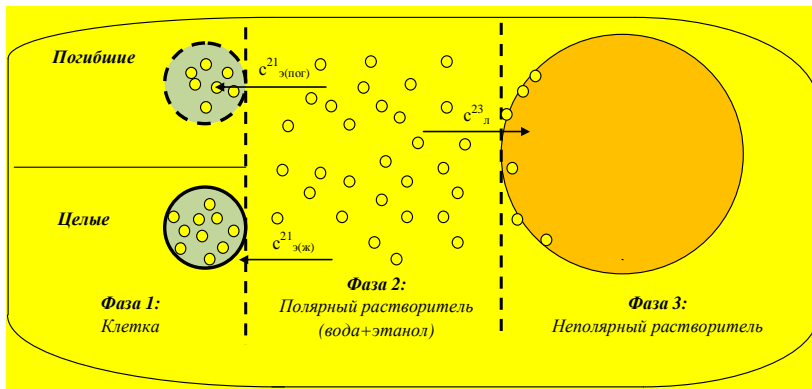


Д. С. ДВОРЕЦКИЙ, С. И. ДВОРЕЦКИЙ, Е. И. АКУЛИНИН,
О. О. ГОЛУБЯТНИКОВ, М. С. ТЕМНОВ

СИСТЕМНЫЙ АНАЛИЗ И ОПТИМИЗАЦИЯ БИОТЕХНОЛОГИЧЕСКИХ ПРОИЗВОДСТВ



Тамбов

◆ Издательский центр ФГБОУ ВО «ТГТУ» ◆

2019

Министерство науки и высшего образования
Российской Федерации

**Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Тамбовский государственный технический университет»**

**Д. С. ДВОРЕЦКИЙ, С. И. ДВОРЕЦКИЙ, Е. И. АКУЛИНИН,
О. О. ГОЛУБЯТНИКОВ, М. С. ТЕМНОВ**

СИСТЕМНЫЙ АНАЛИЗ И ОПТИМИЗАЦИЯ БИОТЕХНОЛОГИЧЕСКИХ ПРОИЗВОДСТВ

Утверждено Учёным советом ТГТУ
в качестве учебного пособия для студентов бакалавриата и
магистратуры высших учебных заведений технологических
направлений и специальностей

Учебное издание



**Тамбов
Издательский центр ФГБОУ ВО «ТГТУ»
2019**

УДК 661.935, 519.633.2

ББК 51.001.57

С40

Рецензенты:

Доктор технических наук, профессор, главный научный сотрудник

ФГБНУ ВНИИТиН

С. А. Нагорнов

Доктор технических наук, профессор, профессор кафедры

«Техника и технологии производства нанопроductов»

ФГБОУ ВО «ТГТУ»

Е. Н. Туголуков

С40 **Системный анализ** и оптимизация биотехнологических производств : учебное пособие / Д. С. Дворецкий, С. И. Дворецкий, Е. И. Акулинин, О. О. Голубятников, М. С. Темнов. – Тамбов : Издательский центр ФГБОУ ВО «ТГТУ», 2019. – 160 с. – 100 экз.

ISBN 978-5-8265-2097-0

Является систематическим введением в современную теорию и практику системного анализа, математического моделирования и оптимизации технологических процессов биотехнологических производств, изложены основные принципы, методология и методы системного анализа аппаратурно-технологического оформления автоматизированных процессов биотехнологии, оптимизации и принятия решений при их разработке и проектировании.

Предназначено для студентов бакалавриата и магистратуры высших учебных заведений технологических направлений и специальностей, может быть полезно для повышения квалификации преподавателей вузов и специалистов, изучающих и использующих методы системного анализа, математического моделирования, оптимизации и проектирования автоматизированных технологических процессов биотехнологических производств.

УДК 661.935, 519.633.2

ББК 51.001.57

ISBN 978-5-8265-2097-0 © Федеральное государственное бюджетное образовательное учреждение высшего образования «Тамбовский государственный технический университет» (ФГБОУ ВО «ТГТУ»), 2019

ПРЕДИСЛОВИЕ

Учебное пособие написано на базе лекций по системному анализу и оптимизации технологических процессов биотехнологических производств, читаемых одним из авторов бакалаврам, магистрантам и аспирантам в Тамбовском государственном техническом университете. Слушатели этих лекций имели в основном высшее образование, хорошо знали «свои» технологии и производства, обладали математическими знаниями в объёме обычного курса высшей математики для технологических специальностей/направлений, имели некоторый опыт работы с отраслевыми информационными и компьютерными технологиями, но при этом недостаточно знали современную вычислительную математику, оптимизацию, автоматизацию, а математическое моделирование технологических процессов рассматривали как способ составления уравнений материального и теплового баланса и последующего «беспроблемного» решения их на ЭВМ. Поэтому наши лекции, помимо изложения основ системного анализа и математического моделирования, всегда содержали дополнительные знания/сведения, способствующие повышению математической культуры слушателей; что естественно отразилось на содержании данного пособия.

Семантически учебное пособие состоит из трёх частей. В первой части (разделы 1, 2) изложены общая характеристика и основные понятия биотехнологического производства как системы (раздел 1), основные принципы, методология и классификация методов системного анализа, оптимизации и принятия решений, характеристика и особенности задач системного анализа биотехнологических производств, характеристика этапов системного анализа, анализ состояния биотехнологической системы, разработка и проектирование автоматизированных биотехнологических процессов (раздел 2).

Во второй части (раздел 3) математическое моделирование трактуется как исторически третий (после эксперимента и физического моделирования) метод получения новых знаний путём численного исследования его математической модели на ЭВМ. Под биотехнологическим процессом (БТП) понимается техническая аппаратура, в которой осуществляются био- и физико-химические преобразования исходных веществ в конечные продукты; такой БТП характеризуется измеримыми координатами и био-физико-химической моделью (БФХМ), содержащей субъективные гипотезы/аксиомы о его свойствах. В разделе 3 показаны конечные, дифференциальные (в обыкновенных и частных производных) и интегральные уравнения, которые можно использовать для создания математической модели БТП, если их свойства соот-

ветствуют гипотезам БФХМ. Иначе необходим синтез математической модели (ММ) согласно процедуре: структурная и параметрическая идентификация уравнений модели, получение экспериментальных данных, анализ и обеспечение адекватности ММ; далее формируются три метода построения математической модели. Понятие «численное исследование математической модели БТП» трансформируется в постановку ряда математических задач моделирования, численные решения которых определяют те или иные свойства БТП. Среди этих задач могут быть некорректно поставленные, обратные, плохо обусловленные и тому подобные задачи, нуждающиеся в дополнительном анализе и регуляризации.

В третьей части (раздел 4) приводятся постановки оптимизационных задач БТП и детерминированной оценки альтернатив БТП, численные методы решения безусловных и условных оптимизационных задач. Изложены методы принятия решений и оценки вариантов БТП в условиях неопределённости.

ВВЕДЕНИЕ

Предлагаемое студентам учебное пособие по дисциплине «Системный анализ и оптимизация биотехнологических производств» рассматривается нами в качестве первого этапа их обучения искусству проектирования оптимальных технологических процессов биотехнологических производств. Биотехнологическое производство, как правило, включает несколько технологических процессов, реализуемых на отдельных технологических схемах, и отличается повышенной сложностью. Важной деталью методологии системного анализа является закрепление в сознании обучаемого понятия *системности* как свойства биотехнологического процесса (БТП) и, собственно, биотехнологического производства: БТП и производство функционируют как единое целое, и поэтому любое вмешательство извне нарушает согласованность в работе сопряжённых БТП (технологических схем и аппаратов) и может привести к неожиданным последствиям. Свойство системности в дальнейшем проиллюстрировано фактическими данными из промышленной практики, ещё раз подчёркивающими тесную связь биотехнологии с аппаратурным оформлением для её осуществления.

Далее будем считать, что техническое проектирование завершено и нам известна следующая информация о процессе, аккумулированная в технологическом регламенте:

- предварительное технико-экономическое обоснование способа получения конечного продукта;
- характеристики и свойства исходных веществ, промежуточных и конечных продуктов;
- био-физико-химические механизмы/основы и принципиальная технологическая схема БТП;
- перечень технического оборудования/аппаратуры и их основные характеристики (технические паспорта);
- исходные данные для составления уравнений баланса (массы, энергии, импульса) по всему БТП;
- рекомендуемые режимы функционирования технологического процесса и критерии их оптимальности;
- разработанная аналитическая модель динамики и статики стационарного БТП;
- рекомендации для проектирования систем автоматизации и контроля (перечни координат, подлежащих автоматическому регулированию и контролю; метрологические «допуски» регулирования, контроля и т.п.);

- желаемые показатели надёжности/восстановления систем автоматизации и БТП;

- методы и параметры очистки/ликвидации вредных отходов и выбросов.

Вместе с тем в технологическом регламенте отсутствуют:

- рекомендации по пускам и плановым/аварийным остановкам БТП;

- результаты оптимизации и расчёта надёжности автоматизированного оборудования и систем автоматизации/стабилизации;

- не показана возможность оптимизации технологических режимов и определения технико-экономических показателей в режиме реального времени и др.

Большинство указанных «неделок», однако, может быть выполнено при наличии математической модели БТП с помощью системного анализа, математического моделирования и вычислительного эксперимента на ЭВМ.

Глава 1

ОБЩАЯ ХАРАКТЕРИСТИКА ПРОЦЕССОВ БИОТЕХНОЛОГИИ

Биотехнология – прикладная наука о наиболее экономичных процессах (проводимых с участием биохимических превращений) производства необходимых человечеству продуктов, предметов и требуемых видов энергии с использованием биохимической деятельности микроорганизмов, изолированных клеток или их компонентов (ферментов). Эти процессы отличаются от химических большей сложностью, так как одновременно с тепло- и массопереносом в них протекает биохимическая реакция на микроуровне (в клетке), что приводит, параллельно с биохимическим превращением вещества, к увеличению (росту) биомассы в реакционном объёме. Таким образом, биотехнология – естественная, прикладная наука о способах и процессах производства биопродуктов (предметов потребления и средств производства), осуществляемых с участием биохимических превращений технически, экономически и социально целесообразным путём. Как наука биотехнология имеет в качестве предмета изучения – биохимическое производство, при этом цель изучения – создание целесообразного способа производства необходимых человеку продуктов, методы исследования – экспериментальный, математическое моделирование, системный анализ.

Объектом исследования биотехнологии является биохимическое производство – совокупность процессов и технологических операций, осуществляемых в машинах и аппаратах и предназначенных для переработки сырья путём биохимических превращений в биопродукты. Основное назначение биохимического производства – получение необходимого продукта. Общая структура биохимического производства включает в себя функциональные части (подсистемы), представленные на рис. 1.1.

Позиции 1 – 3 на рис. 1.1, собственно, биохимическое производство, в котором сырьё перерабатывается в биопродукт при непосредственном участии живых микроорганизмов и выделенных из них ферментов (биологических катализаторов).

Подготовка сырья (поз. 1, рис. 1.1) включает в себя приготовление необходимых видов сырья для БТП: приготовление среды, обычно жидкой, включающей необходимые компоненты питания; стерилизацию среды для асептических БТП, для которых нежелательно попадание посторонней микрофлоры; подготовку и стерилизацию газов (обычно воздуха), очистку их от пыли, влаги и присутствующих микроорганизмов; подготовку посевного материала и биокатализатора; предварительную обработку сырья – измельчение, смешивание компонентов и т.д.

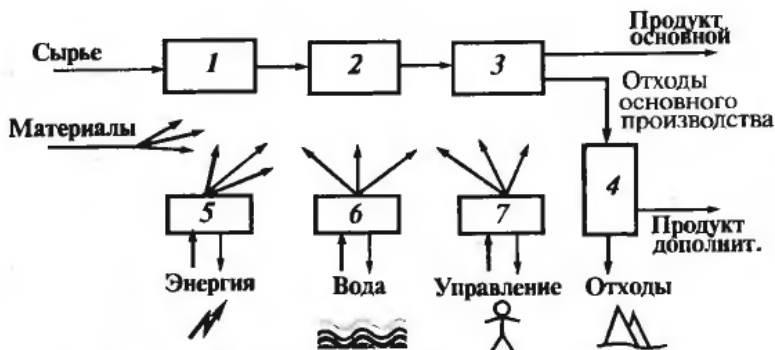


Рис. 1.1. Структура и функциональные подсистемы биохимического производства:

1 – подготовка сырья; 2 – переработка сырья; 3 – выделение основного продукта; 4 – санитарная очистка и утилизация отходов; 5 – энергетическая подсистема; 6 – подготовка вспомогательных материалов и водоподготовка; 7 – подсистема управления

Подготовленное сырьё проходит ряд превращений (поз. 2, рис. 1.1) на биотехнологической стадии, на которой с использованием биологических агентов (микроорганизмов, изолированных клеток, ферментов) происходит преобразование сырья в тот или иной целевой продукт. Биотехнологическая стадия включает в себя [1] биотехнологические процессы:

- *синтез* новых органических соединений;
- *ферментацию* – процесс, осуществляемый с помощью культивирования микроорганизмов;
- *биотрансформацию* – изменение химической структуры вещества под действием ферментативной активности клеток микроорганизмов или готовых ферментов;
- *биокатализ* – химические превращения вещества, протекающие с использованием биокатализаторов-ферментов;
- *биоокисление* – потребление загрязняющих веществ с помощью микроорганизмов в аэробных условиях;
- *метановое брожение* – переработка органических отходов с помощью ассоциации микроорганизмов в анаэробных условиях;
- *биокомпостирование* – снижение содержания вредных органических веществ ассоциацией микроорганизмов в твёрдых отходах;
- *биосорбцию* – удаление вредных примесей из газов или жидкостей микроорганизмами;

– *бактериальное выщелачивание* – перевод нерастворимых в воде соединений металлов в растворённое состояние под действием специальных микроорганизмов;

– *биодеградацию* – процесс деструкции вредных соединений под воздействием микроорганизмов-биодеструкторов.

Обычно биотехнологическая стадия имеет в качестве выходных потоков один жидкостной и один газовый. В случае если процесс протекает в твёрдой фазе (например, созревание сыра или биокомпостирование отходов), выходом является поток переработанного твёрдого продукта.

Чаще всего целевой продукт находится либо в самой биомассе, либо в жидкости. В обоих случаях необходимо сначала разделить эти две фазы. В зависимости от свойств биомассы и жидкости для этих целей могут быть использованы различные процессы [2]:

– *отстаивание* – разделение под действием гравитационных сил (обычно при очистке сточных вод);

– *фильтрация* – пропускание суспензии биомассы через фильтрующий материал;

– *сепарация, центрифугирование* – разделение под действием центробежных сил;

– *микрофильтрация, ультрафильтрация* – пропускание суспензии через мембраны с малым размером пор, обеспечивающее удержание клеток микроорганизмов на мембране и получение раствора, свободного от взвешенных клеток, причём ультрафильтрация задерживает не только клетки, но и крупные молекулы растворённых веществ;

– *коагуляция* – добавление в суспензию реагентов, способствующих образованию и осаждению более крупных клеточных агрегатов и отделению их от жидкости путём отстаивания;

– *флотация* – захват биомассы пузырьками пены и выделение её из пенной фракции.

Кроме того, целевой продукт может находиться либо вне клетки, либо внутри клетки. Для *внутриклеточных продуктов* сначала необходимо разрушить клеточную оболочку физическими, химическими или биотехнологическими методами [1]:

– *дзинтеграция клеток* (с помощью мелющих тел, путём замораживания и продавливания, воздействием ультразвука, методом декомпрессии – резкого сброса давления);

– *гидролиз* – под действием химических реагентов и температуры;

– *ферментализ, автолиз* – под действием ферментов при повышенной температуре.

После проведения предварительной операции разрушения клеток выделение целевого продукта осуществляется из раствора методами, которые являются общими для внеклеточных и внутриклеточных продуктов [2].

Экстракция – переход целевого продукта из водной фазы в несмешивающуюся с водой органическую жидкость (экстрагент). Экстракция прямо из твёрдой фазы (в том числе и биомассы микроорганизмов) называется *экстрагированием*.

Осаждение – выделение целевого продукта путём добавления к жидкости реагента, взаимодействующего с растворённым продуктом и переводящего его в твёрдую фазу.

Адсорбция – перевод растворённого в жидкости продукта в твёрдую фазу путём его сорбции на специальных твёрдых носителях (сорбентах).

Ионный обмен – то же, что и адсорбция, но в этом случае в твёрдую фазу переходят ионы (катионы или анионы), а не целиком молекула целевого продукта или примеси.

Отгонка, ректификация – эти методы используют для выделения растворённой в культуральной жидкости легкокипящих продуктов, например этилового спирта.

Ультрафильтрация, нанофильтрация и обратный осмос применяются для выделения высокомолекулярных соединений (белков, полипептидов, полинуклеотидов). Обратный осмос и нанофильтрация позволяют отделять даже небольшие по размеру молекулы.

Центрифугирование, ультрацентрифугирование используют для выделения вирусов, клеточных органелл, высокомолекулярных соединений.

При выделении продукта главная задача – отделить основную часть продукта, пусть даже и с некоторыми примесями, т.е. получается как бы неочищенный продукт. Поэтому, когда необходимо получать биопродукты высокой кондиции, добавляют ещё стадию их очистки и получения продукта максимальной чистоты. Эта задача решается с помощью уже рассмотренных выше процессов экстракции и экстрагирования, адсорбции, ионного обмена, ультрафильтрации и обратного осмоса, ректификации и ферментолиза, а также:

– *хроматографии* – сбора на твёрдом сорбенте нескольких растворённых веществ, близких по структуре, например смеси белков, нуклеотидов, сахаров, антибиотиков. При хроматографии они выходят из сорбента как бы по очереди, что и позволяет их разделять;

– *диализа* – процесса, в котором через полупроницаемую перегородку могут проходить низкомолекулярные вещества, а высокомо-

лекулярные остаются. Методом диализа осуществляют очистку вакцин и ферментов от солей и низкомолекулярных растворимых примесей;

– *кристаллизации* – процесса, в котором вещества имеют различные растворимости при разных температурах, медленное охлаждение позволяет формировать кристаллы из растворов целевых продуктов (например, кристаллы пенициллина), причём чистота их обычно высока, а вся грязь остаётся в маточном растворе.

После очистки целевой продукт часто находится в растворе с небольшими концентрациями примесей, дальнейшая задача – обеспечить его концентрирование. На стадии концентрирования применяют такие процессы, как *выпаривание, сушка, осаждение, кристаллизация* с фильтрацией образовавшихся кристаллов, *нано- и ультрафильтрация* [2].

