (vestibulo-ocular reflex), является составной частью глазодвигательной системы. Его задачей является обеспечение стабильности визуального образа при поворотах головы за счет вращения глаз. Процесс VOR реализуется премоторными нейронами в вестибулярном центре, которые получают и обрабатывают сигналы поворота головы от вестибулярных сенсорных нейронов и передают результат на моторные нейроны мышц глаз. Механизм VOR хорошо подходит для моделирования, так как входной (поворот головы) и выходной (поворот глаз) сигналы можно точно описать. К тому же это довольно простой рефлекс, а нейрофизические свойства реализующих его нейронов довольно хорошо описаны. Среди трех задействованных в нем типов нейронов премоторные нейроны, входящие в состав вестибулярного центра, являются самыми сложными, а значит, самыми интересными. Ранее механизм VOR моделировался с помощью сосредоточенной линейной системы и теории управления. Эти модели были пригодны для описания некоторых общих свойств VOR, но не давали четкого представления о свойствах самих нейронов. С появлением нейросетевых моделей ситуация в корне изменилась. Рекуррентные модели VOR позволили воспроизвести и описать многие статические, динамические, нелинейные и распределенные аспекты обработки сигналов при реализации рефлекса VOR и, в частности, вестибулярный центр.

Сетчатка (retina), более чем какая-то другая часть мозга, выполняет функции взаимосвязи окружающего мира, представленного визуальным рядом или физическим изображением (physical image), проецируемым на матрицу рецепторов, с первым нейронным изображением (neural image). Сетчатка – это матрица микроскопических рецепторов на внешней лицевой стороне глазного яблока. В ее задачи входит преобразование оптического сигнала в нейронное изображение, передаваемое по оптическим нервам в различные центры для анализа. Принимая во внимание синаптическую ор-

ганизацию сетчатки, задача сложная. В любой сетчатке преобразование изображения из оптического в нейронное проходит три стадии:

- Снятие фотокопии слоем нейронов-рецепторов.
- Передача сформированного сигнала (реакция на свет) химическими синапсами на слой биполярных клеток (bipolar cell).
- Передача этих сигналов (также с помощью химических синапсов) на выходные нейроны.

На двух последних стадиях (при передаче информации от рецепторов на биполярные рецепторы и от последних – на выходные нейроны) в операции участвуют специальные нейроны с латеральным торможением, в том числе так называемые горизонтальные клетки (horizontal cell). Их задачей является преобразование сигнала между разными синаптическими слоями. Также существуют центробежные элементы, обеспечивающие передачу сигнала с внутреннего синаптического слоя на внешний. Некоторые исследователи создавали электронные микросхемы, имитирующие структуру сетчатки. Эти электронные чипы назывались нейроморфными контурами (neuromorphic integrated circuit). Нейроморфные сенсоры представляют собой матрицу фоторецепторов, связанных с соответствующими элементами рисунка (пикселями). Они имитируют сетчатку в том смысле, что могут адаптироваться к изменению освещенности, идентифицировать контуры и движение. Нейробиологическая модель, воплощенная в нейроморфных контурах, обеспечила еще одно преимущество: она вселила надежду на то, что физическое понимание нейробиологических структур может оказать существенное влияние на область электроники и технологию VLSI.

§ 3. Искусственный интеллект и нейронные сети

Основной задачей искусственного интеллекта (artificial intelligence – AI) является разработка парадигм или алгоритмов, обеспечивающих компьютерное решение когнитивных задач, свойственных человеческому мозгу. Следует заметить, что это определение искусственного интеллекта не является единственно возможным.

Системы искусственного интеллекта должны обеспечивать решение следующих трех задач: накопление знаний, применение накопленных знаний для решения проблемы и извлечение знаний из опыта. Системы искусственного интеллекта реализуют три ключевые функции: представление, рассуждение и обучение [43] (рис. 3.11).

1. *Представление* (representation). Одной из отличительных черт систем искусственного интеллекта является использование символического языка (symbol structure) для представления общих знаний о предметной области и конкретных знаний о способах решения задачи. Символы

обычно формулируются в уже известных терминах. Это делает символическое представление относительно простым и понятным человеку. Более того, понятность символических систем искусственного интеллекта делает их пригодными для человеко-машинного общения.



Рис. 3.11. Три ключевые функции систем искусственного интеллекта

Термин «знания», используемый создателями систем искусственного интеллекта, является всего лишь еще одним названием данных. Знания могут иметь процедурный и декларативный характер. В декларативном (declarative) представлении знания – это статический набор фактов. При этом существует относительно малый объем процедур, используемых для манипуляций этими фактами. Характерной особенностью декларативного представления является то, что в глазах человека оно имеет смысл само по себе, независимо от использования в системах искусственного интеллекта. В процедурном (procedural) представлении знания внедрены в процедуры, функционирующие независимо от смысла самих знаний. В большинстве предметных областей требуются одновременно оба типа представления знаний.

2. *Рассуждения* (reasoning). Под рассуждениями обычно понимается способность решать задачи. Для того чтобы систему можно было назвать разумной, она должна удовлетворять следующим условиям [3].