Отходы производства или невостребованные продукты переработки сырья могут содержать как вредные компоненты, которые опасно выбрасывать в окружающую среду, так и полезные, которые целесообразно выбрасывать. Поэтому существенной стадией биохимического производства является санитарная очистка и утилизация отходов производства (поз. 4, рис. 1.1). Санитарная очистка или обезвреживание отходов – перевод токсичных составляющих в безвредные с тем, чтобы их можно было бы удалить из производства, не опасаясь заражения окружающей среды и обеспечивая безвредность существования в той экосистеме, в которой расположено производство. Биотехнологическое производство почти на всех стадиях имеет определённые стоки и выбросы в атмосферу. По существу очистка стоков – это отдельное биотехнологическое производство, имеющее свои подготовительные стадии, биохимическую стадию, стадию отстаивания биомассы активного ила и стадию дополнительной очистки и переработки осадка. Очищенная вода может быть возвращена в основное производство. Так организована, например, безотходная технология получения кормового белка из парафинов нефти. На заводе в г. Кириши после создания такой схемы удалось полностью ликвидировать технологические стоки в реку Волхов, а свежая вода стала забираться из реки Волхов только для компенсации потерь воды за счёт испарения из градилен и с готовым продуктом (кормовой белок имеет влажность до 10%).

Биохимическая промышленность потребляет довольно много энергии, для переработки сырья в биопродукты расходуется около 15% всех энергоресурсов, поэтому *энергетическая подсистема* является важной и сложной подсистемой биохимического производства (поз. 5, рис. 1.1), которая обеспечивает не только распределение энергии по стадиям производства, но и по возможности возвращение её после использования в переработке сырья.

Кроме энергии, в биохимическом производстве используются вспомогательные материалы, например сорбенты для очистки и выделения продуктов; вещества, с помощью которых создается среда, необходимая для осуществления БТП, и др. Особое место занимает вода – она используется для охлаждения технологических потоков, выработки пара, растворения и разбавления технологических потоков. Подготовка вспомогательных материалов и особенно водоподготовка (поз. б, рис. 1.1) – также важная и сложная часть биохимического производства. Поскольку вспомогательные материалы и вода обеспечивают БТП, но, как правило, не входят в целевые продукты производства, то система подготовки должна обеспечивать восстановление их свойств после проведения цикла технологических операций с их участием и возврат их в производство.

Подсистема управления (поз. 7, рис. 1.1) – часть производства для получения информации о его функционировании и для управления им. Обычно это – автоматизированная система управления технологическим процессом.

Биотехнологический процесс предназначен для целенаправленной переработки сырья и некоторых веществ в тот или иной полупродукт или конечный продукт путём осуществления различных видов биохимической деятельности микроорганизмов, совокупности ряда технологических операций и известных физико-химических процессов (диффузия, теплопроводность, химические преобразования, сорбция, десорбция и т.д.) в современной технической аппаратуре (автоматизированные машины и аппараты, биохимические реакторы, разделительные колонны, сушильные и холодильные агрегаты и т.п.). Миссия проектируемого БТП заключается в рациональном или экономически оптимальном и безопасном преобразовании определённого количества исходных веществ в конечные продукты биотехнологии заданного качества с использованием микроорганизмов или ферментов. Из функциональных подсистем биотехнологического производства, показанных на рис. 1.1, к биотехнологическому процессу можно отнести стадии 1 – 4, на которых происходит собственно переработка сырья в продукты.

Исследование и разработка отдельных процессов и их совокупности – биотехнологического процесса – основа профессиональной деятельности биотехнолога.

Определение полезности и эффективности биотехнологического производства и технологического процесса проводится по различным показателям, которые можно сгруппировать следующим образом [3].

Технические показатели определяют качество биотехнологического процесса. *Производительность (мощность) Q* производства – количество получаемого продукта или количество перерабатываемого сырья в единицу времени:

$$Q = G / t,$$

где G – количество получаемого продукта или перерабатываемого сырья за время t .

Обычно производительность выражают в количестве продукта за 1 ч или 1 сут, показывая максимальную производительность в непрерывном режиме. Производительность за длительный срок – один год – учитывает плановые остановки производства. Поэтому для биохимических производств для связи часовой или суточной производительности с годовой принимают, что производство работает 8000 ч или 330 суток в году.

Значение Q зависит от конкретного производства. Крупнотоннажные производства выпускают десятки тысяч тонн в год, а в малотоннажных производствах (реактивы, редкие металлы, продукты тонкого органического синтеза) производительность составляет килограммы и даже граммы продукта в час.

Расходный коэффициент показывает количество затраченного сырья, материалов и энергии на производство единицы продукта, его размерность очевидна: [кг сырья / т продукта], [м³ сырья / кг продукта], [кВт·ч / кг продукта], [Гкал / т продукта] и т.д. Расходный коэффициент показывает количественно затраты на производство продукта, но не отражает эффективности использования расходуемых компонентов.

Выход продукта – отношение реально получаемого количества продукта из использованного сырья к максимальному количеству, которое теоретически можно получить из того же сырья. Например, на получение 1 т HNO_3 реально расходуется 290...296 кг NH_3 , а теоретический расход составляет 270 кг. При этом выход продукта – 91...93%. Неполнота выхода продукта зависит от неполноты превращения, потерь, наличия примесей.

Интенсивность процесса – количество перерабатываемого сырья или образующегося продукта в единице объема аппарата – характеризует совершенство организации процесса.

Удельные капитальные затраты – затраты на технологическое оборудование, отнесенные к единице его производительности, – характеризуют эффективность организации процесса в отдельных аппаратах и в производстве в целом, совершенство используемых аппаратов.

Качество продукта определяет его потребительские свойства и товарную ценность. Показатель индивидуален для каждого продукта, он

может включать содержание (состав и количество) примесей, физические и химические показатели, внешний вид и размеры, цвет, запах и пр. Определяется нормативными документами (ГОСТ – государственный отраслевой стандарт, технические условия, сертификат качества).

Экономические показатели определяют экономическую эффективность производства. *Себестоимость продукции* – суммарные затраты на получение единицы продукции. Себестоимость складывается из следующих расходов: затрат на сырьё, энергию, вспомогательные материалы; единовременные капитальные затраты, распределяемые равномерно на срок эксплуатации оборудования, затраты на оплату труда работников. Общая структура себестоимости C

$$C = \left(\sum_i C_i G_i + E Z_k + Z_t / Q \right),$$

где C_i и G_i – цена и количество израсходованного сырья, энергии и материалов на производство продукта в количестве Q ; Z_k – капитальные затраты; E – коэффициент окупаемости капитальных затрат (их доля, отнесённая на время производства количества продукта Q , в среднем для биохимических производств $E = 0,15$ в расчёте на годовую производительность Q); Z_t – оплата труда.

Себестоимость имеет денежное выражение.

Производительность труда – количество продукции, произведённой в единицу времени (обычно за год) в пересчёте на одного работающего; характеризует эффективность производства относительно затрат труда.

Экономические показатели рассчитываются на основе технических показателей и их также называют техноэкономическими.

Эксплуатационные показатели характеризуют изменения, возникающие в биотехнологическом процессе и производстве во время их эксплуатации при появлении отклонений от регламентных условий и состояний. Влияние отклонений на показатели БТП, возможность управления процессом определяются эксплуатационными показателями.

Надёжность БТП характеризуют средним временем безаварийной работы либо числом аварийных остановов оборудования или производства в целом за определённый отрезок времени. Этот показатель зависит от качества используемого оборудования и правильности его эксплуатации.

Безопасность функционирования – вероятность нарушений, приводящих к нанесению вреда или ущерба обслуживающему персоналу, оборудованию, а также окружающей среде, населению.

Чувствительность БТП к нарушениям режима и изменению условий эксплуатации; определяется отношением изменения его показателей к этим отклонениям.

Управляемость и *регулируемость* характеризует возможность поддерживать показатели БТП в допустимых пределах, определяет величину допустимых изменений условий функционирования БТП, управляющие переменные и их взаимовлияние.

Социальные показатели определяют комфортность работы на данном производстве и его влияние на окружающую среду.

Безвредность обслуживания следует из сопоставления санитарно-гигиенических условий для обслуживающего персонала с соответствующими нормами по загазованности, запылённости, уровню шума и др.

Уровень автоматизации и механизации БТП определяет долю ручного труда в эксплуатации производства.

Экологическая безопасность – степень воздействия производства на окружающую среду и экологическую обстановку в регионе.

Перечень основных показателей биохимического производства свидетельствует о том, насколько высоки требования к качеству его разработки, проектирования, создания и эксплуатации. Нередко одновременное достижение наилучших показателей вступает в противоречие друг с другом. Необходимы компромиссные решения. Поэтому инженер-биотехнолог должен обладать не только обширными, разносторонними знаниями, но и высокой культурой.

Глава 2

СИСТЕМНЫЙ АНАЛИЗ БИОТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ И СИСТЕМ

Как следует из главы 1, *биотехнологическое производство* представляет собой совокупность машин, аппаратов и других устройств, связанных между собой материальными трубопроводами и паропроводами, линиями электрическими, транспортными и связи (для информации и управления) и функционирующих вместе, обеспечивая получение продукции. Такой объект называется *системой*.

Система – сложное образование, состоящее из совокупности элементов и связей между ними и функционирующее как единое целое. В биотехнологическом производстве элементы – это машины, аппараты и другие устройства, связи – материальные трубопроводы, паропроводы и другие линии, которые соединяют машины, аппараты, устройства. В элементах происходит превращение потоков (изменение их состояния – разделение, смешение, сжатие, нагрев, биохимические превращения и т.д.). По связям потоки (материальные, тепловые, энергетические) передаются из одного элемента в другой. Это позволяет представить биотехнологическое производство как биотехнологическую систему.

Биотехнологическая система (БТС) – совокупность аппаратов, машин и других устройств (элементов) и материальных, тепловых, энергетических и других потоков (связей) между ними, функционирующая как единое целое и предназначенная для переработки исходных веществ (сырья) в продукты. Элемент изменяет состояние потока. Это может быть отдельный аппарат (реактор, смеситель, абсорбер, кристаллизатор, теплообменник и т.д.) или их совокупность, производящая какое-то изменение состояния потока. Каскад биореакторов с теплообменниками и смесителями потоков между ними тоже производит изменение химического состава, и эту совокупность аппаратов можно также представить как элемент БТС. Степень детализации элемента: один аппарат или совокупность нескольких аппаратов; состав такой совокупности зависит от задачи исследований (определить те или иные показатели биотехнологического процесса, изучить особенности функционирования и т.д.).

С другой стороны, реакционный узел также можно представить как систему – совокупность биореакторов, теплообменников, смесителей (элементов) и потоков между ними (связей), функционирующую как единое целое. По отношению к БТС в целом это – *подсистема* как часть большой системы.

В зависимости от цели исследования не все аппараты будут влиять на интересующие исследователя свойства БТС. Если цель – определение производительности, выхода продукта и другие материальные показатели, то теплообменники, насосы и другое оборудование, не изменяющее состав потоков, можно не включать в БТС, которая для достижения этих целей будет представлять анализируемое производство. Если цель – обеспечение производства энергией, то в состав БТС включают энергетическое оборудование как её элементы.

Фактически исследование сложных БТС сводится к изучению её подсистем. Подсистемы могут быть выделены как по масштабу, так и функционально. Реакционный узел – по масштабу малая подсистема во всем БТП переработки сырья в продукты. Энергетическая подсистема по масштабу охватывает всё производство, но выполняет определённую функцию.

Потоки между элементами БТС передают вещество, энергию, информацию для управления процессами, осуществляемыми на стадиях биотехнологического производства. Связь в теории систем и поток в БТП – синонимы.

Представив биотехнологическое производство как биотехнологическую систему, можно использовать теорию систем для его исследования, получить научный метод его изучения.

2.1. ОПРЕДЕЛЕНИЕ СИСТЕМНОГО АНАЛИЗА

Системный анализ как дисциплина сформировался в результате возникновения необходимости исследовать и проектировать сложные системы, управлять ими в условиях неполной информации, ограниченности ресурсов и дефицита времени. Системный анализ является дальнейшим развитием целого ряда дисциплин, таких как исследование операций, теория оптимального управления, теория принятия решений, экспертный анализ, теория организации эксплуатации систем и т.д. Перечисленные теоретические дисциплины являются базой и методологической основой системного анализа. Таким образом, *системный анализ* – междисциплинарный курс, обобщающий *методологию исследования сложных технических, природных и социальных систем* [4, 5]. Широкое распространение идей и методов системного анализа, а главное – успешное их применение на практике, стало возможным только с внедрением и повсеместным использованием ЭВМ. В связи с этим академик Н. Н. Моисеев пишет [5], что системный анализ – это совокупность методов, основанных на использовании ЭВМ и ориентированных на исследование сложных систем – технических, экономических, экологических и т.д.

Центральной проблемой системного анализа является проблема принятия решений в условиях неопределённости. Неопределённость обусловлена многокритериальностью задач оптимизации, неопределённостью целей развития систем, неоднозначностью сценариев развития системы, недостаточностью априорной информации о системе, воздействием случайных факторов в ходе динамического развития системы и прочими условиями. Таким образом, главным содержанием дисциплины «Системный анализ» являются сложные проблемы принятия решений, при изучении которых неформальные процедуры, представления здравого смысла и способы описания ситуаций играют не меньшую роль, чем формальный математический аппарат.

В системном анализе можно выделить три главных этапа, которые всегда присутствуют в исследовании сложных систем:

- 1) построение модели исследуемого объекта;
- 2) постановка задачи исследования;
- 3) решение поставленной математической задачи.

Построение модели. При построении модели осуществляется математическое описание явлений и процессов, происходящих в системе. Поскольку знание всегда относительно, описание на любом языке отражает лишь некоторые стороны происходящих процессов и никогда не является абсолютно полным. С другой стороны, следует отметить, что при построении модели необходимо уделять основное внимание тем сторонам изучаемого процесса, которые интересуют исследователя. Построение математической модели есть основа всего системного анализа, центральный этап исследования или проектирования любой системы. От качества модели зависит результат всего системного анализа.

Постановка задачи исследования. На данном этапе формулируется цель анализа, которая должна быть формализована. Задача системного анализа состоит в проведении необходимого анализа неопределённостей, ограничений и формулировании, в конечном счёте некоторой оптимизационной задачи. Анализируя требования к системе, т.е. цели, которые предполагает достигнуть исследователь, и те неопределённости, которые при этом неизбежно присутствуют, исследователь должен сформулировать задачу на языке математики.

Решение поставленной задачи. Следует отметить, что задачи системного анализа могут иметь ряд особенностей, которые приводят к необходимости применения наряду с формальными процедурами эвристических подходов. Причины, по которым обращаются к эвристическим методам, в первую очередь связаны с недостатком априорной информации о процессах, происходящих в анализируемой системе. При этом трудности, возникающие в результате необходимости при-

менения неформальных процедур анализа, зачастую становятся определяющими. Ввиду этого проверка качества решения, его соответствия исходной цели исследования превращается в важнейшую теоретическую проблему.

Рассмотрим основные задачи, на решение которых направлены усилия специалистов и которые нуждаются в дальнейшей разработке. Во-первых, следует отметить задачи исследования системы взаимодействия анализируемых объектов с окружающей средой.

Задачи следующего типа связаны с конструированием альтернатив этого взаимодействия, альтернатив развития системы во времени и в пространстве. Важное направление развития методов системного анализа связано с попытками создания новых возможностей конструирования оригинальных альтернатив решения, неожиданных стратегий, непривычных представлений и скрытых структур. Другими словами, здесь речь идёт о разработке методов и средств усиления индуктивных возможностей человеческого мышления, в отличие от его дедуктивных возможностей, на усиление которых, по сути дела, направлена разработка формальных логических средств. Здесь можно выделить несколько важных направлений, таких как: разработка формального аппарата индуктивной логики, методов морфологического анализа и других структурно-синтаксических методов конструирования новых альтернатив, методов синектики и организации группового взаимодействия при решении творческих задач, а также изучение основных парадигм поискового мышления.

Задачи третьего типа заключаются в конструировании множества имитационных моделей, описывающих влияние того или иного взаимодействия на поведение объекта исследования. Речь идёт о разработке частных моделей, каждая из которых решает свои специфические вопросы.

Задачи четвёртого типа связаны с конструированием моделей принятия решений. Всякое системное исследование связано с исследованием альтернатив развития системы. Задача системных аналитиков выбрать и обосновать наилучшую альтернативу развития.

Рассмотренные задачи системного анализа не охватывают полного перечня задач. Здесь перечислены те, которые представляют наибольшую сложность при их решении. Чтобы успешно решать эти задачи, исследователь должен обладать широким кругозором и владеть богатым арсеналом методов и средств научного исследования.

Конечной целью системного анализа является разрешение проблемной ситуации, возникшей перед объектом проводимого системного исследования. Системный анализ занимается изучением проблем-

ной ситуации, выяснением её причин, выработкой вариантов её устранения, принятием решения и организацией дальнейшего функционирования системы, разрешающего проблемную ситуацию. Приведём пример из теории проектирования систем. Проблема проектирования сложных систем имеет два аспекта. Во-первых, требуется осуществить формализованное описание объекта проектирования, причём на этом этапе решаются задачи формализованного описания как статики системы, так и её поведения во времени, т.е. динамики системы. Во-вторых, требуется формализовать процесс проектирования. Составными частями процесса проектирования являются методы формирования различных проектных решений, методы их инженерного анализа и методы принятия решений, методы выбора наилучших вариантов реализации системы.

Важное место в процедурах системного анализа занимает проблема принятия решений, оптимальности решения. Для этого необходимо решать задачи оптимального управления, оптимального проектирования систем, включающих в себя большое количество элементов и подсистем. Развитие техники достигло такого уровня, при котором создание просто работоспособной конструкции само по себе уже не всегда удовлетворяет ведущие отрасли промышленности. Необходимо в ходе проектирования обеспечить наилучшие показатели по ряду характеристик новых изделий, например, добиться максимального быстродействия, минимальных габаритов, стоимости и прочего при сохранении всех остальных требований в заданных пределах. Таким образом, практика предъявляет требования разработки не просто работоспособного изделия, объекта, системы, а создания оптимального проекта.

2.2. ХАРАКТЕРИСТИКА ЭТАПОВ СИСТЕМНОГО АНАЛИЗА

В предыдущем пункте были сформулированы три укрупнённых этапа проведения системного анализа, которые являются основой решения любой задачи проведения системных исследований. Суть их состоит в том, что необходимо построить модель исследуемой системы, т.е. дать формализованное описание изучаемого объекта, сформулировать критерий решения задачи системного анализа, т.е. поставить задачу исследования и далее решить поставленную задачу исследования. Системотехнику требуется представить методику выполнения каждого из этапов.

Любая задача системного анализа начинается с построения модели исследуемой системы, для чего необходимо провести анализ структуры системы, выполнить анализ её компонентов, исследовать взаимосвязи между отдельными элементами.

Под элементом принято понимать простейшую неделимую часть системы. Понятие элемента условно, так как зависит от уровня иерархии рассмотрения объектов в структуре системы. Принято считать, что элемент – это предел членения системы с точки зрения решения конкретной задачи и поставленной цели.

В качестве следующей структурной компоненты рассмотрим подсистему. Подсистема – совокупность взаимосвязанных элементов, обладающая свойствами системы (в частности, свойством целостности), способная выполнять относительно независимые функции, подцели, направленные на достижение общей цели системы.

Если части системы не обладают свойством целостности и способностью выполнять независимые функции, а представляют собой совокупности однородных элементов, то такие части принято называть компонентами.

Структура отражает определённые взаимосвязи, взаиморасположение составных частей системы, её устройство, строение. При описании системы недостаточно перечислить элементы, из которых она состоит. Требуется отобразить систему путём расчленения её на подсистемы, компоненты и элементы и показать, каким путём обеспечивается в объекте выполнение поставленной цели. Таким образом, структура отражает наиболее существенные взаимоотношения между элементами и их группами, которые мало меняются при изменениях в системе и обеспечивают существование системы и её основных свойств. Структура системы характеризует организованность системы, устойчивую упорядоченность её элементов и связей. Структура системы – состав системы и схема связей между её элементами. Понятие структуры можно определить как совокупность отношений, заданных на множестве подсистем и элементов, образующих систему.

Понятие связи входит в любое определение системы, наряду с понятием элемента, и обеспечивает возникновение и сохранение структуры и целостных свойств системы. Понятие связи характеризует одновременно и статическое строение системы, и динамическое её поведение. Связь определяют как ограничение степени свободы элементов. Связь характеризуется направлением, силой и характером. По первому признаку связи делят на направленные и ненаправленные. По второму признаку различают сильные и слабые связи. По характеру выделяют связи подчинения, равноправные, генетические, связи управления. Различают также связи по направленности процессов – прямые и обратные. Обратные связи могут быть положительными, сохраняющими тенденции происходящих в системе изменений той или иной координаты, и отрицательными – противодействующими

тенденциям изменения выходной координаты. Обратная связь направлена на сохранение, стабилизацию требуемого значения координаты. Обратная связь является основой приспособления систем к изменяющимся условиям существования, основой саморегулирования и развития систем.

Следующее понятие – цель системы – важное понятие, лежащее в основе развития систем. Цели системы – желательные состояния или результаты поведения системы. Глобальная цель системы – конечное состояние, к которому стремится система в силу своей структурной организации. Цель достигается путём решения задач. Задачи системы – цели, которые желательно достичь к определённом моменту времени в пределах определённого периода функционирования системы.

Для описания системы создается её модель – отражение структуры системы, её элементов и взаимосвязей, направленное на отображение определённой группы свойств. Создание модели системы позволяет предсказывать её поведение в определённом диапазоне условий.

Структурные представления являются средством исследования систем. Одну и ту же систему можно представить различными структурами, выбор которых обусловлен содержанием исследований, проводимых на данном этапе. Принятый способ описания структур – графическое изображение, в котором элементы, компоненты, подсистемы и прочие объекты системы отображаются в виде вершин графа; связи между объектами представляют в виде дуг.

Сетевые структуры применяют для представления организационных структур, для изображения структурных схем систем, для представления информационного обеспечения. С помощью сетевых структур отображаются пространственные взаимосвязи между элементами, как правило, одного иерархического уровня.

Различают следующие виды сетевых структур:

- линейные структуры со строго упорядоченным взаимоотношением элементов «один к одному»;
- древовидную структуру, представляющую собой объединение многих линейных структур;
- кольцевую структуру, имеющую замкнутые контуры в соответствующих графах.

Иерархические структуры представляют собой декомпозицию системы в пространстве и применяются, прежде всего, для описания подчинённости элементов в структурах управления. Термин «иерархия» означает соподчинённость, порядок подчинения низших по должности лиц высшим. В иерархических структурах важно лишь выделение уровней соподчинённости, а между уровнями и между компонентами в пределах уровня, в принципе, могут быть любые взаимоотношения.

Таким образом, рассмотрены основные понятия, с помощью которых осуществляется решение задачи анализа системы. На данном этапе необходимо произвести изучение структуры системы, анализ её компонентов, выявление взаимосвязей между отдельными элементами, т.е. осуществить структурную декомпозицию системы. Основное содержание процедур, выполняемых на данном этапе, состоит в том, чтобы подготовить информацию к проведению работ по построению модели системы.

Цель изучения структуры системы, анализа её компонентов, выявления связей между отдельными элементами – отразить статическое состояние системы. Свойства системы реализуются в процессе её функционирования, т.е. в процессе динамического поведения системы. Чтобы построить модель системы, которая имела бы возможность отражать свойства и характеристики системы, реализующиеся в процессе её функционирования во времени, необходимо помимо структуры системы знать её параметры, поэтому следующим этапом работ при проведении системного анализа является сбор данных о функционировании системы и исследование информационных потоков.

Основное содержание этапа сбора данных о функционировании системы состоит в идентификации параметров системы в целях последующего включения их числовых значений в модель. Параметры отражают свойства системы.

Для определения параметров системы проводят наблюдения с целью сбора данных в процессе функционирования системы, либо для сбора данных организуются специальные экспериментальные исследования. В первом случае говорят, что данные получены в результате пассивного эксперимента, во втором случае имеет место активный эксперимент. Активный эксперимент проводится по специально составленному плану с использованием методов планирования эксперимента. При этом предусматривается возможность изменения входных координат объекта, влияющих на процесс функционирования системы, и проводится исследование выходных координат объекта в зависимости от уровней изменения входных координат. Результаты испытаний фиксируются с помощью измерения – алгоритмической операции, ставящей данному наблюдаемому состоянию системы или процесса в соответствие определённое обозначение: число, номер или символ. Такое соответствие обозначает то, что результаты измерений содержат информацию об исследуемой системе. Требуемая информация в виде оценок координат системы получается путём преобразования результатов измерения или, как ещё говорят, с помощью обработки экспериментальных данных.

Современное понимание измерений существенно шире только количественных измерений. Есть наблюдаемые явления, в принципе не допускающие числовой меры, но которые можно фиксировать в «слабых» шкалах, и эти результаты учитывают в моделях, получая качественные, но вполне научные выводы.

Выявление основных структурных элементов, динамических и информационных компонентов системы является предварительным этапом работы по построению модели системы.

2.3. БИОТЕХНОЛОГИЧЕСКИЙ ПРОЦЕСС КАК СИСТЕМА

Исследование сложных биотехнологических систем с использованием методов и средств теории систем получило название системного анализа БТС. Биохимическое производство состоит из десятков разнообразных аппаратов и устройств, связанных между собой разнообразными потоками. Исследовать его в целом при огромном многообразии его составных частей – задача не только сложная, но и малоэффективная. Представив биохимическое производство как биотехнологическую систему, проведём дальнейшую систематизацию частей производства, представленных в структуре БТС. Будем выделять подсистемы по двум признакам – функциональному и масштабному.

Функциональные подсистемы обеспечивают выполнение функций производства и его функционирование в целом. Технологическая подсистема – часть производства, где осуществляется собственно переработка сырья в продукты, биотехнологический процесс.

Энергетическая подсистема – часть производства, служащая для обеспечения энергией БТП. Соответствующая подсистема может быть представлена в зависимости от вида энергии: тепловая, силовая, электрическая.

Сложное биохимическое производство невозможно эксплуатировать без автоматизированной системы управления (АСУ) производством и АСУ БТП [6]. Они обеспечивают оптимальное (в определённом смысле) функционирование объекта. Степень оптимальности функционирования объекта количественно оценивается величиной одного или нескольких *показателей (критериев) качества*, имеющих очевидную технологическую или экономическую трактовку, таких как: производительность, степень превращения сырья, выход целевого продукта, себестоимость, прибыль и т.п. Для достижения этой цели АСУ должна выполнять следующие функции:

1) сбор информации о текущем состоянии объекта управления и вычисление критериев качества;

2) нахождение оптимальных режимов работы объекта (*оптимальных управлений*), при которых достигается максимальное (минимальное) значение одного из критериев качества;

3) реализация найденных оптимальных управлений на объекте.

В АСУ сбор и оперативная обработка информации, вычисление критериев, а также нахождение оптимальных управлений осуществляются техническими средствами, главным из которых является ЭВМ (компьютер). Задачи анализа найденных управлений и их реализация на объектах возлагаются на управленческий персонал, использующий технические устройства.

В АСУ оптимальные управления находят путём решения на ЭВМ некоторой математической задачи оптимизации, состоящей из следующих компонентов: критерия качества функционирования объекта, математической модели объекта (уравнений связи между выходными, управляющими и возмущающими переменными), ограничений на выходные и управляющие переменные. Во многих случаях качество работы объекта характеризуется набором разнородных критериев (показателей). В подобных случаях осуществляют «свёртку» (преобразование) этих показателей к одному критерию, называемому *основным*. Иногда основным критерием считают один из показателей, а остальные полагают на уровне заданных плановых показателей. Процедура решения задачи оптимизации заключается в нахождении с помощью ЭВМ таких управлений, при которых основной критерий достигает максимума (минимума) при соблюдении уравнений связи, ограничений и условий, налагаемых на остальные показатели качества работы объекта.

Помимо названных основных функций (сбор и обработка информации, нахождение оптимальных управлений), АСУ решает ряд других задач, например обеспечение защиты производства от нежелательных (аварийных) ситуаций, пуск и плановые/аварийные остановки БТП, расчёт надёжности оборудования и систем автоматизации производства.

Масштабные подсистемы выполняют определённые функции в последовательности процессов переработки сырья в продукты как отдельные части БТП. Их можно систематизировать в виде иерархической последовательности – иерархической структуры БТС. В структуре БТС минимальный элемент – отдельный аппарат (биореактор, абсорбер, ректификационная колонна, насос и пр.) (рис. 2.1). Это – низший масштабный уровень IV. Несколько аппаратов, выполняющих вместе какое-то преобразование потока, – элементы подсистемы III масштабного уровня (реакционный узел, система разделения многокомпонентной смеси и т.д.) (рис. 2.1). Совокупность подсистем второ-

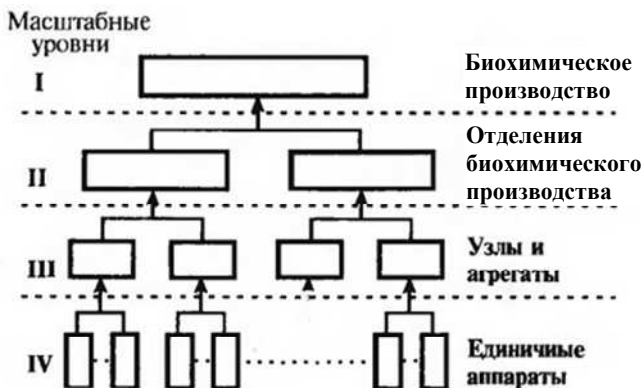


Рис. 2.1. Иерархическая структура биотехнологической системы

го уровня как элементы образует подсистему II уровня (отделения или участки производства) (рис. 2.1). К этим же подсистемам могут относиться водоподготовка, регенерация отработанных вспомогательных материалов, утилизация отходов. Совокупность отделений, участков образует БТС производства в целом (I уровень иерархии, рис. 2.1). Описанное выделение подсистем условно. В каких-то задачах выделение подсистем, элементов может быть иным.

Элементы и связи БТС. В описанной выше иерархической структуре (рис. 2.1) отдельные аппараты или агрегаты предназначены для определённого изменения состояния потока.

Потоки между аппаратами (связи между элементами) классифицируют по их содержанию. Материальные потоки переносят вещества и материалы по трубопроводам различного назначения, транспортёрами и другими механическими устройствами.

Энергетические потоки переносят энергию в любом её проявлении – тепловую, механическую, электрическую, топливо. Тепловая энергия и топливо для энергетических элементов передаются обычно по паропроводам (пар, горячие потоки, горючие газы и жидкости), механическая энергия – также по трубопроводам (в виде газов под давлением) или через вал двигателей и другие элементы электропривода. Провода, силовые кабели передают электрическую энергию.

Информационные потоки используются в системах контроля и управления процессами и производством. Используются электрические провода и тонкие капиллярные трубки в пневматических системах.

Последовательность прохождения потоков через элементы БТС определяет структуру системы и обеспечивает необходимые условия работы её элементов. Основные структуры связей показаны на рис. 2.2.

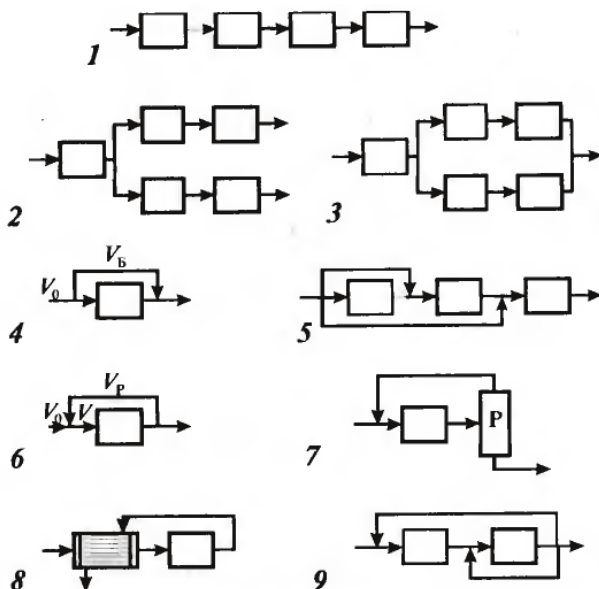


Рис. 2.2. Связи в биотехнологической системе:
 1 – последовательная; 2 – разветвлённая; 3 – параллельная;
 4, 5 – обводная (байпас); 6 – 9 – обратная (рециркуляционная) –
 рецикл полный (6, 9) и фракционный (7, 8)

Здесь прямоугольниками обозначены элементы, линии со стрелками – связи и направления потоков. *Последовательная связь* (схема 1, рис. 2.2): последовательная переработка сырья при применении разных технологических операций, углубление переработки сырья последовательными воздействиями на него, управление процессом путём необходимого управляющего воздействия на каждый элемент БТС.

Разветвлённая связь (схема 2, рис. 2.2): после некоторой операции поток разветвляется и далее отдельные потоки перерабатываются различными способами (используется для получения разных продуктов).

Параллельная связь (схема 3, рис. 2.2): поток разветвляется, отдельные части его проходят через разные аппараты, после чего потоки объединяются. Если мощность некоторых аппаратов ограничена, то устанавливают несколько аппаратов параллельно, обеспечивая суммарную производительность всей системы. Другое применение такой связи – использование периодических стадий в непрерывном процессе. В этом случае поочерёдно работает один из параллельных аппаратов. После завершения рабочего цикла одного аппарата поток переключается

ют на другой аппарат, а отключённый подготавливают к очередному циклу. Так включены адсорберы в схеме короткоцикловой безнагревной адсорбции: пока в одном адсорбере происходит поглощение, в другом регенерируют адсорбент. Ещё одно назначение параллельной схемы – резервирование на случай выхода из строя одного из аппаратов, когда такое нарушение может привести к резкому ухудшению работы всей системы и даже к аварийному состоянию. Такое резервирование называют «холодным», в отличие от резервирования, обусловленного периодичностью процесса, – «горячего».

Обводная связь, или *байпас* (схемы 4, 5, рис. 2.2): часть потока, не поступая в аппарат, «обходит» его. Такая схема используется в основном для управления процессом.

Обратная связь, или *рецикл* (схемы 6 – 9, рис. 2.2): часть потока после одного из аппаратов возвращается в предыдущий. Через аппарат, в который направляется поток G_p , проходит поток G больший, чем основной G_0 , так что $G = G_0 + G_p$. Количественно величину рецикла характеризуют двумя величинами: кратностью рециркуляции $K_p = G / G_0$ и отношением циркуляции $R = G_p / G$. Очевидно $R = (K_p - 1) / K_p$. Если выходящий из аппарата поток разветвляется, и одна его часть образует обратную связь (схема 6), то такая связь образует *полный рецикл* – составы выходящего потока и рециркулирующего одинаковы. Такую схему используют для управления процессом, создания благоприятных условий для его протекания.

Возможен возврат (рецикл) части компонентов после системы разделения Р (схема 7, рис. 2.2) для более полного использования сырья. Это – *фракционный рецикл* (возвращается фракция потока).

Приведённые типы связей присутствуют практически во всех БТС, обеспечивая необходимые условия их функционирования.

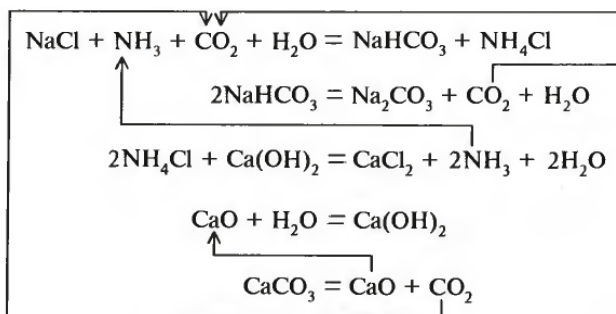
Модели БТС. Исследование системы, в том числе БТС, предполагает, что вначале она будет представлена моделью. Из определения системы как сложного образования, состоящего из совокупности элементов и связей между ними, представляется целесообразным представить её в виде схемы, графически. С другой стороны, качественные и количественные показатели её функционирования могут быть отражены словесными и математическими описаниями (моделями) происходящих в них процессов. Модели можно разделить на две группы: графические (в виде схем и других графических изображений) и описательные (в виде формул, уравнений). В каждой из названных групп также можно выделить несколько видов моделей, различающихся по форме и назначению:

– графические модели: функциональная, технологическая, структурная, специальная;

– описательные модели: химическая, операционная, математическая.

Здесь перечислены не все виды моделей, применяемые в исследовании БТС, а только те, которые будут использованы далее.

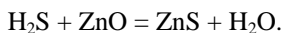
Химическая модель (схема) представляется основными реакциями (химическими уравнениями), которые обеспечивают переработку сырья в продукт. Например, производство соды Na_2CO_3 из поваренной соли NaCl и известняка CaCO_3 :



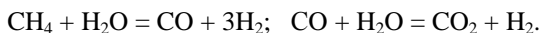
Химическая схема показывает генеральный путь превращения сырья в продукт. Но реализация этого превращения не ограничивается только данными уравнениями – необходимы ещё стадии, обеспечивающие эти химические преобразования или детализирующие их. Они представлены в других моделях процесса.

Операционная модель представляет основные стадии (операции) переработки сырья в продукт, в том числе обеспечивающие протекание основных превращений. Например, производство аммиака можно описать следующей операционной моделью.

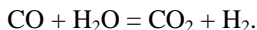
1. Очистка природного газа от серосодержащих соединений адсорбцией сероводорода, который мешает дальнейшим превращениям:



2. Конверсия метана с водяным паром. Природный газ (CH_4) и вода являются сырьём для получения одного из компонентов для синтеза аммиака – водорода H_2 . В этом превращении протекают одновременно две реакции:

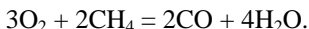


3. Конверсия оксида углерода с водяным паром (в предыдущем процессе оксид углерода CO не полностью превращается в CO_2 из-за равновесных ограничений):



В результате этого процесса достигается максимально возможное извлечение водорода из исходного сырья – метана CH_4 и воды.