- Описывать и решать широкий спектр задач.
- Понимать явную (explicit) и неявную (implicit) информацию.
- Иметь механизм управления (control), определяющий операции, выполняемые для решения отдельных задач.

Решение задач можно рассматривать как некоторую задачу поиска (searching problem). В процессе поиска используются правила (rules), данные (data) и управляющие воздействия (control). Правила действуют на области данных, а управляющие воздействия определяются для правил. Для примера рассмотрим известную «задачу коммивояжера». В ней требуется найти кратчайший маршрут из одного города в другой. При этом все города, расположенные по маршруту, необходимо посетить

только один раз. В этой задаче множество данных состоит из всех возможных маршрутов и их стоимостей, представленных в форме взвешенного графа. Правила определяют пути движения из одного города в другой, а модуль управления решает, когда и какие правила применять.

Во многих практических задачах (например, в медицинской диагностике) доступный набор знаний является неполным или неточным. В таких ситуациях используются вероятностные рассуждения (probabilistic reasoning), позволяющие системам искусственного интеллекта работать в условиях неопределенности.

3. *Обучение* (learning). В простейшей модели машинного обучения (рис. 3.12) информацию для обучаемого элемента (learning element) предоставляет сама среда. Обучаемый элемент использует полученную информацию для модернизации базы знаний (knowledge base), знания из которой функциональный элемент (performance element) затем использует для выполнения поставленной задачи. Информация, поступающая из внешней среды, является несовершенной, поэтому обучаемый элемент заранее не знает, как заполнить пробелы или игнорировать несущественные детали. Машина действует наугад, после чего получает сигнал обратной связи (feedback) от функционального элемента. Механизм обратной связи позволяет системе проверять рабочие гипотезы и пересматривать их по мере необходимости.

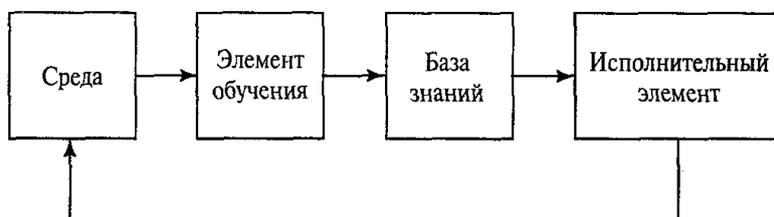


Рис. 3.12. Простейшая модель машинного обучения [43]

Машинное обучение может включать два совершенно разных способа обработки информации: *индуктивный* (inductive) и *дедуктивный* (deductive). При индуктивной обработке информации общие шаблоны и правила создаются на основании практического опыта и потоков данных. При дедуктивной обработке информации для определения конкретных фактов используются общие правила. Обучение на основе подобиия представляет собой индуктивный процесс, а доказательство теорем – дедуктивный, поскольку оно опирается на известные аксиомы и уже доказанные теоремы. В обучении на основе объяснения используется как индукция, так и дедукция.

Возникающие при обучении сложности и накопленный при этом опыт привели к созданию различных методов и алгоритмов пополнения баз зна-

ний. В частности, если в данной предметной области работают опытные профессионалы, проще получить их обобщенный опыт, чем пытаться дублировать экспериментальный путь, который они прошли в процессе его накопления. Эта идея и положена в основу *экспертных систем* (expert system).

Возникает вопрос: как сравнить когнитивные модели нейронных сетей с символьными системами искусственного интеллекта? Для такого сравнения разобьем проблему на три части: уровень объяснения, стиль обработки и структуру представления.

1. *Уровень объяснения* (explanation level). Классические системы искусственного интеллекта (artificial intelligence – AI) основаны на символьном представлении. С точки зрения познания AI предполагает существование ментального представления, в котором познание осуществляется как *последовательная обработка* (sequential processing) символьной информации.

В центре внимания нейронных сетей находятся *модели параллельной распределенной обработки* (parallel distributed processing или PDP). В этих моделях предполагается, что обработка информации происходит за счет взаимодействия большого количества нейронов, каждый из которых передает сигналы возбуждения и торможения другим нейронам сети. Более того, в теории нейронных сетей большое внимание уделяется нейробиологическому описанию процесса познания.

2. *Стиль обработки* (processing style). В классических системах искусственного интеллекта обработка происходит последовательно (sequential), как и в традиционном программировании. Даже если порядок выполнения действий строго неопределен (например, при сканировании правил и фактов в экспертных системах), операции все равно выполняются пошагово. Такая последовательная обработка, скорее всего, объясняется последовательной природой естественных языков и логических заключений, а также структурой машины фон Неймана. Нельзя забывать о том, что классические системы искусственного интеллекта зародились практически в ту же интеллектуальную эру, что и машина фон Неймана.