4. Получение азота N_2 – второго исходного компонента для синтеза аммиака. В современных схемах его получают из воздуха «выжиганием» из него кислорода:

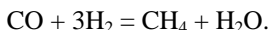


Это наиболее простой способ освободить азот воздуха от кислорода, тем более что сжигание части природного газа всё равно необходимо для обеспечения теплом всего процесса. Продукты горения – CO и H_2O – участники получения водорода.

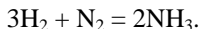
5. Абсорбция диоксида углерода – удаление CO_2 , образовавшегося при получении водорода. Его поглощают раствором моноэтаноламина:



6. Очистка газа от оксида углерода CO . После конверсии CO небольшое количество CO остаётся, и он мешает дальнейшим превращениям. Освобождаются от него путём превращения в метан:



7. Синтез аммиака (после всех стадий получена чистая азотоводородная смесь; примесь CH_4 , полученная в предыдущей стадии, мала):



Химическая и операционная схемы дают первое описание и представление о производстве и его основных стадиях. Для дальнейшего рассмотрения БТС удобнее использовать графические модели.

Функциональная модель (эскизная схема) строится на основе химической и операционной и наглядно отражает основные стадии биотехнологического процесса и их взаимосвязи. На эскизной технологической схеме стадии БТП изображаются прямоугольниками с их наименованиями, а также с указанием выходов по ним и условий осуществления процессов (температуры, давления, pH-сред и т.д.; связи (материальные потоки) изображают сплошными линиями со стрелками.

На рисунке 2.3 показана функциональная схема бактериального выщелачивания металлов из руд.

Руду перед началом производства измельчают, затем смешивают с раствором минеральных солей, и под воздействием микроорганизмов *Thyobacillus ferrooxidans* происходит переход ионов металла из руды в раствор. Далее суспензия отстаивается, твёрдая фаза возвращается на стадию выщелачивания, а жидкая фаза поступает на стадию биоокис-

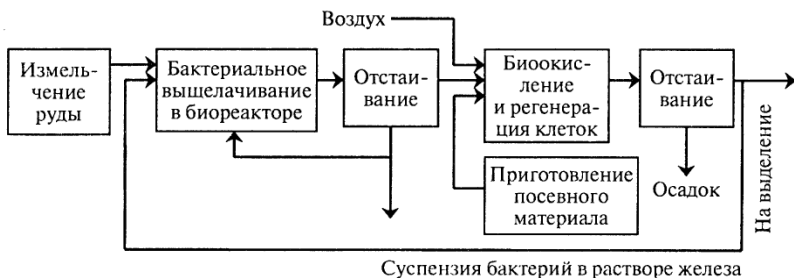


Рис. 2.3. Эскизная схема процесса бактериального выщелачивания металлов из руд

ления и регенерации клеток. Здесь двухвалентное железо под действием кислорода воздуха и бактерий превращается в трёхвалентное, а количество микроорганизмов возрастает. Далее раствор подвергается отстаиванию. При этом образовавшийся осадок, содержащий уже малые концентрации металла, поступает в отвал. Жидкий раствор с регенерированным трёхвалентным железом и ионами цветных металлов частично возвращается в биореактор, а частично идёт на стадию выделения.

Руководствуясь эскизной технологической схемой, приступают к составлению и расчёту уравнений материального и теплового балансов для всего производства. Расчёт материальных балансов стадий, связанных с биохимическими превращениями, проводят на основании стехиометрических уравнений реакции. Исходными данными для проведения расчёта являются: эскизная технологическая схема биохимического производства с указанием основных и побочных реакций; степень превращения сырья (выход целевого продукта); состав исходных веществ и состав реакционной массы, поступающей с предыдущей стадии; данные регламента о соотношении реагирующих веществ для стадий, связанных с химическими превращениями, и состав получаемых потоков для стадий фильтрации, сушки, ректификации и т.д.

Технологическая модель (принципиальная технологическая схема) отображает элементы БТС, порядок их соединения и последовательность осуществляемых в них типовых процессов и технологических операций. Принципиальную технологическую схему разрабатывают биотехнологи и специалисты по автоматизации на основе функциональной эскизной схемы и чертежей общего вида выбранного оборудования. При этом описываются способы доставки сырья в цех и отгрузки готовой продукции, обезвреживания и удаления отходов производства, обеспечения экологической безопасности и охраны труда, автоматизации производства.

Предварительный вариант технологической схемы вычерчивают с соблюдением определённых правил [7]. Аппараты можно изображать без соблюдения масштаба, но с учётом соотношения размеров. Обязательным является распределение их по высотным отметкам. Каждый аппарат изображается упрощённо в виде эскиза, отражающего его принципиальное устройство. Связи изображают обычно линиями со стрелками. Расположение аппаратов соответствует их примерной расстановке в цехе. На технологической схеме могут быть приведены данные о технологических параметрах процесса. Приборы и средства автоматизации, электрические, вычислительные и микропроцессорные устройства на принципиальной технологической схеме показывают в соответствии с ГОСТ 21.404–85.

Математическая модель. Приведённые выше модели (описания, схемы) дают общее представление о БТС. Для количественных выводов о её функционировании необходимо построить математическую модель. Математическая модель процесса в элементе устанавливает связь параметров выходящих потоков y_k из k -го элемента и входящих в него x_k . Переменные состояния потока – это его величина, состав (концентрации), температура, давление, теплосодержание и другие параметры. На состояние потока на выходе могут влиять некоторые входные (управляющие) переменные u_k , с помощью которых управляют процессом. В общем случае можно записать

$$y_k = F(x_k, u_k, k = 1, 2, 3, \dots). \quad (2.1)$$

Уравнения (2.1) – математические модели реактора, абсорбера, компрессора и других аппаратов и машин БТС.

Связи в БТС определяют, из какого элемента в какой передаётся поток. Поскольку передача потока происходит без его изменения, то уравнения связи в общем виде выглядят так:

$$x_k = \alpha_{l-k} y_l, \quad (2.2)$$

где $\alpha_{l-k} = 1$ для потока, выходящего из l -го элемента и входящего в k -й элемент; $\alpha_{l-k} = 0$, если между элементами l и k нет связи.

Система уравнений (2.1), (2.2) довольно громоздка и решается, как правило, численными методами.

Экспериментальное исследование. При проектировании БТП обычно используются опытные данные: это могут быть результаты специально выполненных исследований, сведения, собранные при изучении работы действующих установок, или информация, полученная из литературных источников. Часто предварительный выбор биотехнологической концепции метода производства основывается на результатах собственных исследований.

После предварительного анализа возможных вариантов концепции выбирается наиболее прогрессивный из них и проводятся исследования с целью экспериментальной проверки концепции и определения оптимальных условий осуществления процесса. При этом планирование эксперимента должно отвечать следующим требованиям.

1. Эксперимент необходимо ставить так, чтобы полученные данные можно было обобщить, т.е. выводы из опыта распространить на условия, в которых будет проводиться промышленный процесс.

2. Следует изучить влияние изменения тех параметров, которые существенным образом воздействуют на ход процесса, но уменьшить при этом число исследуемых параметров до возможного минимума.

3. План эксперимента должен быть составлен так, чтобы можно было получить нужную информацию, проводя минимально необходимое число опытов.

4. Результаты опытов следует подвергнуть критической оценке и представить в виде, удобном для их использования при проектировании.

Первые два указанных требования можно удовлетворить с помощью теории подобия. Для выполнения третьего требования нужно использовать статистические методы планирования эксперимента. Применение различных математических методов для вычисления ошибок эксперимента, составления диаграмм и номограмм, обобщения результатов в виде уравнений (в том числе критериальных, полученных на основе теории подобия) обусловлено четвёртым требованием.

Как уже упоминалось, исследование должно проводиться так, чтобы можно было получить требуемую (достаточно представительную) информацию, выполнив как можно меньшее число соответствующим образом запланированных опытов. Результаты этих экспериментов могут стать основой для решения интересующей нас технологической задачи. Обычно мы стремимся найти оптимальное решение и, следовательно, определить значения независимых параметров процесса (например, температуры, давления, концентрации), при которых зависимая переменная (например, выход целевого продукта, производительность, потери и т.д.) имеет оптимальное (максимальное или минимальное) значение.

Полное изучение процесса достигается редко. Оно основывается на нахождении системы уравнений, описывающих зависимости интересующих нас зависимых переменных от изменений независимых переменных. Проведение систематических исследований, называемых также однофакторными, с целью полного изучения процесса, в котором зависимая переменная y является функцией независимых переменных x_1, x_2, \dots, x_n

$$y = f(x_1, x_2, \dots, x_n)$$

реализуется таким образом, что исследуется ряд зависимостей с поочерёдным сохранением постоянных значений всех независимых переменных за исключением одной. Например, когда имеются только две независимые переменные x_1 и x_2 , то принимая, что обоим независимым переменным задаётся одинаковое число m значений, сначала исследуется ряд зависимостей $y = f(x_1, x_2 = \text{const})$, $x_2 = x_2^{(1)}$, $x_2 = x_2^{(2)}$, ..., $x_2 = x_2^{(m)}$, а затем $y = f(x_2, x_1 = \text{const})$, $x_1 = x_1^{(1)}$, $x_1 = x_1^{(2)}$, ..., $x_1 = x_1^{(m)}$, где $x_1^{(1)}$, $x_1^{(2)}$, ..., $x_1^{(m)}$, $x_2^{(1)}$, $x_2^{(2)}$, ..., $x_2^{(m)}$ – постоянные, m – число различных значений, заданных независимым переменным.

Предположим, что искомое значение y в рассматриваемом случае лежит на поверхности геометрической фигуры, показанной на рис. 2.4, и мы находим форму кривых пересечения этой поверхности плоскостями, перпендикулярными плоскости x_1, x_2 и параллельными либо плоскости x_1, y (рис. 2.4, *a*) либо плоскости x_2, y (рис. 2.4, *б*). Чем детальней нам нужно изучить зависимость $y = f(x_1, x_2)$, тем больше сечений m необходимо исследовать и тем больше будет число необходимых измерений.

Обычно только на основе многочисленных измерений можно при систематическом исследовании установить физико-химический закон, описывающий зависимость переменной y от независимых переменных x_1, x_2, x_3, \dots и являющийся результатом достаточно полного изучения исследуемого процесса.

Может случиться, что даже собранный обширный численный материал недостаточен для определения закона, которому подчиняется ход явления (например, часто не удаётся точно установить кинетическое уравнение сложного биохимического превращения). В этих случаях мы вынуждены аппроксимировать исследуемую зависимость эмпирическим уравнением в виде, например полинома

$$y = b_0 + \sum_{i=1}^n b_i x_i + \sum_{i=1}^n \sum_{j=1}^n b_{ij} x_i x_j + \sum_{i=1}^n b_{ii} x_i^2 + \dots$$

$$(i \neq j; i = 1, 2, \dots, n; j = 1, 2, \dots, n),$$

где n – число независимых переменных.

Чем больше число измерений x , тем с большей точностью найдутся значения постоянных коэффициентов b_0, b_i, b_{ij}, b_{ii} . Чем больше членов полинома принимается во внимание, тем точнее уравнение

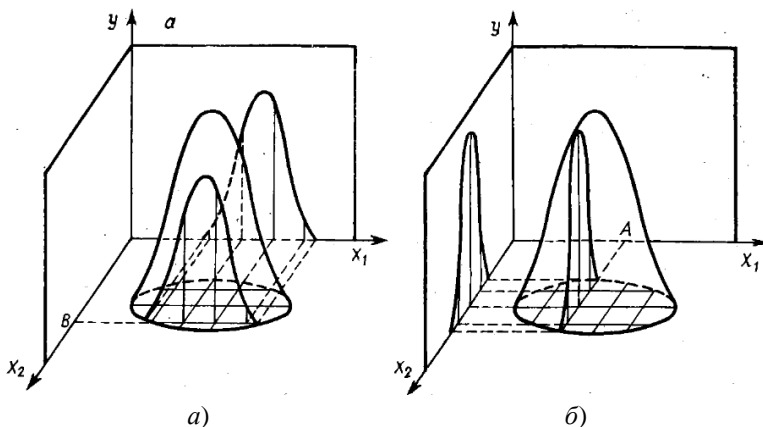


Рис. 2.4. Исследование зависимости $y = f(x_1, x_2)$:

a – $y = f(x_1, x_2 = x_2^{(m)})$ (кривая на плоскости x_1, y);

б – $y = f(x_1 = x_1^{(m)}, x_2)$ (кривая на плоскости x_2, y)

описывает исследуемое явление. Часто в технических исследованиях достаточно изучить небольшую область изменения y и использовать полиномиальное уравнение с двумя первыми членами (линейная аппроксимация) или с тремя (неполное квадратное уравнение).

Вследствие больших затрат труда и времени на полное систематическое изучение процесса инженер чаще всего довольствуется выборочным исследованием и, следуя интуиции, ведёт эксперимент в тех областях, где надеется отыскать оптимальное решение.

Математическая теория планирования эксперимента даёт возможность избежать излишне длительных и дорогостоящих систематических исследований и составить план решения задачи при минимально необходимом числе опытов [8].

2.4. СОСТОЯНИЕ БИОТЕХНОЛОГИЧЕСКОЙ СИСТЕМЫ

Чтобы судить о свойствах БТС, показателях её функционирования, эффективности управления процессами в ней, необходимо определить состояние БТС.

Состояние или *режим* БТС определяется переменными состояниями потоков (связей) и состоянием аппаратов (элементов). *Переменные состояния потоков* – химические, физико-химические и физические данные о потоке. К ним относятся переменные состояния – расход потока, фазовый состав, химический состав, температура, давление, теп-

лосодержание и физические свойства – теплоёмкость, плотность, вязкость и др.

Свойства потока могут быть определены из переменных его состояния и свойств индивидуальных компонентов.

Для каждой цели исследования и вида БТП (подсистемы) необходимо использовать не все данные. Если представляет интерес распределение веществ в системе, то достаточно определить величину потоков, их химический состав и, возможно, фазовый состав, давление. Энергетические потери в БТС будут зависеть от величины потоков и скорости их движения, давления, плотности, вязкости и, очевидно, теплосодержания.

Состояние элемента (аппарата, машины, агрегата) включает данные, от которых зависит изменение в элементе переменных состояния потока. Это могут быть управляющие воздействия или изменяющиеся в процессе эксплуатации характеристики аппарата и условия осуществления процесса в нём, например изменение условий осуществления процесса в реакторе вследствие дезактивации катализатора или в теплообменнике из-за загрязнения его поверхности.

Управляющие воздействия входят в описание элемента, их влияние на процесс в аппарате (элементе) проявляется при анализе БТС как эффект преобразования в элементе входящего в него потока в выходящий.

Расчёт БТС – определение переменных состояния потоков в БТС заданной структуры и заданного состояния её элементов. Очевидно, что при любых изменениях состояния потоков в элементе БТС, её подсистеме или БТС в целом должны соблюдаться законы сохранения массы и энергии.

Пусть имеется БТС, состоящая из теплообменника, биореактора и системы разделения (рис. 2.5) [3].

В ней пунктирными линиями окружены один элемент – реактор (I), подсистема – реактор с теплообменником (II) и вся БТС (III). В каждую из этих «оболочек» какие-то потоки входят и какие-то выходят. Внутри каждой оболочки происходят изменения входящих потоков в выходящие. Фактически каждая «оболочка» есть подсистема БТС. В соответствии с законами сохранения массы и энергии в стационарном состоянии должны соблюдаться соотношения для подсистем:

$$\text{для всех потоков} - \sum_j G_{j, \text{вх}} = \sum_j G_{j, \text{вых}};$$

$$\text{для } i\text{-го компонента} - \sum_j G_{ij, \text{вх}} + \sum_k G_{ik, \text{ист}} = \sum_j G_{ij, \text{вых}};$$

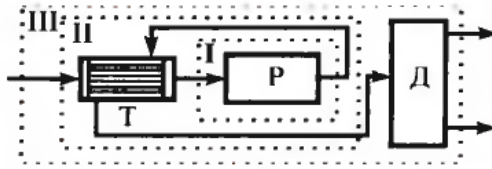


Рис. 2.5. Функциональная модель БТС:

T – теплообменник; P – реактор; Д – система разделения

$$\text{для тепловой энергии} - \sum_j Q_{j, \text{вх}} + \sum_k Q_{k, \text{ист}} = \sum_j Q_{j, \text{вых}},$$

где $G_{j, \text{вх}}, G_{j, \text{вых}}$ – массы j -го входящего в «оболочку» и выходящего из неё потоков; $G_{ij, \text{вх}}, G_{ij, \text{вых}}$ – массы i -го компонента в j -м входящем и выходящем потоках; $G_{ik, \text{ист}}$ – источник i -го компонента внутри «оболочки» в основном вследствие биохимических превращений, которых может быть несколько (k – номер источника), для образующихся веществ $G_{ik, \text{ист}} > 0$, для исчезающих – $G_{ik, \text{ист}} < 0$, для не реагирующих – $G_{ik, \text{ист}} = 0$; $Q_{j, \text{вх}}, Q_{j, \text{вых}}$ – теплосодержание j -х входящего и выходящего потоков; $Q_{k, \text{ист}}$ – источники теплоты внутри «оболочки», которыми могут быть химическая реакция, фазовые превращения, изменение импульса потока (сжатие, расширение).

В самом общем виде уравнения материального и теплового балансов выглядят следующим образом:

$$\frac{dM}{dt} = \sum M_{\text{вх}} - \sum M_{\text{вых}} + \sum M_{\text{ист}}; \quad \frac{dQ}{dt} = \sum Q_{\text{вх}} - \sum Q_{\text{вых}} + \sum Q_{\text{ист}},$$

где $\frac{dM}{dt}$ и $\frac{dQ}{dt}$ – накопление вещества и теплоты в выделенном элементарном объёме.

Для многофазных процессов уравнения составляют для каждой фазы и учитывают массо- и теплообмен между ними.

Поскольку состояния потоков меняются в элементах, рассмотрим различные типы элементов и составление для них материального баланса между входными и выходными потоками.

Элемент без химических и фазовых превращений (теплообменник, насос, дробилка). Естественно, что ни величина потока, проходящего через такой элемент, ни количество компонентов в потоке меняться не будут, и $G_{\text{вх}} = G_{\text{вых}}, G_{i, \text{вх}} = G_{i, \text{вых}}$.