В отличие от них, концепция обработки информации в нейронных сетях проистекает из принципа параллелизма (parallelism), который является источником их гибкости. Более того, параллелизм может быть массовым (сотни тысяч нейронов), что придает нейронным сетям особую форму робастности. Если вычисления распределены между множеством нейронов, практически не важно, что состояние отдельных нейронов сети отличается от ожидаемого. Зашумленный или неполный входной сигнал все равно можно распознать; поврежденная сеть может продолжать выполнять свои функции на удовлетворительном уровне, а обучение не обязательно должно быть совершенным. Производитель-

ность сети в пределах некоторого диапазона снижается достаточно медленно. Кроме того, можно дополнительно повысить работоспособность сети, представляя каждое свойство группой нейронов.

3. *Структура представления* (representational structure). В классических системах искусственного интеллекта в качестве модели выступает язык мышления, поэтому символьное представление имеет квазилингвистическую структуру. Подобно фразам обычного языка, выражения классических систем искусственного интеллекта, как правило, сложны и составляются путем систематизации простых символов. Учитывая ограниченное количество символов, новые смысловые выражения строятся на основе композиции символьных выражений и аналогии между синтаксической структурой и семантикой.

С другой стороны, в нейронных сетях природа и структура представления являются ключевыми проблемами. В марте 1988 года в специальном выпуске журнала *Cognition* [4] были опубликованы критические замечания по поводу вычислительной адекватности нейронных сетей при решении когнитивных и лингвистических задач. Они аргументированы тем, что нейронные сети не удовлетворяют двум основным критериям процесса познания: природе мысленного представления (mental representation) и мыслительных процессов (mental process). В соответствии с этой работой следующие свойства присущи именно системам искусственного интеллекта и не присущи нейронным сетям:

- Мысленное представление характеризуется комбинаторной избирательной структурой и комбинаторной семантикой.
- Мыслительные процессы характеризуются чувствительностью к комбинаторной структуре представления, с которым они работают.

Таким образом, символьные модели искусственного интеллекта – это формальные системы, основанные на использовании языка алгоритмов и представлении данных по принципу «сверху вниз» (top down), а нейронные сети – это параллельные распределенные процессоры, обладающие естественной способностью к обучению и работающие по принципу «снизу вверх» (bottom up). Поэтому при решении когнитивных задач целесообразно создавать структурированные модели на основе связей (structured connectionist models) или гибридные системы (hybrid system), объединяющие оба подхода. Это обеспечит сочетание свойств адаптивности, робастности и единообразия, присущих нейронным сетям, с представлениями, умозаключениями и универсальностью систем искусственного интеллекта. Для реализации этого подхода были разработаны методы извлечения правил из обученных нейронных сетей. Эти результаты не только позволяют интегрировать нейронные сети с интеллектуальными машинами, но и обеспечивают решение следующих задач.

- Верификация нейросетевых компонентов в программных системах. Для этого внутреннее состояние нейронной сети переводится в форму, понятную пользователям.
- Улучшение обобщающей способности нейронной сети за счет выявления областей входного пространства, не достаточно полно представленных в обучающем множестве, а также определения условий, при которых обобщение невозможно.
- Выявление скрытых зависимостей на множестве входных данных.
- Интеграция символьного и коннекционистского подходов при разработке интеллектуальных машин.
- Обеспечение безопасности систем, для которых она является критичной.

§ 4. Нейросетевое моделирование

Постановка задачи: выявление причинно-следственной связи между факторами, влияющими на функцию отклика.

Методы исследований. Современные инструментальные средства обработки статистической информации для количественных переменных (регрессионный, корреляционный или дискриминантный анализ) дают низкую достоверность обработки псевдоколичественных данных [25], когда возникает трудность установления между ними отношений эквивалентности и предпочтения, что приводит к трудности в определении мер сходства и т. п. Поэтому выберем в качестве методов исследования менее чувствительные к выполнению вероятностных предпосылок нейросетевые методы, а в качестве инструментов исследования возьмём нейропакеты НейроПро и Excel Neural Package. Такой подбор инструментальных средств обеспечит перекрёстную проверку получаемых результатов.

Этапы исследований.

1. Определение признаков, наиболее существенно влияющих на выбранную выходную переменную (отклик).
2. Построение модели, отражающей причинно-следственные связи между откликом и входными признаками.
3. Интерпретация модели.

Полученные результаты.

1. Результаты первого этапа, характеризующие степень влияния факторов на отклик x_{14} (качественные признаки x_4, x_5, x_7 исключены из рассмотрения), представлены на рис. 3.13 (пакет НейроПро) и рис. 3.14 (пакет Excel Neural Package).

Из рис. 3.13 следует, что наиболее значимыми факторами являются $x_2, x_{11}, x_9, x_{10}, x_{13}$, а из рис. 3.14 – x_1, x_{10} (седьмой по счету), x_{11}, x_3, x_6

Обобщая результаты, получим следующую упорядоченную по степени влияния на x_{14} последовательность признаков: $x_1, x_{11}, x_{10}, x_6, x_{13}$

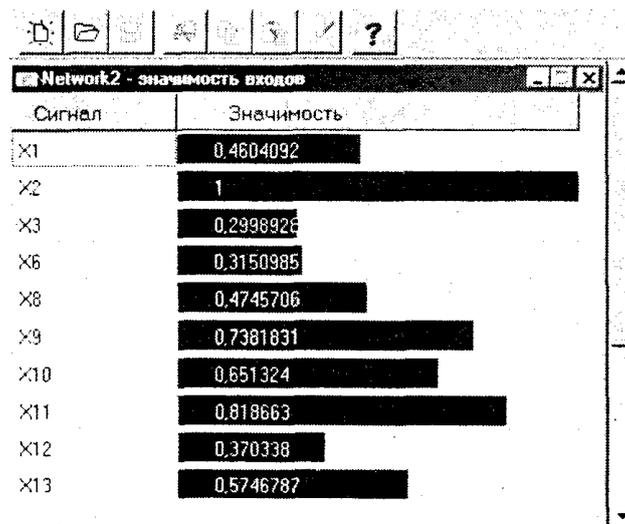


Рис. 3.13. Оценка значимости факторов (пакет НейроПро) [25]

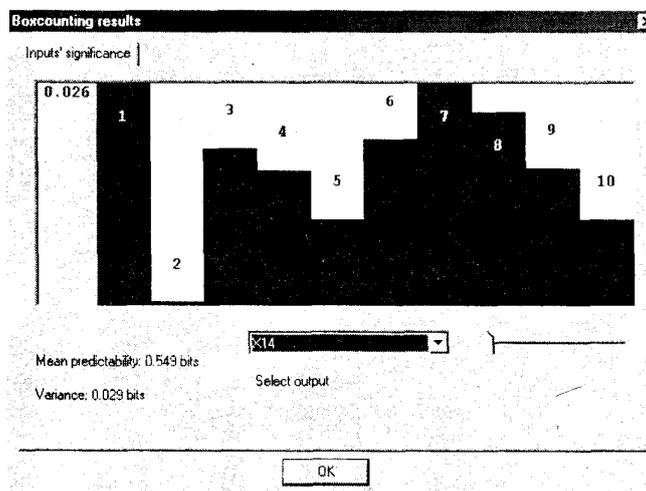


Рис. 3.14. Оценка значимости факторов (пакет Excel Neural Package) [25]

2. Результаты второго этапа исследований, характеризующие проверку возможностей использования различных моделей для описания имеющихся данных с учётом полученной значимости признаков, показали, что в качестве таких моделей не подходят модели количественного характера типа регрессионных или классических нейросетевых.

Наиболее подходящей представляется модель в виде совокупности кластеров. Такие кластеры, соответствующая информация о которых представлена в табл. 3.1, выявлены с помощью пакета Excel Neural Package при использовании самоорганизующейся карты Кохонена.

Таким образом, наиболее подходящей моделью для рассматриваемого типа задач является модель в виде совокупности кластеров. Выявление наиболее влияющих на функцию отклика x_{14} признаков и кластеров может быть использовано для дальнейшего прогнозирования в поведении систем.

Таблица 3.1

Информация о выявленных кластерах [25]

Центр кластера	Кластер				
	1	2	3	4	5
x_1	39,2	39,6	32,3	40,5	66,9
x_2	1,45	1,78	1,56	1,48	1,51
x_3	4,1	4,1	3,3	2,38	2,76
x_6	3,25	2,3	2,5	1,58	1,75
x_8	10,4	11,7	3,58	3,24	9,88
x_9	4,05	3,06	3,54	3,72	1,07
x_{10}	3,84	4,15	4,7	3,15	1,05
x_{11}	1480	492	609	348	480
x_{12}	3,06	1,53	1,99	1,18	1,47
x_{13}	1,1	1,53	1,27	1,91	1,63
x_{14}	2,9	3,29	2,42	4,03	3,91
Количество элементов в кластере	204	377	279	256	384

§ 5. Применение искусственных нейронных сетей

Процесс разработки нейросетевых приложений состоит из определенной последовательности операций, которая не зависит от выбранного пакета программ. Большое значение имеет исходная формулировка проблемы. Самая трудоемкая для аналитика часть работы включает сбор и предварительную обработку данных. По некоторым оценкам это может занимать до 90 % времени.

Многие операции по подготовке данных удобнее выполнять не в NetMaker, а, например, в Excel. Более того, для интенсивного использования нейронной сети требуются специальные программы, которые выбирают из поступающего потока нужные данные, преобразуют их и передают на вход сети. Решение этой проблемы весьма актуально для хранилищ данных.

Нейронные сети позволяют строить комплексную систему прогноза как на базе технических, так и макроэкономических показателей, находя компромисс между сторонниками двух подходов.

Рассмотрим ряд задач с успехов решаемых с помощью искусственных нейронных сетей.

Сжатие информации. Проиллюстрируем применение многослойных персептронов на примере решения задачи сжатия информации для представления исходных данных в компактной и удобной форме. Результаты этапа сжатия зачастую определяют успех решения задачи распознавания в целом.