Смеситель объединяет два потока:

$$G_{1, \text{ВХ}} + G_{2, \text{ВХ}} = G_{\text{ВЫХ}}; \quad G_{i1, \text{ВХ}} + G_{i2, \text{ВХ}} = G_{i, \text{ВЫХ}},$$

здесь индексы «1» и «2» относятся к двум входным потокам, индекс i – к i -му компоненту потока. Обозначим g_i , g_{i1} и g_{i2} – массовые концентрации i -го компонента соответственно в смешанном и в двух входящих в смеситель потоках:

$$g_{i1} = \frac{G_{i1, \text{ВХ}}}{G_{1, \text{ВХ}}}; \quad g_{i2} = \frac{G_{i2, \text{ВХ}}}{G_{2, \text{ВХ}}}; \quad g_i = \frac{G_{i, \text{ВЫХ}}}{G_{\text{ВЫХ}}} = \frac{G_{i1, \text{ВХ}} + G_{i2, \text{ВХ}}}{G_{1, \text{ВХ}} + G_{2, \text{ВХ}}}.$$

Определив соотношение величин смешиваемых потоков $\alpha = \frac{G_{1, \text{ВХ}}}{G_{2, \text{ВХ}}}$, преобразуем последнее равенство

$$g_i = \frac{G_{i1, \text{ВХ}} / G_{2, \text{ВХ}} + G_{i2, \text{ВХ}} / G_{2, \text{ВХ}}}{G_{i, \text{ВХ}} / G_{2, \text{ВХ}} + 1} = \frac{\alpha G_{i1, \text{ВХ}} / G_{2, \text{ВХ}} + G_{i2, \text{ВХ}} / G_{2, \text{ВХ}}}{\alpha + 1}$$

и получим связь весовых концентраций i -го вещества в смешиваемых (g_{i1} и g_{i2}) и выходящем (g_i) потоках:

$$g_i = (\alpha g_{i1} + g_{i2}) / (\alpha + 1).$$

Из последнего уравнения можно получить соотношения, в которых надо смешать два потока, чтобы получить заданную концентрацию g_i :

$$\alpha = (g_i - g_{i2}) / (g_{i1} - g_i).$$

Последние выражения позволяют решать задачи, часто встречающиеся в практике расчёта БТС. Имеется некоторый поток $G_{1, \text{ВХ}}$ с концентрацией i -го компонента g_{i1} . Необходимо получить поток с концентрацией этого компонента g_i . Сколько надо добавлять другого потока $G_{2, \text{ВХ}}$ с концентрацией этого компонента в нём g_{i2} ? Или с какой концентрацией g_{i2} надо добавлять второй поток, чтобы образующийся поток был определённой величины $G_{\text{ВЫХ}}$?

Делитель простой разделяет поток на два. Полагаем, что один из выходящих из делителя потоков (пусть это будет 1-й) составляет долю α входного потока:

$$G_{1, \text{ВЫХ}} = \alpha G_{\text{ВХ}}; \quad G_{2, \text{ВЫХ}} = (1 - \alpha) G_{\text{ВХ}}.$$

Соответственно, разделятся по потокам все компоненты:

$$G_{i1, \text{ВЫХ}} = \alpha G_{i, \text{ВХ}}; \quad G_{i2, \text{ВЫХ}} = (1 - \alpha) G_{i, \text{ВХ}}.$$

Реакционный элемент (биореактор или ферментер, реакторный узел), в котором протекают процессы микробиологического синтеза. Биохимический реактор-ферментер является основным аппаратом в любом биотехнологическом процессе, поэтому его работа в значительной степени определяет экономическую эффективность производства.

Стехиометрия биохимических превращений

В химии вещества вступают во взаимодействие в определённых соотношениях, которые находят отражение в химических уравнениях. Из них определяют изменение состава реагирующей смеси при протекании реакции, т.е. результат химического превращения. Стехиометрические уравнения показывают, в каких соотношениях вещества вступают в химическое взаимодействие. Так, в результате взаимодействия реагентов A и B получают продукты C и D и выделяется тепло Δh , то можно записать с помощью стехиометрических коэффициентов $\nu_A, \nu_B, \nu_C, \nu_D$ следующее стехиометрическое уравнение:

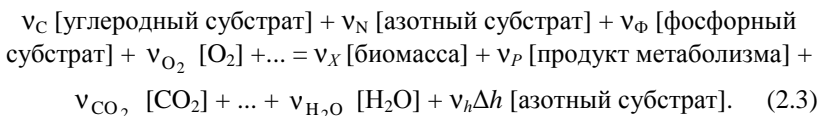


Как известно, стехиометрические коэффициенты подбирают таким образом, чтобы выразить фундаментальный закон природы – закон сохранения материи. Количество атомов любого элемента, входящего в составы веществ A, B, C, D , не должно изменяться в процессе превращения веществ.

В биологии также действует закон сохранения материи. В ходе биологических превращений в клетке перегруппировываются атомы углерода, азота, фосфора, водорода, кислорода и других жизненно важных химических элементов. Но общее количество каждого из этих элементов, включённое в структуры клетки, в точности равно количеству, взятому клеткой из питательной среды.

Процессы ферментации можно представить как систему, в которой происходит преобразование исходных реагентов (субстратов) в продукты (клетки и продукты метаболизма). В аэробных процессах в число субстратов входит кислород O_2 , а в число продуктов – диоксид углерода CO_2 . Другие субстраты и продукты в разных процессах различные, но стехиометрические соотношения между ними должны соблюдаться.

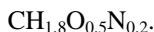
По аналогии со стехиометрией в химии можно записать стехиометрическое уравнение для микробиологического процесса [1]:



Разница между «истинной» стехиометрией и микробиологической в том, что наряду с веществами, имеющими определённую химическую формулу, в неё входит биомасса, состоящая из множества индивидуальных веществ – белков, нуклеиновых кислот, липидов и т.д, многие из которых даже не идентифицированы. Все эти вещества записывают целой совокупностью, «связкой». Отсюда следует, что какой-то существующей в природе истинной формулы биомассы нет, а состав биомассы микроорганизмов различных классов не очень-то и различается. Интересно, что «молекулярная масса» для каждой из этих формул будет 100 или около того (неучтённой золой можно пренебречь).

В стехиометрических расчётах обычно пренебрегают элементами, составляющими малую часть состава биомассы. В формуле отбрасывают, таким образом, фосфор и серу, а иногда и азот [1].

Часто оказывается удобным принимать такую «молекулярную массу», чтобы в ней оказался один атом (грамм-атом) углерода. Для этого в ранее вычисленных формулах биомассы достаточно все индексы при атомах разделить на индекс при атоме углерода. Такой условный моль, приведённый к атому углерода, называют С-моль. Для «усреднённой биомассы» часто используется формула, предложенная Стоутхамером для С-моля:



Из-за простоты эту формулу часто применяют в расчётах. Рассчитаем «молекулярную массу» С-моля:

$$M = 1 \cdot 12 + 1,8 \cdot 1 + 0,5 \cdot 16 + 0,2 \cdot 14 = 24,6.$$

Имея теперь брутто – формулу биомассы, можно проводить и стехиометрические расчёты, как в химических уравнениях.

Расчёт стехиометрического выхода на углеродный субстрат

Обычно наибольший интерес с учётом выхода составляет самый дорогой субстрат – углеродный. В качестве углеродного субстрата могут использоваться разные вещества: глюкоза, крахмал, этанол, метанол, парафины нефти, метан и др. Эти вещества можно также пере считать на С-моль. Например, для глюкозы с формулой $\text{C}_6\text{H}_{12}\text{O}_6$ С-моль будет иметь формулу CH_2O , для крахмала с формулой $(\text{C}_6\text{H}_{12}\text{O}_6)_n$ вид С-моля не изменится. Вот почему эти вещества определяются словом «углеводы». Рассчитав молекулярную массу С-моля субстрата и сравнив её с массой С-моля биомассы, можно найти теоретический выход биомассы, если весь углерод субстрата перейдёт в углерод биомассы (табл. 2.1).

2.1. Расчёт теоретического выхода биомассы для разных субстратов [1]

Субстрат	Химическая формула	Молекулярная масса субстрата	Молекулярная масса С-моля субстрата	Стехиометрический выход биомассы	Фактически измеренный выход биомассы, г/г
Глюкоза	$C_6H_{12}OH$	180	30	0,82	0,5
Крахмал	$(C_6H_{12}O_6)_n$	$180n$	30	0,82	0,5
Целлюлоза	$(C_6H_{12}O_6)_n$	$180n$	30	0,82	0,5
Этанол	C_2H_5OH	46	23	1,07	0,75
Метанол	CH_3OH	32	32	0,77	0,5
Парафины (<i>n</i> -Алканы)	$(CH_2)_nH_2$	$14n + 2$	~14	~1,76	1,0
Метан	CH_4	16	16	1,54	0,62

Из таблицы 2.1 следует, что различные субстраты дают различный стехиометрический выход по биомассе. Последний столбец таблицы включает в себя данные о фактически измеренном выходе биомассы в реальных микробиологических процессах, проведённых с использованием указанных в таблице субстратах. Из сравнения теоретического (стехиометрического) и фактического выходов видно, что тенденция прослеживается: максимальный выход – на парафинах, минимальный – на углеводах. При этом фактический выход гораздо ниже теоретического.

Чтобы как-то объяснить эти расхождения, ввели понятие «энергетический выход» биомассы. Это понятие базируется на том, что в каждом субстрате заключена энергия, которая зависит от степени восстановленности субстрата, которую легко подсчитать по известной формуле вещества. При этом, как и в биомассе, учитывают основные элементы – углерод, кислород и азот. Обобщённая формула субстрата – $C_mH_nO_pN_q$.

Степень восстановленности зависит от числа так называемых «доступных электронов», или «редоксонов». Установлено, что один атом углерода имеет четыре доступных электрона, один атом водорода –

один доступный электрон. Кислород доступных электронов не имеет, а наоборот, как бы забирает на себя два электрона, т.е. имеет отрицательное число (-2) доступных электронов. То же и с азотом, который забирает три доступных электрона.

Итак, для обобщённой формулы субстрата $C_mH_nO_pN_q$ степень восстановленности рассчитывают по формуле

$$\gamma_S = 4m + n - 2p - 3q .$$

Рассмотрим примеры:

Метан (CH_4) – $\gamma_S = 1 \cdot 4 + 4 = 8$.

Гексан (C_6H_{14}) – $\gamma_S = 4 \cdot 6 + 14 = 38$.

Метанол (CH_3OH) – $\gamma_S = 4 \cdot 1 + 4 - 2 \cdot 1 = 6$.

Глюкоза ($C_6H_{12}O_6$) – $\gamma_S = 4 \cdot 6 + 12 - 2 \cdot 6 = 24$.

Этанол (C_2H_5OH) – $\gamma_S = 4n + 2n + 2 = 6n + 2 = 32$.

n-Алканы ($(CH_2)_nH_2$) – $\gamma_S = 4 \cdot n + 4 = 8$.

Щавелевая кислота ($H_2C_2O_4$) – $\gamma_S = 4 \cdot 2 + 2 - 2 \cdot 4 = 2$.

Степень восстановленности различна прежде всего из-за разного числа углеродных атомов, а также и их соотношения с другими атомами (кислорода, азота, водорода). Для субстратов, как и для биомассы, можно ввести понятие С-моля и пересчитать для таких С-молей степень восстановленности. Это позволит сравнивать между собой различные субстраты.

Биомасса, как и другие вещества, согласно «формуле» может быть оценена по степени восстановленности (γ_X). Биомасса $CH_{1,8}O_{0,5}N_{0,2}$:

$$\gamma_X = 4 \cdot 1 + 1,8 - 2 \cdot 0,5 - 3 \cdot 0,2 = 4,2 .$$

Ерошкиным В. К. введено понятие «энергетический выход биомассы», представляющий собой отношение энергии в субстрате к энергии, заключённой в биомассе:

$$\eta = \gamma_S / \gamma_X .$$

Величина η показывает, какое количество С-молей биомассы можно получить из одного С-моля субстрата исходя из соотношения энергий.

Примеры:

Метан (CH_4) – $\eta = 8,0/4,2 = 1,9$ С-моль биомассы / С-моль субстрата.

Гексан (C_6H_{14}) – $\eta = 6,3/4,2 = 1,5$ С-моль биомассы / С-моль субстрата.

Метанол (CH_3OH) – $\eta = 6,0/4,2 = 1,43$ С-моль биомассы / С-моль субстрата.

Глюкоза ($C_6H_{12}O_6$) – $\eta = 4,0/4,2 = 0,95$ С-моль биомассы / С-моль субстрата.

Этанол (C_2H_5OH) – $\eta = 6,0/4,2 = 1,43$ С-моль биомассы / С-моль субстрата.

n-Алканы ($(CH_2)_nH_2$) – $\eta = 6,0/4,2 = 1,43$ С-моль биомассы / С-моль субстрата.

Щавелевая кислота ($H_2C_2O_4$) – $\eta = 1,0/4,2 = 0,24$ С-моль биомассы / С-моль субстрата.

В формуле для η выход выражен в С-молях биомассы на С-моль субстрата. От него можно перейти к обычному массовому выходу:

$$Y_{XS} = \eta \frac{M_X}{M_S},$$

где M_X и M_S – молекулярные массы С-молей биомассы и субстрата соответственно.

Примеры:

Метан (CH_4) – $Y_{XS} = 1,9/(24,6/16) = 2,92$ г биомассы / г субстрата.

Гексан (C_6H_{14}) – $Y_{XS} = 1,5/(24,6/14,3) = 1,72$ г биомассы / г субстрата.

Метанол (CH_3OH) – $Y_{XS} = 1,43/(24,6/32) = 1,1$ г биомассы / г субстрата.

Глюкоза ($C_6H_{12}O_6$) – $Y_{XS} = 0,95/(24,6/30) = 0,78$ г биомассы / г субстрата.

Этанол (C_2H_5OH) – $Y_{XS} = 1,43/(24,6/23) = 1,53$ г биомассы / г субстрата.

n-Алканы ($(CH_2)_nH_2$) – $Y_{XS} = 1,43/(24,6/14) = 2,54$ г биомассы / г субстрата.

Щавелевая кислота ($H_2C_2O_4$) – $Y_{XS} = 0,24/(24,6/45) = 0,13$ г биомассы / г субстрата.

Если теперь сравнить полученные выходы с фактическими для этих субстратов (см. табл. 2.1), то можно убедиться, что только для глюкозы энергетический выход биомассы даёт значения чуть ниже стехиометрического (0,78 против 0,82), но при этом отличие от фактического (0,5) все ещё велико. Для других субстратов получается, что биомассы может быть даже больше, чем по стехиометрическим расчётам, а это противоречит здравому смыслу. Ведь из одного атома углерода не может получиться два, как это предсказывает, например, энергетический выход по метану.

Таким образом, исходные предпосылки «энергетической» теории материального баланса неверны! Они не учитывают того, что в клетке одновременно протекают как бы два процесса:

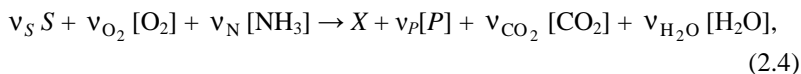
1) *катаболизм* – преобразование субстратов в элементарные биохимические соединения, в том числе и содержащие энергию, которые

в дальнейшем используются при биосинтезе. Есть здесь и «отходы» – диоксид углерода, тепло – то, что напрямую не идёт для синтеза биомассы и других продуктов;

2) *анаболизм* – построение макромолекулярных соединений (белков, нуклеиновых кислот, компонентов клеточной стенки), а также продуктов метаболизма с использованием универсальных биохимических «кирпичей».

При катаболизме энергия выделяется, при анаболизме – расходуется. Затраты энергии (да и материи – молекул, атомов) на анаболизм никогда не бывают равны разности энергетических уровней исходных и конечных веществ. Прежде всего это вытекает из второго закона термодинамики (нужна энергия для на осуществление биохимических превращений). С другой стороны, биохимические механизмы не идеальны, часть энергии и вещества расходуется на побочные продукты и тепло. Эти потери вещества учитываются реальными стехиометрическими (экономическими) коэффициентами, которые как раз и определяются из эксперимента – сравнением кривых изменения концентрации биомассы, субстрата и продуктов.

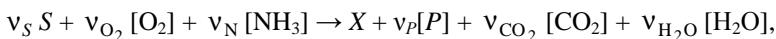
При этом общее стехиометрическое соотношение для объединённого процесса, включающего катаболизм и анаболизм, может быть записано в виде [1]



здесь биомасса выражена в С-молях, а субстрат – в обычных молях.

Другое соображение, поясняющее различия между стехиометрическим и фактическим выходом биомассы, заключается в следующем. В ходе культивирования наряду с «конструктивными» биохимическими превращениями (образование биомассы и различных продуктов метаболизма) протекают также биохимические процессы, которые в совокупности можно назвать «поддержанием жизнедеятельности» микроорганизмов. При этом происходит синтез клеточных компонентов, которые естественным образом деградируют, и для сохранения жизнеспособности клеток нужна их «репарация». В принципе мы можем так подавать субстрат, чтобы его хватало на поддержание жизнедеятельности, а новая биомасса микроорганизмов и продукты метаболизма при этом не образовывались. Ясно, что в таких процессах выход биомассы $Y_{XS} = 0$. И он (то, что в табл. 2.1 определено как фактический выход) не является постоянным и зависит от условий и скорости роста биомассы. Представленные цифры характеризуют максимально достигнутый выход, но и он далёк от стехиометрического, потому что затраты субстрата на поддержание жизнедеятельности существуют всегда.

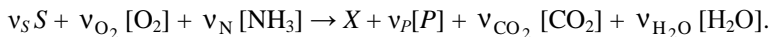
Рассмотрим теперь вопрос о нахождении стехиометрических коэффициентов, указанных в объединённом стехиометрическом уравнении (2.4) [1]:



если известны из эксперимента фактические данные по потреблению субстрата и образованию продуктов биохимического взаимодействия. Необходимо знать по меньшей мере количество израсходованного субстрата (в реальных мерах – граммы, килограммы и т.д.) и соответствующие данные по количеству образовавшейся биомассы микроорганизмов или продукта метаболизма – соответственно G_S , G_X и G_P (в пересчёте на весь аппарат), или для периодического процесса приращения их концентраций – соответственно ΔS , ΔX и ΔP .

Разделив эти величины на молекулярные массы соответственно субстрата, биомассы или продукта, получаем для этих веществ количества г молей (или кг молей), которые и являются основой для последующих стехиометрических расчётов.

Первый из этих расчётов – приведение всех количеств к одному С-молю биомассы. Мы уже упоминали, что все стехиометрические коэффициенты можно умножать или делить на одно и то же число. Поэтому условимся находить такие стехиометрические коэффициенты, которые дают стехиометрический коэффициент при биомассе (выраженной в С-молях), равный 1, как это записано в уравнении (2.4):



Таким образом, сразу находим v_S и v_P .

Для определения коэффициентов по другим веществам (O_2 , CO_2 , NH_3 и H_2O) необходимо составить и решить систему уравнений элементного баланса, имея в виду, что каждое вещество (субстрат, биомасса, продукт, вода, кислород, диоксид углерода) может быть описано общей элементной формулой $C_m H_n O_p N_q$.

При этом индексы в формуле субстрата будем обозначать как m_S , n_S , p_S , q_S , а в формуле продукта – соответственно m_P , n_P , p_P , q_P . Для биомассы соответствующие индексы $m_X = 1$; $n_X = 1,8$; $p_X = 0,5$; $q_X = 0,2$.