Популярный метод сжатия информации был предложен в 1987 г. G. Cottrell, P. Munro, D. Zipser. Рассмотрим трехслойный персептрон,

у которого число нейронов входного и выходного слоев одинаково, а число нейронов скрытого слоя значительно меньше. Предположим, что в результате обучения персептрон может воспроизводить на выходе тот же самый вектор X , который подается на входной слой. Такой персептрон автоматически осуществляет сжатие (компрессию) информации: на нейронах скрытого слоя возникает представление каждого вектора, которое значительно короче, чем длина вектора, подаваемого на вход. Предположим, что некоторый набор векторов нужно передавать по линии связи, предварительно сжимая информацию и, тем самым, уменьшая число каналов, необходимых для ее передачи. Поместим на одном конце линии входной и скрытый слои персептрона, а результат работы нейронов скрытого слоя (короткие векторы) будем передавать по линии связи. На другом конце линии поместим копию скрытого слоя и выходной слой, тогда переданный короткий вектор с нейронов скрытого слоя перейдет на нейроны выходного слоя, где и будет воспроизведен входной вектор (декомпрессия).

Обработка видеоизображений. Одной из наиболее сложных и актуальных проблем обработки видеоизображений, представленных последовательностью оцифрованных кадров, является проблема выделения и распознавания движущихся объектов в условиях действия различного рода помех и возмущений. Для этого должны быть решены задачи выделения изображений движущихся объектов на сложном зашумленном фоне, фильтрации помех, скоростной фильтрации, отделения объектов от фона, оценки скорости каждого объекта, его идентификации и сопровождения. Системы обработки видеоизображений, построенные с применением нейросетевых методов, представляют собой, как правило, программно-аппаратные комплексы на персональных компьютерах, позволяющие работать с данными телевизионной системы в реальном времени (не менее 25 кадров/с, 320×200 пикселей).

Выделение изображений движущихся объектов осуществляется путем построения оценки поля скоростей с помощью многослойной локально-связной нейронной сети. Размерность сети для изображения 320×200 пикселей составляет несколько миллионов нейронов и примерно вчетверо больше синапсов.

Распознавание выделенных силуэтов производится на самоорганизующейся нейронной сети, предварительно обученной на изображениях объектов рассматриваемых классов. Обеспечивается инвариантность к произвольному движению фона, устойчивость к зашумлению до 10 %. Вероятность правильного распознавания не менее 90 %.

Обработка статических изображений. Не менее сложными являются задачи выделения и распознавания объектов на статическом тоновом изображении. В частности, подобные задачи возникают при автоматиче-

ской обработке спутниковых изображений земной поверхности. Для их решения разработан и реализован на персональных компьютерах ряд автоматизированных систем анализа изображений земной поверхности. Системы в автоматическом режиме обеспечивает выделение обрабатываемых изображениях объектов заданных классов: дорожной сети, кварталов с характерной застройкой, аэродромов стоящих на них самолетов. Нейросетевые принципы, заложенные в их основу, обеспечивают инвариантность к яркостным характеристикам выделяемых и распознаваемых объектов, а также позволяют проводить обучение и адаптацию систем.

Обнаружение и классификация объектов по звуковым и гидроакустическим сигналам. Использование нейросетевых технологий для анализа акустического излучения демонстрируют системы обнаружения и распознавания летательных аппаратов по звуку, а также надводных подводных объектов по гидроакустическим сигналам. Сигналы с объектов подвергаются предобработке и в оцифрованном виде подаются на вход предварительно обученной нейронной сети для распознавания. Исследования систем показали высокую вероятность правильного распознавания (до 90 %).

Задачи комбинаторной оптимизации. Высокая степень параллельности обработки информации позволяет успешно применять нейросетевые технологии для решения задач комбинаторной оптимизации, к которым, в первую очередь следует отметить задачи транспортно-ориентированной оптимизации, например, задача коммивояжера и ее модификации и задачи распределения ресурсов (задача о назначениях, о перераспределениях и другие).

Решение таких задач традиционными методами математического программирования, большинство из которых изначально ориентировано на вычислительную технику с последовательно архитектурой, сопряжено с большими временными затратами, не приемлемыми для многих приложений. При соответствующей аппаратной поддержке нейросетевые методы позволяют значительно повысить оперативность решения данного класса задач, сохраняя высокую точность результата.

Медицинская диагностика. В настоящее время разработано достаточно много нейродиагностических комплексов, позволяющих сократить необходимое время диагностики различных заболеваний, а также снизить потребность в квалифицированном медицинском персонале.

Например, компанией «НейроПроект» создана система диагностики слуха у грудных детей. Общепринятая методика диагностики состоит в том, что в процессе обследования регистрируются отклики мозга в ответ на звуковые раздражители. Для достаточно уверенной диагностики слуха ребенка опытному эксперту-аудиологу необходимо провести около 2000 тестов, что

занимает около часа. Нейронная сеть способна с той же достоверностью определить уровень слуха уже по 200 наблюдениям в течение всего нескольких минут, причем без участия квалифицированного персонала.