Тогда уравнение баланса по углероду можно записать следующим образом:

$$v_S m_S = 1 + v_P m_P + v_{CO_2} \cdot$$

То же для водорода:

$$v_S n_S + v_{NH_3} \cdot 3 = 1,8 + v_P n_P + v_{H_2O} \cdot 2 \cdot$$

Уравнение баланса по кислороду:

$$v_S p_S + v_{O_2} \cdot 2 = 0,5 + v_P p_P + v_{CO_2} \cdot 2 + v_{H_2O} \cdot 1.$$

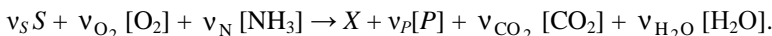
Уравнение баланса по азоту:

$$v_S q_S + v_{NH_3} = 0,2 + v_P q_P.$$

Решение составленной системы уравнений позволит найти все стехиометрические коэффициенты для данного процесса ферментации.

Пример. Процесс производства лимонной кислоты. Известно, что в процессе ферментации на 1 кг потреблённой сахарозы получается 0,6 кг лимонной кислоты и 0,3 кг сухой биомассы.

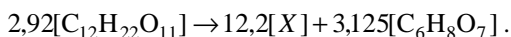
Рассчитаем стехиометрические коэффициенты уравнения (2.4)



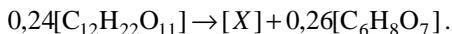
Переведём все величины в г/моли соответствующих веществ.

Молекулярная масса сахарозы ($C_{12}H_{22}O_{11}$) равна 342, лимонной кислоты ($C_6H_8O_7$) – 192, С-моля биомассы – 24,6. Отсюда получаем в молях:

$$\frac{1000 \text{ г}}{342 \text{ г}} [C_{12}H_{22}O_{11}] \rightarrow \frac{300 \text{ г}}{24,6 \text{ г}} [X] + \frac{600 \text{ г}}{192 \text{ г}} [C_6H_8O_7];$$



Пересчитаем все стехиометрические коэффициенты на С-моль биомассы, т.е. разделим обе части равенства на 12,2. Получим



Иначе говоря, для полного уравнения уже имеем $v_S = 0,24$ и $v_P = 0,26$.

Для нахождения оставшихся коэффициентов записываем систему уравнений элементного баланса для данного процесса.

При этом для субстрата (сахарозы) значения индексов $m_S = 12$, $n_S = 22$, $p_S = 11$, $q_S = 0$.

Для продукта (лимонной кислоты) $m_P = 6$, $n_P = 8$, $p_P = 7$, $q_P = 0$.

Баланс по углероду ($v_S m_S = 1 + v_P m_P + v_{CO_2}$)

$$0,24 \cdot 12 = 1 + 0,26 \cdot 6 + v_{CO_2}, \text{ откуда получаем } v_{CO_2} = 0,32.$$

Баланс по водороду ($v_S n_S + v_{NH_3} \cdot 3 = 1,8 + v_P n_P + v_{H_2O} \cdot 2$)

$$0,24 \cdot 22 + v_{NH_3} \cdot 3 = 1,8 + 0,26 \cdot 8 + 2v_{H_2O} \text{ или } 2v_{H_2O} - v_{NH_3} \cdot 3 = 1,4.$$

Баланс по кислороду ($v_S p_S + v_{O_2} \cdot 2 = 0,5 + v_P p_P + v_{CO_2} \cdot 2 + v_{H_2O} \cdot 1$)

$$0,24 \cdot 11 + 2 \cdot v_{O_2} \cdot 3 = 0,5 + 0,26 \cdot 7 + 2 \cdot 0,32 + v_{H_2O} \text{ или}$$

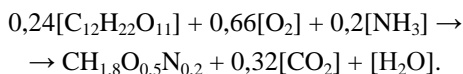
$$2v_{O_2} - v_{H_2O} = 0,32 .$$

Баланс по азоту ($v_S q_S + v_{NH_3} = 0,2 + v_P q_P$)

$$0 + v_{NH_3} = 0,2 + 0 , \text{ откуда } v_{NH_3} = 0,2 .$$

Далее из $2v_{H_2O} - v_{NH_3} \cdot 3 = 1,4$ получаем $v_{H_2O} = 1,0$, а из $2v_{O_2} - v_{H_2O} = 0,32$ следует $v_{O_2} = 0,66$.

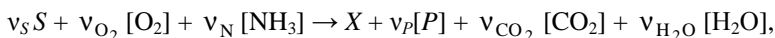
Таким образом, стехиометрическое уравнение для данного процесса можно записать следующим образом:



То, что в итоговом уравнении большинство коэффициентов оказались меньше единицы, не должно смущать. Главное – полученное уравнение даёт чёткую количественную связь между участвующими в биохимических превращениях веществами.

Расчёт тепла, выделяемого в биохимическом процессе

В стехиометрическом соотношении (2.4)



если провести расчёты степени восстановленности веществ, указанных слева (субстратов) и справа (продуктов), часто получается всё же определённая разница в пользу левой части. Это говорит о том, что энергия в субстрате больше энергии, заключённой в биомассе и продуктах метаболизма. Именно эта разность энергий и преобразуется в тепловыделение, имеющие место в ходе микробиологического процесса.

В биотехнологии определение величины тепловыделений является важной задачей, которая нужна для расчёта биореакторов. Это определение легко выполнить с помощью сравнения суммарной степени восстановленности субстратов, продуктов и образовавшейся биомассы с учётом соответствующих стехиометрических коэффициентов.

Формула для расчёта следующая:

$$\Delta h = \left(\sum_{i=1}^m v_i \gamma_{S_i} - \gamma_X - \sum_{j=1}^n v_j \gamma_{P_j} \right) \cdot 115 \frac{\text{кДж}}{\text{С - моль биомассы}} , \quad (2.5)$$

где Δh – тепловыделение; v_i – стехиометрический коэффициент для i -го субстрата; v_j – стехиометрический коэффициент для j -го продукта; m – общее количество субстратов; n – общее количество продуктов метаболизма; γ_{S_i} – степень восстановленности i -го субстрата; γ_{P_j} – степень восстановленности j -го продукта; γ_X – степень восстановленности биомассы.

Пример. Производство лимонной кислоты. Определить тепловыделение в пересчёте на С-моль биомассы. Подставим в формулу (2.5) ранее вычисленные значения v_S , γ_S , v_P и γ_P :

$$\begin{aligned}\Delta h &= v_S \gamma_S - \gamma_X - v_P \gamma_P = (0,24 \cdot 48 - 4,2 - 0,26 \cdot 18) \cdot 115 = \\ &= (11,52 - 4,2 - 4,68) \cdot 115 = 303,6 \text{ кДж/С-моль биомассы.}\end{aligned}$$

Чтобы получить количество тепловыделений на 1 кг израсходованной сахарозы, нужно учесть соотношение стехиометрических коэффициентов. Из стехиометрического уравнения видно, что на 1 кг сахарозы получено 12,2 С-молей биомассы. Следовательно, тепловыделение составляет

$$\Delta h = 303,6 \cdot 12,2 = 3704 \text{ кДж/кг сахарозы.}$$

Обычная же схема расчёта тепловыделения, принятая в химии, включает определение теплот сгорания всех входящих в реакцию веществ:

$$\Delta h = \left(\sum_{i=1}^m v_i (\Delta h_{S_i}) - \Delta h_X - \sum_{j=1}^n v_j (\Delta h_{P_j}) \right),$$

где Δh_S , Δh_X , Δh_P – теплота сгорания субстратов, биомассы, продуктов; m – общее количество субстратов; n – общее количество продуктов метаболизма.

Для химических соединений (субстрата, продукта) есть данные по Δh в справочниках или их можно определить экспериментально в «калориметрической бомбе». Для биомассы такие измерения сделаны:

$$\Delta h_{\text{дрожжей}} = 20,4 \text{ кДж/г биомассы;}$$

$$\Delta h_{\text{бактерий}} = 21,0 \text{ кДж/г биомассы.}$$

Стоит запомнить также теплоты сгорания для протеина (13,5 кДж/г), моносахаров – 15,7 кДж/г, дисахаров – 16,5 кДж/г, глицерина – 18,1 кДж/г. В среднем для биомассы принимают 22,0 кДж/г.

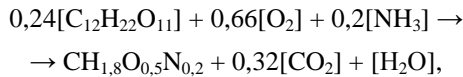
И ещё один способ расчёта теплового эффекта. Если известно количество кислорода, израсходованного на окисление субстрата, то тепловыделение можно определить по формуле

$$\Delta h = \Delta G_{O_2} \cdot 460 \text{ кДж},$$

где ΔG_{O_2} – количество израсходованного кислорода, г моль O_2 .

Как известно, в 1 г моле O_2 содержится 2 г атома кислорода, следовательно, это соответствует степени восстановленности, а точнее, степени окисленности $\gamma_S = -4$. Отсюда $460 = 4 \cdot 115$, где 115 – коэффициент в формуле (2.5).

В данном примере мы можем определить ΔG_{O_2} из найденных ранее стехиометрических коэффициентов. Интересно рассчитать количество г молей кислорода, полученное на С-моль биомассы. Оно численно равно стехиометрическому коэффициенту в итоговом уравнении



т.е. 0,66 молей O_2 на 1 г моль биомассы. Тогда из формулы $\Delta h = \Delta G_{O_2} \cdot 460$ кДж получим: $\Delta h = 0,66 \cdot 460 = 303,6$ кДж/С – моль биомассы.

Таким образом, расчёт по обеим формулам (2.5) и $\Delta h = \Delta G_{O_2} \cdot 460$ кДж даёт одинаковые результаты. Однако использование последней формулы более предпочтительно в случаях, когда легко определяется потребление кислорода культурой.

Стехиометрия химических превращений

В стехиометрическом уравнении соблюдается баланс по химическим элементам: количества каждого из них во всех исходных веществах и во всех продуктах равны между собой. Поэтому стехиометрические уравнения можно рассматривать как алгебраические и представить в виде

$$\sum_i v_i A_i = 0,$$

где A_i – наименование веществ (v_i – стехиометрический коэффициент i -го вещества).

В формуле $\sum_i v_i A_i = 0$ принято: $v_i < 0$ – для исходных веществ, которые расходуются; $v_i > 0$ – для исходных веществ, которые образуются. И можно ввести в уравнение вещество-инерт (не участвующее в превращении), для которого $v_i = 0$.

Из смыслового определения стехиометрического уравнения следует

$$(N_1 - N_{10})/v_1 = (N_2 - N_{20})/v_2 = \dots = (N_i - N_{i0})/v_i = \text{const}, \quad (2.6)$$

где $N_{10}, N_{20}, \dots, N_{l0}$ – начальные количества компонентов; N_1, N_2, \dots, N_l – количества компонентов после превращения; $\nu_1, \nu_2, \dots, \nu_l$ – соответствующие стехиометрические коэффициенты.

Подчеркнём, что стехиометрические уравнения показывают только соотношение количеств реагирующих веществ, но не отвечают действительно протекающим взаимодействиям.

Степень превращения. Соотношение (2.6) связывает между собой количества N_i всех реагентов. Конечно, удобнее ввести один показатель, через который можно выразить количества и концентрации всех реагентов. Таким показателем является *степень превращения* x какого-либо исходного вещества, показывающая, какая доля вещества превратилась. Для вещества A

$$x_A = (N_{A0} - N_A)/N_{A0}. \quad (2.7)$$

Используя выражения (2.6) и (2.7), можно определить количество любого вещества N_i в реагирующей смеси:

$$N_i = N_{i0} - (\nu_i/\nu_A)N_{A0}x_A. \quad (2.7a)$$

Степень превращения может быть определена для любого исходного реагента, например $x_B = (N_{B0} - N_B)/N_{B0}$. Из уравнения (2.6) и определений x_A и x_B получим

$$x_B = x_A \frac{\nu_A N_{A0}}{\nu_B N_{B0}}.$$

Если исходная реакционная смесь имеет неэквимолярный состав, то состав реакционной смеси удобнее выражать через степень превращения того вещества, которое в недостатке. Его называют *ключевым компонентом*, когда ключевой компонент превратится полностью, другие исходные компоненты ещё остаются.

Если компонент A участвует в нескольких реакциях, определим x_{A_j} – его степень превращения в j -й реакции; уравнение (2.7a) примет вид

$$N_i = N_{i0} - \frac{\sum \nu_{ij}}{\nu_{Aj}} N_{A0} x_{A_j}.$$

Концентрации компонентов реакционной смеси. Концентрация компонента в реакционной смеси $c_i = N_i/\sum_i N_i$. Общее количество реакционной смеси $\sum_i N_i$ найдём суммированием выражений (2.7a):

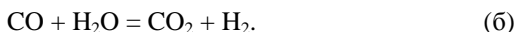
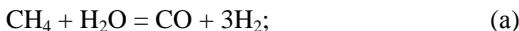
$$\sum_i N_i = \sum_i N_{i0} - x_A \sum_i (\nu_i/\nu_A).$$

Рассчитаем начальные концентрации компонентов как $c_{i0} = N_{i0}/\sum_i N_{i0}$ и из последнего выражения получим

$$c_i = \frac{c_{i0} - (v_i/v_A)c_{A0}x_A}{1 - (v_i/v_A)c_{A0}x_A}. \quad (2.8)$$

Числитель в уравнении (2.8) отвечает за изменение концентрации вследствие превращения веществ, а знаменатель – вследствие изменения объёма реакционной смеси.

Пример. Конверсия метана с водяным паром описывается уравнениями:



Ключевой компонент, который в недостатке, – метан, и использование его степени превращения в первом уравнении оправдано. Во втором уравнении ключевой компонент – CO, но использовать его степень превращения невозможно, так как начальная концентрация CO равна нулю. Суммирование последних уравнений даёт систему уравнений (а) и (б)



Определим x_1 и x_2 – степени превращения CH_4 по этим уравнениям. Исходная смесь содержит N_{10} и N_{20} молей CH_4 , H_2O , H_2 , CO , CO_2 , (N_1, N_2, \dots, N_5 соответственно) в реагирующей смеси:

$$N_1 = N_{10} - N_{10}(x_1 + x_2); \quad N_2 = N_{20} - N_{10}x_1 - 2N_{10}x_2;$$

$$N_3 = 3N_{10}x_1 + 4N_{10}x_2; \quad N_4 = N_{10}x_1; \quad N_5 = N_{10}x_2.$$

Далее, суммируя все N_i , определим общее число молей $\sum_i N_i$, а отношение $c_i = N_i/\sum_i N_i$ есть концентрация каждого компонента.

Селективность процесса и выход продукта. Превращение исходного вещества в сложной реакции протекает по разным направлениям. Не во всех реакциях образуется желаемый продукт, и для сложной реакции вводят кроме степени превращения следующие характеристики процесса.

Селективность процесса показывает, какая доля превратившегося исходного вещества пошла на образование заданного. Пусть превращение исходного вещества A описано несколькими уравнениями. Образование продукта R описано только одним. Стехиометрические коэффициенты для A и R – v_A и v_R . Селективность превращения A и R :

$$S_R = \frac{v_A(N_R - N_{R0})}{v_R(N_A - N_{A0})}. \quad (2.9)$$

Выход продукта показывает, какая доля всего вещества превратилась в заданное. Используя приведённые обозначения, определим E_R – выход продукта R :

$$E_R = v_A(N_R - N_{R0})/(-v_R N_{A0}). \quad (2.10)$$

Из определения степени превращения (2.7) и выражений (2.9), (2.10) получим

$$E_R = S_R x_A.$$

Селективность процесса и выход продукта характеризуют эффективность химического процесса по использованию сырья для получения нужного продукта.

Макростехиометрические характеристики процессов биотехнологии

Простейшей макростехиометрической характеристикой процесса ферментации является *выход по субстрату*, или *экономический коэффициент* (или *коэффициент выхода*). Его определяют следующим образом:

$$Y = X_k / S_0,$$

где X_k – количество биомассы, выросшей за весь цикл ферментации; S_0 – количество загруженного субстрата.

Аналогичный коэффициент может быть получен и по продукту метаболизма:

$$Y = P_k / S_0.$$

Ясно, что обозначения коэффициентов должны различаться, первый коэффициент будем обозначать как Y_{XS} , второй – как Y_{PS} .

Уточним формулы для расчёта введённых коэффициентов. Ясно, что в начале процесса присутствует некоторое количество биомассы, определяемое её концентрацией X_0 . Так что прирост её за время ферментации меньше, чем X_k , и равен $(X_k - X_0)$.

В то же время не весь субстрат до конца расходуется за время процесса; какая-то часть его, определяемая конечной концентрацией S_k , остаётся, так что потребление субстрата будет не S_0 , а $(S_0 - S_k)$.

Таким образом, сам биологический процесс более правильно характеризовать коэффициентом, найденным по формуле

$$Y_{XS} = \frac{X_k - X_0}{S_0 - S_k} = \frac{\Delta X}{\Delta S} \left[\frac{\text{г биомассы}}{\text{г субстрата}} \right].$$

Аналогичным образом для продукта метаболизма

$$Y_{PS} = \frac{P_k - P_0}{S_0 - S_k} = \frac{\Delta P}{\Delta S} \left[\frac{\text{г продукта}}{\text{г субстрата}} \right].$$

Коэффициенты выхода можно вычислять не только по начальным и конечным значениям параметров, но и за любой произвольно взятый промежуток времени Δt , используя для их вычисления соответственно этому промежутку определённые значения ΔX , ΔS и ΔP . В этом случае получим значения коэффициентов Y_{XS} , Y_{PS} за данный промежуток времени Δt .

Термодинамика химических превращений

Направленность химической реакции. Химический процесс принципиально осуществим, если реакция протекает с уменьшением химического потенциала, который есть изобарный потенциал, или энергия Гиббса, т.е. возможность протекания реакции определяется неравенством

$$\Delta G_{P,T} < 0,$$

где $\Delta G_{P,T}$ – изменение энергии Гиббса при превращении исходных веществ в продукты.

Для расчёта $\Delta G_{P,T}$ используют следующие данные и соотношения. В справочной литературе по термодинамике приведены значения стандартных энергий Гиббса $(\Delta G_{298})_{\text{обр}}$ образования веществ при температуре 298 К. Для химической реакции

$$\Delta G_{298}^0 = \sum_i \nu_i (\Delta G_{298}^0)_{\text{обр}i}.$$

Зависимость Гиббса от температуры определяется соотношением

$$\Delta G_T^0 = \Delta H_T^0 - T\Delta S_T,$$

где ΔH_T^0 , ΔS_T – изменения энтальпии и энтропии.

Зависимость энергии Гиббса от состава реакционной смеси (для газовых смесей) рассчитывается по формуле

$$\Delta G_{P,T} = \Delta G_T^0 + RT \ln \Pi p_i^{\nu_i},$$

где Π – знак произведения; p_i – парциальные давления компонентов.

Неравенство $\Delta G_{P,T} < 0$ позволяет определить возможность использования процесса для получения желаемого продукта – с этого начинают разработку нового способа производства. Другое использование неравенства – нахождение условий, предотвращающих протекание нежелательных реакций.