Распознавание речи. Распознавание речи – одно из наиболее популярных применений нейронных сетей. Достаточно квалифицированный пользователь может создать свою нейросетевую систему распознавания речи, используя, к примеру, двухкаскадную иерархическую нейронную сеть, где первый уровень осуществляет грубое распознавание слов, относя их к одному из классов, а второй уровень точно классифицирует слово внутри каждого из классов.

Обнаружение фальсификаций. В США введена в действие система обнаружения мошенничеств в области здравоохранения. Подсчитано, что потери бюджета от такого рода фальсификаций составляют около 730 млн долларов в год. Создание нейросетевой специализированной системы заняло у фирмы ИТС более года и обошлось в 2,5 млн долларов. Тестирование показало, что нейронная сеть позволяет обнаруживать 38 % мошеннических случаев, в то время как существующая экспертная система – только 14 %. Для настройки системы были использованы также методы нечеткой логики и генетической оптимизации.

Анализ потребительского рынка. Фирма IBM Consulting создала нейросетевую систему, прогнозирующую свойства потребительского рынка. Одним из основных маркетинговых механизмов является распространение купонов, дающих право покупки определенного товара со скидкой. Так как затраты на рассылку купонов довольно велики, решающим фактором является эффективность рассылки, т. е. доля клиентов, воспользовавшихся скидкой.

Для повышения эффективности купонной системы важно было провести предварительную сегментацию рынка, а затем адресовать клиентам каждого сегмента именно те купоны, которыми они с большей вероятностью воспользуются. В терминах анализа данных здесь требовалось решить задачу кластеризации, что и было успешно проделано с помощью сетей Кохонена. На втором этапе для потребителей каждого из кластеров подбирались подходящие коммерческие предложения, а затем строился прогноз объема продаж для каждого сегмента.

Другой вариант решения этой же задачи избрала компания GoalAssist Corp., исследуя механизм предоставления поощрительных призов за приобретенные покупки. Обычные методы прогнозирования откликов потребителей оказались недостаточно точны. Так, спрос на одни призы оказался слишком велик, в то время как другие призы остались невостребованными. Для повышения точности прогнозирования было решено использовать нейронные сети. Первая нейронная сеть, по-

строенная с помощью пакета NeuroShell Classifier, решала задачу классификации откликов. Вторая самоорганизующаяся сеть, реализованная в нейропакете NeuroShell Predictor, осуществляла количественное прогнозирование. Средняя ошибка прогноза составила всего около 4 %.

Фирма Neural Innovation Ltd. использует при работе с маркетинговыми компаниями конкретную стратегию прямой рассылки. Вначале рассылается 25 % от общего числа предложений и собирается информация об откликах потребителей. Затем эта информация обрабатывается нейронной сетью, которая осуществляет поиск оптимального сегмента потребительского рынка для данного товара. Затем остальные 75 % предложений рассылаются в указанный сегмент. При этом эффективность рассылки существенно возрастает.

Проектирование и оптимизация сетей связи. Одна из важнейших задач в области телекоммуникаций – нахождение оптимального пути пересылки трафика между узлами, может быть успешно решена с помощью нейронных сетей. В данном случае важны две особенности: во-первых, решение должно быть адаптивным, то есть учитывать текущее состояние сети связи и наличие сбойных участков, а во-вторых, необходимо найти оптимальное решение нужно очень быстро, в реальном времени. Нейронные сети прекрасно приспособлены для решения такого рода задач.

Прогнозирование потребления энергии. Фирма Solutions разработала нейросетевую систему анализа данных об энергопотреблении, фиксирующую измерение потребляемой энергии для каждого клиента через каждые 15 минут, с учетом того, что некоторые из измерений ошибочны. Система позволяет выявлять ошибочные измерения, а также прогнозировать потребление энергии в каждый момент времени. Знание точного прогноза позволило энергетической компании использовать гибкую тарифную политику и увеличить свою прибыль.

Приведем перечень промышленных применений нейронных сетей, на базе которых либо уже созданы коммерческие продукты, либо реализованы демонстрационные прототипы:

Промышленное производство:

- управление манипуляторами;
- управление качеством;
- управление процессами;
- обнаружение неисправностей;
- адаптивная робототехника;
- управление голосом.

Нефтяная и химическая промышленность:

- анализ геологической информации;
- идентификация неисправностей оборудования;

- разведка залежей минералов по данным аэрофотосъемок;
- анализ составов примесей;
- управление процессами.

Военная промышленность и авиация:

- обработка звуковых сигналов (разделение, идентификация, локализация, устранение шума, интерпретация);
- обработка радарных сигналов (распознавание целей, идентификация и локализация источников);
- обработка инфракрасных сигналов (локализация);
- обобщение информации;
- автоматическое пилотирование.

Биомедицинская промышленность:

- анализ рентгенограмм;
- обнаружение отклонений в ЭКГ.

Банки и страховые компании:

- автоматическое считывание чеков и финансовых документов;
- проверка достоверности подписей;
- оценка риска для займов;
- прогнозирование изменений экономических показателей.