Пример. При конверсии метана с водяным паром возможно образование сажи. Одна из вероятных реакций: $\text{CO} + \text{H}_2 = \text{C} + \text{H}_2\text{O}$. Из справочных данных и формул получено для этой реакции: $\Delta G_T^0 = -132 + 0,133T$ (кДж/моль).

Конверсия метана в промышленном реакторе протекает вблизи равновесия. Соответствующее содержание компонентов, ответственных за образование сажи (углерода) для исходной смеси с соотношением пар:метан = 2:1 при температуре 873 К и давлении 0,1 МПа следующее: $C_{\text{CO}} = 0,071$; $C_{\text{H}_2} = 0,53$; $C_{\text{H}_2\text{O}} = 0,24$. Можно получить такие данные и для других температур и начального состава и рассчитать для них $\Delta G_{p,T}$ (рис. 2.6).

При начальном составе пар:газ = 1:1 во всём температурном интервале $\Delta G_{p,T} < 0$, и потому возможно выделение сажи. Для начальной смеси пар:газ = 2:1 в реакторе (при температурах ниже 900 К) возможно сажеобразование. И только при четырехкратном избытке водяного пара выделение сажи становится невозможным ($\Delta G_{p,T} > 0$).

Термодинамическая возможность протекания реакции не означает, что она обязательно должна протекать. Это зависит от скорости реакции. Но условия, определяющие принципиальную невозможность её протекания, гарантируют, что реакция протекать не будет.

Тепловой эффект реакции. Другой важной для технологических исследований термодинамической величиной является *тепловой эффект* Q_p , или *изменение энтальпии* ΔH реакции. Они противоположны по знаку: $Q_p = -\Delta H$. Изменение энтальпии рассчитывается через энтальпии образования веществ, участвующих в реакции. В зависимости от знака ΔH или Q_p реакции могут быть экзо- или эндотермическими:

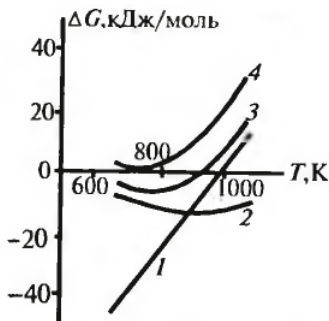
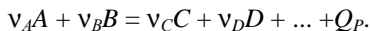


Рис. 2.6. Зависимость от температуры T энергии Гиббса ΔG реакции конверсии метана и реагирующей смеси в конверторе метана (1) с начальным соотношением пар: газ = 1:1 (2), 2:1 (3) и 4:1 (4)

$\Delta H < 0$; $Q_p > 0$ – реакция экзотермическая;

$\Delta H > 0$; $Q_p < 0$ – реакция эндотермическая.

Тепловой эффект реакции входит в запись термохимического уравнения:

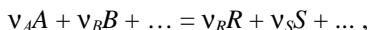


Величина Q_p зависит от записи химического уравнения. Например, значения Q_p в уравнениях $N_2 + 3H_2 = 2NH_3$ и $0,5N_2 + 1,5H_2 = NH_3$ различаются в 2 раза, поскольку различаются стехиометрические коэффициенты. В справочной литературе Q_p приводят прямо в уравнениях, или указывают изменение энтальпии при превращении 1 моля вещества.

Количество выделившейся (или поглощённой) теплоты h зависит от количества превращённого вещества ΔN . Если Q_p приводят в уравнениях, то $h = Q_p \Delta N_A / \nu_A$.

Равновесие. Одними из первых количественных расчётов в химической технологии были термодинамические. Они включали определение равновесия химически реагирующих систем: это остаётся актуальным и сегодня.

Равновесный состав смеси, реагирующей в соответствии с уравнением



связан с константой равновесия K_p :

$$K_p = \left(\frac{c_R^{\nu_R} c_S^{\nu_S}}{c_A^{\nu_A} c_B^{\nu_B}} \right)_{\text{равн}} = \prod_i (c_i^{\nu_i})_{\text{равн}}, \quad (2.11)$$

где ν_i – стехиометрические коэффициенты в уравнении химической реакции.

Концентрации c_i в уравнении – равновесные. Если реагирующая смесь описывается несколькими стехиометрическими уравнениями (сложная реакция), то химическое равновесие в такой смеси будет представлено таким же числом уравнений вида (2.11).

Константа равновесия K_p – термодинамическая функция:

$$\Delta G_T^0 = \Delta H_T^0 - T \Delta S_T = -RT \ln K_p.$$

Она зависит только от температуры. Если реакционная смесь идеальна (небольшие температуры и давления, разбавленные растворы), то, используя справочные значения ΔG_T , можно рассчитать K_p из последнего уравнения. Многие реальные смеси не являются идеальными и для них используют формулы расчёта K_p в виде

$$K_p = K_{p0} e^{-\Delta H / RT} = K_{p0} e^{Q_p / RT},$$

или более сложные зависимости, в том числе от давления.

Для газофазной реагирующей смеси используют равновесные парциальные давления компонентов p_i : $K_p = \prod_i (p_i^{v_i})_{\text{равн}}$. Для неидеальной смеси используют активности вместо концентраций, а вместо парциальных давлений – летучести.

Расчёт состояния биотехнологической системы. После того как определены структура БТС (элементы и связи между ними), описание элементов (математические модели аппаратов или балансовые соотношения) и их состояние, параметры входных или выходных потоков (какие – зависит от постановки задачи расчёта), можно переходить к расчёту состояния БТС (параметров потоков).

Подготовка данных – важнейший этап расчётов, от которых зависит достоверность конечных результатов.

Размерность используемых величин должна быть проверена на их «совместимость» в балансовых уравнениях. Для материального баланса удобно использовать размерность потоков в массовых величинах (кг/ч, т/ч и т.д.). Для характеристики химической реакции традиционно используют измерения в молях. Задаваемые концентрации могут иметь, например, размерности [%] или [доля]. «Доля» бывает массовая, мольная или объёмная, численно они различаются. Следует использовать основной способ проверки правильности использования размерностей: подставить в балансовое уравнение на место величин их размерности. Если размерности подобраны правильно, то все слагаемые уравнения должны получиться одинаковой размерности. В противном случае в каждый член уравнения надо вводить пересчётные коэффициенты.

Последовательность расчёта зависит от структуры БТС. *Линейная структура* – поток (потоки) проходит последовательно один элемент за другим (см. рис. 2.2).

Возможны разветвления потока, параллельное прохождение элементов, байпасирование. Последовательность расчёта очевидна. На рисунке 2.7 показан пример линейной БТС, на которой пронумерованы аппараты (элементы) и потоки. Известно состояние потока 0 на входе в систему, в первый по ходу элемент I. Вычисляют состояние потока 1 на выходе из него, который входит в элемент II. Из его описания определяют состояние потока 2. После его разветвления, определив состояние входного потока в элемент III (поток 3), продолжают такие же вычисления далее по нижней ветви схемы, а затем – по верхней. Расчёт состояния системы линейной структуры проводится от

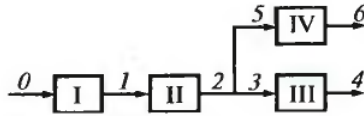


Рис. 2.7. Линейная структура БТС

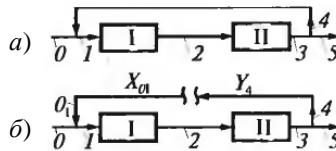


Рис. 2.8. Структура БТС с рециклом: исходная (а) и вычислительная (б) схемы

входа в систему последовательно от элемента к элементу по мере прохождения потока.

Структура с рециклом (рис. 2.8, а). Состояние потока на входе в элемент I зависит от входного потока 0 и состояния потока 4 после отвлечения его от потока 3 за элементом II. Чтобы вести последовательный расчёт, начиная от элемента I, надо уже знать результат расчёта последующего элемента II, что невозможно. Общий подход к расчёту состояния БТС с рециклом следующий. Разрывают связи рецикла, как показано на рис. 2.8, б. Структура становится линейной с неизвестным состоянием X_{O_1} входного потока O_1 .

Априори задают какие-то параметры его состояния и проводят расчёт полученной линейной системы. Получают состояние потока 3 на выходе. В «разорванной» схеме состояния потоков X_{O_1} и Y_4 должны совпадать, т.е. $Y_4 = X_{O_1}$. Если это не так, то задают новые значения X_{O_1} , и расчёты повторяют до получения указанного совпадения. Стараются минимизировать разность между X_{O_1} и Y_4 , т.е. $\Delta = |X_{O_1} - Y_4| < \varepsilon$, где ε – точность расчёта. Минимизация функции $F = \Delta$ есть типовая задача оптимизации.

Если рециклов несколько, то разрывают их все и минимизируют функцию $F = \sum \Delta_i^2$, где индекс i соответствует нумерации рециклов.

Форма представления состояния БТС. Наиболее полная форма представления состояния БТС – сводная таблица, в левой части которой перечислены, а затем просуммированы все статьи прихода (пока-

затели входных потоков), а в правой части – все расходные статьи (показатели выходных потоков). Подобные таблицы приводят для отдельных элементов, подсистем и БТС в целом. Достоинствами представления балансов в форме таблиц являются полнота и точность представленного материала, удобство практической работы с данными. Таблицы – основная форма представления балансовых расчётов, входящая обязательно в техническую документацию производства.

2.5. АНАЛИЗ БИОТЕХНОЛОГИЧЕСКОЙ СИСТЕМЫ

Анализ заключается в получении сведений о состоянии БТС, показателях эффективности её функционирования, а также о влиянии на эти показатели структуры технологических связей, свойств и состояния элементов подсистем, условий их эксплуатации. Он осуществляется при разработке и проектировании нового биотехнологического производства, при эксплуатации действующего производства, для сравнения различных вариантов реализации процесса, при модернизации и реконструкции производства.

Первым шагом анализа БТС является определение её состояния, т.е. расчёт БТС. Зная изменение состава и количества потоков, энергетические расходы, можно провести и другие расчёты – эффективности использования сырья и энергии как технологических показателей, экономических показателей, некоторых социальных показателей, определяемых свойствами всех компонентов биотехнологического процесса, в том числе отходов производства. Эксплуатационные показатели определяются в основном из реакции системы на те или иные возмущения в процессе (изменение состава и количества сырья, энергетического обеспечения, состояния аппаратов, включая выход из строя некоторых фрагментов оборудования, а также воздействия на режимы отдельных аппаратов и узлов).

Свойства биотехнологической системы. БТС функционирует как единое целое, как совокупность взаимосвязанных элементов. Поэтому во-первых, биотехнологический процесс характеризуют рядом показателей функционирования БТС в целом и, во-вторых, могут появиться свойства системы, не присущие её элементам.

Зависимость режима одного аппарата (элемента) от режимов других. Свойство очевидное: изменение режима аппарата, т.е. изменение состояния потоков, выходящих из него и входящих в последующий, изменяет условия работы последнего. Изменение условий работы последующего аппарата сказывается двояко: влияет на эффективность осуществления в нём ТП и на эксплуатационные свойства аппарата.

Режим аппарата может зависеть от режимов не только в предыдущем, но и в последующем аппаратах. Это возможно, во-первых, при наличии рецикла, когда поток согласно схеме передаётся в предыдущий аппарат. Во-вторых, обратное относительно направления потока воздействие может иметь физическую причину. Чаще всего это связано с изменением давления, которое распространяется во всех направлениях, в том числе и в противоположном движению потоке. Если гидравлическое сопротивление какого-либо узла возрастает, то увеличивается давление на входе в этот узел (при поддержании расхода потока через него) и, следовательно, давление в предыдущем аппарате. Может возникнуть аварийная ситуация.

Взаимосвязь условий работы элементов в системе необходимо учитывать как в «штатном» режиме (при задаваемых условиях эксплуатации), так и в изменяющихся условиях эксплуатации.

Усовершенствование одного узла (элемента) улучшает эффективность БТС в целом за счёт выигрыша в другом узле системы. Это свойство вытекает из предыдущего и обусловлено взаимозависимостью режимов БТС. Усовершенствование одного элемента связано с определёнными затратами. Положим, в реакторе использован новый катализатор, увеличивающий селективность сложного биохимического превращения, точнее – уменьшающий выход некоторых побочных продуктов реакции, количество которых незначительно. На общем выходе полезного продукта это скажется мало. Новый, более дорогой катализатор требует дополнительных затрат в реакционном узле. Но в системе очистки основные затраты связаны с удалением именно малых количеств примесей. Уменьшение примесей приведёт к существенному уменьшению затрат на очистку от них и в конечном итоге – к уменьшению затрат в системе в целом.

Оптимальные режимы элемента одиночного и в системе могут различаться. Подтвердить это положение можно, если найдётся пример, опровергающий противоположное утверждение: «оптимальные режимы элемента одиночного и в системе совпадают». Рассмотрим БТС, состоящую из биореактора и системы разделения. Заданы производительность $\Pi = V_0 c_0 x$ и состав сырья – c_0 ; V_0 – количество необходимого сырья. Определим, при какой степени превращения x затраты Z будут минимальны. Они складываются из капитальных затрат и расходов на сырьё:

$$Z = K_1 C_{\text{кап}} + C_c V_0,$$

где $C_{\text{кап}}$ – стоимость капитальных вложений (на оборудование); K_1 – коэффициент окупаемости, показывающий долю стоимости оборудования, входящую в затраты на производство продукции за определённый

ное время, обычно за 1 год (затраты также рассчитывают на 1 год; полагая, например, срок службы оборудования 5 лет, получаем коэффициент $K_1 = 0,2$); C_c – стоимость единицы расхода сырья; $V_0 = \Pi/(c_0x)$.

Одиночный реактор. Затраты на него зависят от объёма. Для простоты рассматриваем вариант режима идеального смешения и реакции первого порядка. Из описания процесса в таком реакторе $x/\tau = k(1 - x)$ получим

$$v_p = V_0\tau = (V_0/k) [x/(1 - x)].$$

Полагаем, что стоимость реактора пропорциональна его объёму – $C_{\text{кап}} = \alpha v_p$. Подставляя выражение для V_0 и v_p в выражение для Z , получим формулу для расчёта затрат на процесс в одиночном реакторе:

$$Z_p = A/(1 - x) + B/x.$$

Коэффициенты A и B включают все постоянные – k , α , c_0 , V_0 , K_1 и др. Первое слагаемое – капитальные затраты на реактор Z_k – возрастает с увеличением степени превращения: чем больше x , тем больше должны быть объём реактора и затраты на него. Второе слагаемое – затраты на сырьё Z_c . Чем больше x , тем полнее используется сырьё и тем его нужно меньше. Зависимости Z_k , Z_c и Z_p показаны на рис. 2.9 сплошными линиями. Из анализа графиков следует, что существует некая степень превращения x , при которой достигаются минимальные затраты Z_p на единственный реактор.

Реактор в БТС. К затратам на реактор надо добавить затраты Z_d на систему разделения. Предположим, что они пропорциональны примесям к продукту, т.е. количеству непрореагировавшего сырья $V_0c_0(1 - x)$. Не

усложняя расчётную формулу, примем простейший вариант: $Z_d = D(1 - x)$, где D включает постоянные в расчёте Z_d . Суммарные затраты на БТС составят

$$Z_{\text{БТС}} = \frac{A}{1 - x} + \frac{B}{x} + D(1 - x).$$

К затратам на биореактор добавляются затраты на разделение – прямая пунктирная линия на рис. 2.9, и оптимальная степень превращения $x_{\text{БТС}}$ становится отличной от x_p . В данном случае $x_{\text{БТС}} > x_p$.

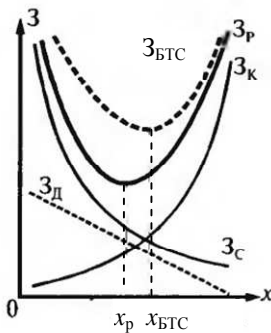


Рис. 2.9. К определению оптимальной степени превращения x_p для одиночного биореактора и биореактора, работающего в составе БТС ($x_{\text{БТС}}$)

- значительная часть метана сжигается для выработки тепла, необходимого для эндотермической реакции конверсии метана и обеспечения температурного режима процессов;
- часть образовавшегося водорода расходуется на очистку природного газа от сероводородсодержащих примесей и азотводородной смеси, CO, деструкцию высших углеводородов в природном газе до CH₄;
- конверсия метана из-за обратимости реакции протекает не полностью (потери).

Значения расходного коэффициента входят в технологический регламент производства. Напомним читателю, что расходный коэффициент определяют не только по сырью, но и по другим расходуемым компонентам – по энергии всех видов, по вспомогательным материалам.

Энергетическая и эксергетическая эффективность БТС

Тепловой коэффициент полезного действия. Состав энергетических потоков БТС показан на рис. 2.10.

Источниками энергии являются входящие потоки, несущие энергию топлива $Q_{\text{топл}}$, электроэнергию $Q_{\text{эл}}$, теплосырья $Q_{\text{сыр}}$ и вспомогательных материалов $Q_{\text{мат}}$, а также экзотермические реакции $Q_{\text{экс}}$. Их суммарное количество – потребление, или затраты энергии $Q_{\text{затр}}$. Из БТС энергия выводится с потоками продуктов – $Q_{\text{прод}}$ и отходов производства – $Q_{\text{отх}}$, с энергетическими потоками – $Q_{\text{эн}}$, затрачивается на проведение эндотермических реакций – $Q_{\text{энд}}$ и теряется естественным путём – $Q_{\text{пот}}$. Из выходящих энергетических потоков полезно используется $Q_{\text{эн}}$ в виде пара котлов-утилизаторов, нагреваемых в теплообменниках-утилизаторах внешних потоков (например, воды), энергии турбин и электрогенераторов. Исходя из сказанного, энергетическую эффективность БТС характеризуют тепловым коэффициентом полезного действия – тепловым КПД $\eta_{\text{тепл}}$:

$$\eta_{\text{тепл}} = \frac{\sum_i Q_{\text{п}, i} + \sum_k (\alpha_1 + \alpha_2 \eta_{\text{пз}}) Q_{\text{пз}, k}}{\sum_l B_l Q_{\text{н}} + \sum_p Q_{\text{экс}, p} + \sum_j Q_{\text{н}, j}},$$

где $\sum_i Q_{\text{п}, i}$ – тепло, полезно затраченное на технологические цели и определяемое из условий теоретического расхода для осуществления физических и химических процессов, происходящих в элементах БТС; α_1, α_2 – доли тепла вторичных энергоресурсов, генерированных в элементах ХТС (котлах-утилизаторах) и расходуемых вне и внутри системы соответственно; $Q_{\text{пз}, k}$ – тепло вторичных энергоресурсов, генерированное в k -м энергетическом элементе БТС; $\eta_{\text{пз}}$ – КПД преобразова-

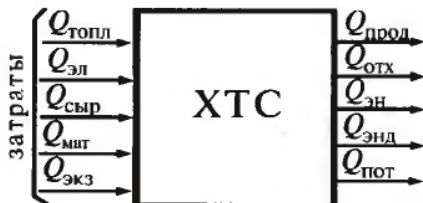


Рис. 2.10. Состав энергетических потоков в БТС

ния тепла вторичных энергоресурсов, используемых внутри БТС; B_l – расход топлива на технологические цели БТС в l -м элементе; Q_n – низшая теплота сгорания топлива; $Q_{экз, p}$ – теплота, выделяющаяся при экзотермических и физических превращениях в p -м элементе БТС; $Q_{п, j}$ – тепло, подведенное в j -й элемент БТС с энергоносителями (например, водяным паром различных параметров с ТЭЦ, горячей водой и др.).