Административное обслуживание:

- автоматическое считывание документов;
- автоматическое распознавание штриховых кодов.

Телевидение и связь:

- адаптивное управление сетью связи;
- сжатие и восстановление изображения.

Служба безопасности:

- распознавание лиц, голосов, отпечатков пальцев

Таким образом, нейронные сети идеально приспособлены для обнаружения сложных зависимостей в отсутствие априорных знаний об исследуемой системе или процессе. Нейронные сети также можно использовать везде, где обычно применялись линейные методы и алгоритмы, и производилось оценивание при помощи статистических методов анализа, таких как регрессионный, кластерный, дискриминантный анализ, временные ряды.

Технологии нейронных сетей, нечеткой логики и генетических алгоритмов применимы практически в любой области. В технической физике, в процессах разделения возможности использования могут быть направлены на решение комплекса задач управления динамическими объектами, например каскадом по разделению изотопов и его оптимизацией, построения адаптивных и интеллектуальных систем управления.

Главный недостаток существующих разработок – это их малая серийность из-за узкой специализации и, в связи с этим, высокая их стоимость при сравнительно небольшой сложности нейронной сети.

Контрольные вопросы и задания

1. Назовите основные элементы модели искусственного нейрона и их назначение.
2. Как можно математически описать функционирование нейрона?
3. Что собой представляет функция активации, её типы? Запишите математически.
4. Чем отличаются детерминистская и стохастические нейросетевые модели? Какая из этих моделей и когда предпочтительнее?
5. Что называется нейронной сетью? Какие типы нейронных сетей Вам известны?
6. Что Вам известно об алгоритмах обучения нейронной сети? Объясните на примере.
7. Охарактеризуйте особенности RBF-сетей, назовите их преимущества и недостатки.
8. Что собой представляют обобщенно-регрессионные нейронные сети?
9. Назовите задачи и этапы исследований при нейросетевом моделировании.
10. Назовите преимущества и недостатки в использовании нейронных сетей.
11. Охарактеризуйте ключевые функции систем искусственного интеллекта.
12. В чём отличие символьных моделей искусственного интеллекта и нейросетевых моделей, как они взаимодействуют?
13. Охарактеризуйте область применения искусственных нейронных сетей и пути их дальнейшего развития.

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Churchland P.S. and T.J. Sejnowski. The Computational Brain. – Cambridge, MA: MIT Press, 1992. – 560 pp.
2. Cybenko G. Approximation by superposition of a sigmoidal function // Mathematics of Control, Signals and Systems, 1989. – Vol. 2. – P. 303–314.
3. Fischler M.A. and O. Firschein. Intelligence: The Eye, The Brain and The Computer Reading – MA: Addison-Wesley, 1987. – 331 pp.
4. Fodor J.A. and Z.W. Pylyshyn. Connectionism and cognitive architecture: a critical analysis // Cognition, 1988 – Vol. 28. – P. 3–72.
5. Hecht-Nielsen. R. Kolmogorov's Mapping Neural Network Existence Theorem // IEEE First Annual Int. Conf. on Neural Networks. – San Diego, 1987. – Vol. 3. – P. 11–13.
6. Hornik K., Stinchcombe M., Wite H. Multilayer Feedforward. Networks are Universal Approximators // Neural Networks, 1989. – Vol. 2. – P. 359–366.
7. Kerlirzin P. and F. Vallet. Robustness in multilayer perceptrons // Neural Computation, 1993. – Vol. 5. – P. 473–482.
8. McCulloch W.S. and W. Pitts. A logical calculus of the ideas immanent in nervous activity // Bulletin of Mathematical Biophysics, 1943. – Vol. 5. – P. 115–133.
9. Авчухов В.В., Паюсте Б.Я. Задачник по процессам тепломассообмена. – М.: Энергоатомиздат, 1986. – 144 с.
10. Адлер Ю.П., Маркова Е.В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий. – М.: Наука, 1976. – 280 с.
11. Ахназарова С.Л., Кафаров В.В. Методы оптимизации эксперимента в химической технологии. – М.: Высш. школа, 1985. – 327 с.
12. Вергун А.П., Савостина Н.В. Оптимизация разделительных процессов. – Томск, 2002. – 36 с.
13. Галушкин А.И. Современные направления развития нейрокомпьютеров // Зарубежная радиоэлектроника. Успехи современной радиоэлектроники. – 1998. – № 1. – С. 3–15.
14. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности/ Г.К. Вороновский, К.В. Махотило, С.Н. Петрашев, С.А. Сергеев. – Харьков: Основа, 1997. – 112 с.
15. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере. – Новосибирск: Наука, 1996. – 276 с.
16. Грачёв Ю.П., Плаксин Ю.М. Математические методы планирования эксперимента. – М.: ДеЛи принт, 2005. – 296 с.

17. Гухман А.А. Применение теории подобия к исследованию процессов тепломассообмена. – М.: Высш. школа, 1974. – 328 с.
18. Ермаков С.М. Жиглявский А.А. Математическая теория оптимального эксперимента. – М.: Наука, 1987. – 320 с.
19. Захаров Е. И., Рябчиков Б. Е., Дьяконов В. С. Ионнообменное оборудование атомной промышленности. – М.: Энергоатомиздат, 1987. – 248 с.
20. Зенкин В.И. Практический курс математического и компьютерного моделирования. – Калининград: Изд-во РГУ им. И. Канта, 2006. – 151 с.
21. Интеллектуальное управление динамическими системами / С.Н. Васильев, А.К. Жерлов, Е.А. Федосов, Б.Е. Федунюв. – М.: Физматлит, 2000. – 352 с.
22. Кафаров В.В., Глебов М.Б. Математическое моделирование основных процессов химических производств. – М.: Высш. школа, 1991. – 400 с.
23. Косарев А. Б., Лисенков А. Н. Основы методологии планирования эксперимента и построения аппроксимационных моделей при внедрении АСКУЭ на электрифицированных железных дорогах ОАО «РЖД» // Вестник ВНИИЖТ. – 2005 – № 3.
24. Круглов В.В., Борисов В.В. Гибридные нейронные сети. – Смоленск: Русич, 2001. – 224 с.
25. Круглов В.В., Борисов В.В. Искусственные нейронные сети. Теория и практика. – М.: Горячая линия-Телеком, 2001. – 382 с.
26. Круглов В.В., Дли М.И., Голунов Р.Ю. Нечеткая логика и искусственные нейронные сети. – М.: Физматлит, 2001. – 224 с.
27. Кутателадзе С.С. Анализ подобия в теплофизике. – Новосибирск: Наука, 1982. – 280 с.
28. Луговской В.И. Синявский К.С. Дубс Р.В. Математическое моделирование химико-технологических процессов. – Одесса: ОПУ, 2004. – 35 с.
29. Луканин В.Н. Теплотехника. – М.: Высш. школа, 1999. – 671 с.
30. Математическое моделирование и оптимизация химико-технологических процессов. / В.А. Холоднов, В.П. Дьяконов, Е.Н. Иванова, Л.С. Кирьянова. – СПб.: Профессионал, 2003. – 478 с.
31. Медведев В.С., Потемкин В.Г. Нейронные сети. MATLAB 6. – М.: Диалог-МИФИ, 2002. – 496 с.
32. Миркес Е.М. Нейрокомпьютер. Проект стандарта. – Новосибирск: Наука, 1998. – 343 с.
33. Нейроинформатика /А. Н. Горбань, В. Л. Дунин-Барковский, Е. М. Миркес и др. – Новосибирск: Наука, 1998. – 296 с.

34. Осовский С. Нейронные сети для обработки информации. – М.: Финансы и статистика, 2002. – 344 с.
35. Островский Г.М., Волин Ю. М., Зиятдинов Н. Н. Методы оптимизации химико-технологических процессов. – М.: КДУ, 2008. – 422 с.
36. Пантелеев А. В. Методы оптимизации в примерах и задачах. – М.: Высш. школа, 2005. – 544 с.
37. Перегудов Ф.И., Тарасенко Ф.П. Основы системного анализа. – Томск: Изд-во НТЛ, 2001. – 396 с.
38. Резниченко С.С. Математическое моделирование в горной промышленности. – М.: Недра, 1981. – 216 с.
39. Розен А. М. Теория разделения изотопов в колоннах. – М.: Атомиздат, 1960. – 438 с.
40. Терехов В.А., Ефимов Д.В., Тюкин И.Ю. Нейросетевые системы управления. Кн. 8. / Под общ. ред. А.И. Галушкина. – М.: ИПРЖР, 2002. – 480 с.
41. Уоссермен Ф. Нейрокомпьютерная техника. – М.: Мир, 1992. – 184 с.
42. Усков А.А. Интеллектуальные технологии управления. Искусственные нейронные сети и нечётная логика. – М.: Горячая линия-Телеком, 2004. – 143 с.
43. Хайкин Саймон. Нейронные сети. – М.: Вильямс, 2006. – 1103 с.
44. Хехт-Нильсен Р. Нейрокомпьютинг: история, состояние, перспективы // Открытые системы. – 1998. – № 4. – С. 24–28.

Учебное издание

ДОРОФЕЕВА Людмила Ивановна

МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ РАЗДЕЛИТЕЛЬНЫХ ПРОЦЕССОВ

Учебное пособие

Научный редактор,
доктор физико-математических наук,
профессор *А.П. Вергун*

Редактор
Верстка
Дизайн обложки

Подписано к печати 28.06.2011. Формат 60х84/8. Бумага «Снегурочка».
Печать XEROX. Усл.печ.л. 16. Уч.-изд.л. 7.
Заказ 000-11. Тираж 35 экз.



Томский политехнический университет
Система менеджмента качества
Томского политехнического университета сертифицирована
NATIONAL QUALITY ASSURANCE по стандарту ISO 9001:2000



ИЗДАТЕЛЬСТВО  ТПУ. 634050, г. Томск, пр. Ленина, 30.