В таком виде КПД не учтено следующее. Продукт обладает некоторым энергетическим потенциалом, который будет использован при его потреблении. В выражении для КПД потенциал продукта определен как затраты на его образование $(\sum_i Q_{п, i})$. Если продукт образуется в результате протекания экзотермической реакции, то энергетические затраты на него меньше, и $\eta_{тепл}$ уменьшается независимо от потребительской ценности продукта. Таким образом, тепловой КПД всегда хуже для экзотермической реакции, чем для эндотермической.

Эффективность организации процесса в биотехнологической системе

Для оценки эффективности организации процесса в БТС необходимо выяснить, насколько режим процесса в БТС приближается к теоретически возможному, и насколько рассматриваемая БТС оправдана экономически. Такой анализ в совокупности с оценками эффективности использования сырья и энергии является частью характеристики биотехнологического производства, включающей в себя многие показатели.

Предельная эффективность БТС. Теоретический расходный коэффициент по сырью определен из условия полного превращения исходного компонента в продукт. Особенности термодинамики и кинетики протекающих в БТС процессов могут ограничить полную конверсию исходных веществ в продукты. К ним относятся обратимость реакций и фазовые равновесия.

Для заданных условий процесса можно определить равновесную (теоретически возможную) конверсию сырья, равновесные концентрации компонентов в продуктах превращения и селективность биохимического процесса, а также равновесные концентрации компонентов в продуктах превращения и селективность биохимического процесса, а также равновесные концентрации компонентов в контактирующих фазах. Они и определяют предельные показатели биотехнологического процесса.

Предельная эффективность БТС – это эффективность системы, в которой при принятой совокупности условий технологического процесса достигаются максимально возможные (предельные) значения конверсии сырья и селективности, основанные на данной конкретной биохимической схеме превращения исходного сырья. Степень приближения к этим предельным показателям является характеристикой совершенства разрабатываемой БТС.

Эффективность БТС определяется не только технологическими, но и экономическими показателями. Основные экономические показатели – приведённые затраты, себестоимость продукта, прибыль, рентабельность.

Приведённые затраты состоят из полных текущих расходов на производство продукта и части капитальных затрат, т.е. это есть затраты на собственно производство.

Себестоимость продукта включает в себя приведённые затраты, оплату труда и дополнительные (непроизводственные) цеховые и общезаводские расходы, т.е. полные затраты предприятия на производство продукта.

Разность между ценой продукта (естественно, после его продажи) и себестоимостью есть *прибыль* – важнейший показатель эффективности производства.

Отношение прибыли к себестоимости – *рентабельность производства*.

Эти показатели являются категориями экономическими, они рассчитываются специалистами по экономике и используются в хозяйственной деятельности предприятия.

При разработке биотехнологического процесса используют, как правило, стоимостную его оценку в виде приведённых затрат (ПЗ). Она включает:

условно-переменные затраты $Z_{уп}$ – стоимость сырья, вспомогательных материалов и энергии;

условно-постоянные затраты $Z_{ук}$ – часть стоимости капитальных затрат а оборудование.

Таким образом,

$$ПЗ = З_{уп} + З_{ук}.$$

Экономическая оценка БТС используется для сопоставления альтернативных вариантов при разработке и анализе БТС.

Исходная информация включает:

1) структурную (или технологическую) схему БТС; на ней должна иметься спецификация на все аппараты и потоки; таблицу величин всех потоков, поступающих в БТС (сырьё, вспомогательные материалы, пар, вода, топливо, электроэнергия); материальный и тепловой баланс (расчёт) БТС;

2) стоимость каждого входящего потока (сырья, пара, материалов и т.п.), эти данные берутся из соответствующих справочников цен;

3) стоимость используемого оборудования (реакторов, колонн, теплообменников и т.д.).

Конечно, не имея готового проекта производства, трудно точно определить стоимость каждого аппарата. Поэтому прибегают к оценкам стоимости оборудования будущего производства на основе корреляционных зависимостей типа:

$$\text{стоимость оборудования} = \text{базовая стоимость} \times F_c,$$

где F_c – корректирующий фактор (поправка на материал, давление и т.д.).

Такие корреляции приходится получать из справочных данных по ценам аналогичного оборудования.

Стоимость потока (вещества, энергии) определяют исходя из его цены $\Pi_{п}$ и количества $G_{п}$, потребляемого в единицу времени (обычно за год):

$$З_{уп} = \Pi_{п} G_{п}.$$

Стоимость оборудования – единовременные затраты, они постепенно переносятся на стоимость продукции исходя из срока эксплуатации аппарата. Обычно этот коэффициент использования аппарата рассчитывают исходя из годовых затрат на производство продукции. В химической промышленности этот коэффициент E в среднем равен $E = 0,12 \dots 0,15$. В этом случае приведённая стоимость $З_{ук}$ аппарата равна $З_{ук} = E \Pi_{ап}$ и является частью капитальных затрат, входящей в приведённые затраты.

Суммарные затраты на все потоки есть условно-переменные затраты: $З_{уп} = \sum_i \Pi_{пi} G_{пi}$, здесь индекс относится к номеру потока. Суммарный расход на все аппараты есть капитальные вложения $\Pi_{кз} = \sum_j \Pi_{ап, j}$, а приведённая стоимость всех аппаратов

$Z_{\text{ук}} = \sum_j E_j C_{\text{ап}, j}$. Разделив приведённые затраты ПЗ на годовую производительность, получим ПЗ на единицу продукции.

Затраты в биотехнологическом производстве складываются в основном из трёх статей: сырьё, энергия всех видов, капитальные вложения. Расходы по этим статьям в значительной степени зависят от степени конверсии, селективности и интенсивности процессов. В свою очередь эти показатели связаны с основными термодинамическими и кинетическими закономерностями биохимических реакций и физических процессов, положенных в основу данной конкретной БТС.

2.6. РАЗРАБОТКА И ПРОЕКТИРОВАНИЕ АВТОМАТИЗИРОВАННЫХ BIOTECHNOLOGICAL ПРОЦЕССОВ И ПРОИЗВОДСТВ

Разработка (синтез) или проектирование БТС заключается в определении основных технологических операций и их последовательности, выборе аппаратов и установлении связей между ними, определении параметров технологических режимов отдельных аппаратов и системы в целом, обеспечивающих наилучшие условия функционирования БТС. Постановка задачи следующая.

Дано:

- вид и качество продукта, который необходимо производить;
- вид, состав и состояние исходных веществ (сырья);
- основные стадии биотехнологического процесса переработки сырья в продукты и их характеристики;
- возможные аппаратурные решения для осуществления биотехнологического процесса.

Определить:

- элементный состав БТС (аппараты, машины, устройства и пр.) для переработки сырья в продукты;
- структуру связей между элементами БТС, обеспечивающими последовательность стадий биотехнологического процесса и функционирование всей системы;
- режимы отдельных элементов и системы в целом, обеспечивающие необходимые показатели эффективности функционирования производства.

При такой формулировке предполагается, что какие-то основы биотехнологического процесса заданы на ранней стадии его проработки и их необходимо реализовать в некоей БТС. Конечно, возможна корректировка данных предварительной проработки процесса, придётся рассмотреть использование альтернативного сырья или источника

энергии, рассмотреть иные стадии процесса или типы аппаратурного оформления.

Совокупность трёх стадий – подготовки исходных данных, предварительного синтеза БТС и рабочего проектирования производства – называют разработкой биотехнологического процесса или разработкой технологии производства какого-либо продукта.

При анализе подходов к синтезу БТС сопоставим два подхода – научный (получение новых знаний) и инженерный (разработка нового производства) (рис. 2.11). По структуре своей они очень похожи, только мотивировки несколько различаются: в первом случае – любознательность, во втором – практическая необходимость. В обоих процессах движущим началом является стремление к прогрессу, развитию. Оба процесса включают аналогичные стадии, характерные для творчества: поиск на каждой стадии, возврат к предыдущим для нахождения оптимальных решений. Конечно, методическая и техническая база различны в разных объектах исследований, да и время на инженерную разработку ограничено, жёстко лимитировано. Но в обоих процессах требуется сочетание науки и искусства – как в открытии, так и в создании нового.

Основная цель инженерного процесса – создание производства жизненно важных продуктов (новых или уже существующих). Генери-



Рис. 2.11. Синтез БТС как творческий процесс

рование идей и последующая их реализация в производстве безусловно предусматривают не только полезность продукции и эффективность производства, но и обеспечение его экологической безопасности и комфортности в эксплуатации. По данным некоторых исследователей, успех реализации нового биотехнологического процесса на стадии предпроектных научных исследований составляет 1...3%, на стадии разработки – 10...25%, на стадии опытно-промышленной апробации – 40...60%. Добиться успеха в такой ситуации возможно, прорабатывая десятки тысяч (по оценкам экспертов) путей достижения конечной цели, т.е. разработка процесса – действительно искусство (вспомните первоначальный смысл слова «технологи»), процесс творческий.

Инженер-технолог начинает работу с рассмотрения альтернативных вариантов решений и представления технологических схем. Он прорабатывает аппаратные решения, дополняет разработку количественными расчётами, меняет состав и компоновку аппаратов и узлов, проводит экспертизу достигнутого. Творческий процесс совершенствования разработки продолжается до завершения работы. Знания фундаментальных основ и глубокое понимание происходящих процессов, методы системного анализа и математического моделирования сокращают процесс синтеза (разработки) БТС.

Схема основных этапов разработки промышленного биотехнологического процесса и производства представлена на рис. 2.12 [7].



Рис. 2.12. Основные этапы разработки биотехнологического процесса и производства

Эта схема отражает реальную структуру работ: указанные этапы в том или ином виде входят в план работ по созданию проекта производства и нередко являются основой планово-распорядительных документов. В схемы выделены разделы – предпроектные научно-исследовательские работы (НИР), опытно-конструкторские работы (ОКР), проектирование и ввод в эксплуатацию. В отраслевых научно-исследовательских и проектных фирмах такое планирование работ есть основа их организации и деятельности.

Современные вычислительная техника и информационные технологии, казалось бы, могут помочь из большого разнообразия типов аппаратного оформления (элементов) и связей выбрать такие, которые образуют БТС с необходимыми показателями. Но это нереально (по крайней мере в ближайшем будущем), так как система очень разнородна по составу оборудования (биореакторы, экстракторы, адсорберы, насосы и т.д.), варианты простого перебора и требований к системе невероятно много, многие показатели противоречивы.

Здесь используются некоторые наборы решений, обоснованных теорией био- и химико-технологических процессов, созданных на основе инженерных проработок, проверенных практикой. Среди них – максимальное использование сырьевых и энергетических ресурсов, минимизация отходов, эффективное использование оборудования. Поскольку эти требования всегда должны выполняться, назовём их *принципами создания БТС* [7].

Принцип полного использования сырьевых ресурсов. В большинстве биотехнологических и химических производств затраты на сырьё составляют значительную часть себестоимости продукции. Принцип полного использования сырьевых ресурсов направлен на максимальный выход целевого продукта, максимальное превращение сырья в полезные продукты. К методам достижения этой цели следует отнести следующие:

- 1) выбор схемы производства;
- 2) противоточный контакт фаз, обеспечивающий максимальную движущую силу процесса переноса;
- 3) фракционный рецикл при неполном превращении исходных реагентов, в системе разделения после реактора выделяют непрореагировавшие реагенты и возвращают их на переработку;
- 4) регенерация с рециклом, вспомогательный материал после использования регенерируют и возвращают в процесс;
- 5) утилизация отходов позволяет получить дополнительные полезные продукты из побочных продуктов реакции, практически всегда образующихся вместе с целевым продуктом (непрореагировавшие ис-

ходные вещества, дополнительные продукты реакции и примеси, имеющиеся в сырье);

б) альтернативное сырьё, использование невозобновляемого сырья заставляет искать не только способы его полного использования, но и другие сырьевые источники (использование отходов других производств, альтернативные источники первичного сырья);

7) комбинированные био- и химико-технологические процессы и комплексное использование сырья.

Принцип полного использования энергетических ресурсов направлен на минимизацию энергетических и тепловых затрат, на полное использование их в процессе.

К методам достижения этой цели следует отнести следующие:

1) регенерация тепла и энергии, в ряде случаев для проведения процесса (например, в биореакторе) необходимо нагреть поток, а после этого охладить его; можно входящий поток нагревать за счёт тепла выходящего потока;

2) утилизация тепла и энергии, тепло или энергия потока используется для выработки тепловых (пар, горячая вода), электрических и других энергетических ресурсов, применяемых не в самом производстве;

3) альтернативные источники энергии; традиционные источники энергии – пар, горячие теплоносители, электроэнергия, вырабатываемые при потреблении топливных ресурсов; перспективными являются легко возобновляемые и более дешёвые источники, такие, например, как тепло, вырабатываемое атомными котлами, солнечная энергия и др.;

4) вторичные энергетические ресурсы; энергетический потенциал продукции, отходов, дополнительных и промежуточных продуктов, образующихся в технологических агрегатах, который не используется в самом агрегате, но может быть частично или полностью использован для энергоснабжения других производств;

5) энерготехнологическая система.

Принцип минимизации отходов направлен на защиту окружающей среды; она связана с экологической обстановкой в регионе, где расположено производство. Реализация этого принципа в первую очередь связана с улучшением использования сырьевых и энергетических ресурсов и включает следующие специальные мероприятия:

1) санитарная очистка и обезвреживание выбросов; предполагает установку подсистем БТС, в которых вредные (токсичные) примеси удаляются из выбрасываемых потоков (санитарная очистка) или переводятся в нетоксичное состояние (обезвреживание);

2) замкнутый водооборот; вода в большом количестве используется как вспомогательный материал в биотехнологических производствах;

3) повышение надёжности оборудования и элементов БТС, позволяющее исключить потери в виде утечек и выбросов любых компонентов.

Принцип эффективного использования оборудования направлен на минимизацию капитальных затрат на технологическое оборудование путём создания условий протекания в нём процессов с максимальной возможной интенсивностью. К методам достижения этой цели следует отнести следующие:

1) для минимизации капитальных затрат следует выбирать технологический процесс, протекающий с большей скоростью;

2) оптимизация БТП – очевидный приём увеличения его интенсивности;

3) организация процесса в аппарате; почти всегда возможно один и тот же процесс провести разными способами: теплообмен и контакт фаз – в противотоке или прямотоке; гетерогенно-каталитическую реакцию – в неподвижном или движущемся слое катализатора; разделение жидкостей – ректификацией или дистилляцией и т.д.;

4) организация БТП в подсистеме БТС;

5) конструктивные решения аппаратов, позволяющие уменьшить объём его непроизводительных частей;

6) увеличение единичной мощности аппарата;

7) совмещение процессов.

Глава 3

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ БИОТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

3.1. МАТЕМАТИЧЕСКИЕ МОДЕЛИ БИОТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

Под биотехнологическим процессом или объектом понимается совокупность разнообразных технологических операций и физико-биохимических процессов (диффузия, тепло- и массообмен, биохимические превращения, теплопроводность, экстракция, адсорбция/десорбция и т.д.) с использованием того или иного биологического агента (микроорганизмов, изолированных клеток, ферментов или клеточных органелл), реализуемых в современной технической аппаратуре (автоматизированные машины и аппараты, биохимические реакторы, разделительные колонны, сушильные и холодильные агрегаты, котлы-утилизаторы и т.п.) и обеспечивающих целенаправленную переработку сырья и материалов в полупродукт или конечный продукт [3, 9].

Современные крупные БТП, реализуемые на нескольких взаимосвязанных технологических установках, часто называют производствами, а получаемые при этом конечные продукты характеризуются себестоимостью и товарной ценой. Далее мы будем использовать только термин «биотехнологический процесс» и/или объект.

Структурная схема типового биотехнологического процесса/объекта представлена на рис. 3.1.

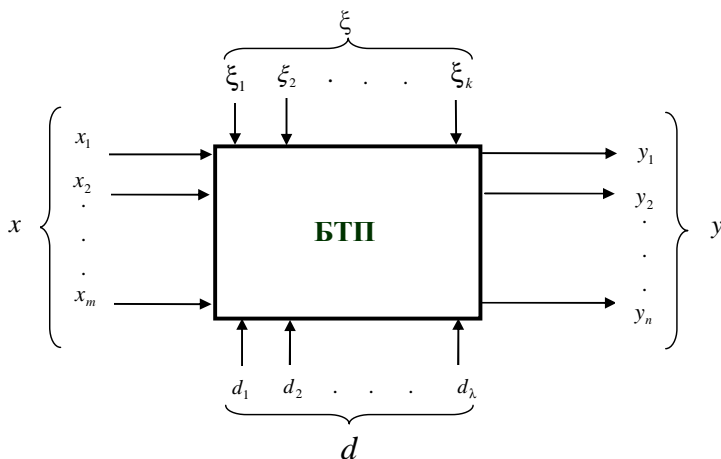


Рис. 3.1. Структурная схема БТП

На этой схеме $x(t)$, $y(t)$ – входные (независимые) и выходные (зависимые) вектор-функции, называемые далее координатами или переменными БТП; d – вектор конструктивных параметров; $\xi(t)$ – вектор-функция случайных помех, воздействующих на объект, и, следовательно, переменную $y(t)$.

Вернёмся к переменным x , y и помехе ξ . Будем далее полагать, что $x(t)$ и $y(t)$ измеримы с помощью измерительных устройств (датчиков, приборов, лабораторных анализов и т.д.), а сам БТП допускает проведение активных экспериментов или в крайнем случае наблюдений за изменениями входных/выходных переменных. Результаты таких экспериментов записываются так:

$$x^3(t), y^3(t) = \bar{y}(t) + \xi(t), t \in [0, T], \quad (3.1)$$

здесь сигнал $x^3(t)$ измерен точно; сигнал $y^3(t)$ состоит из точной реакции БТП $\bar{y}(t)$ на воздействие $x^3(t)$ и аддитивной (суммируемой) помехи $\xi(t)$, приведённой к выходу объекта, такой, что её среднее значение равно нулю, а дисперсия $\sigma_\xi^2 < \infty$. Условие $\sigma_\xi^2 < \infty$ означает, что помеха $\xi(t)$ не является физически реализуемым «белым шумом», а оценка дисперсии $\hat{\sigma}_\xi^2$ (приближенное значение σ_ξ^2 , найденное по конечному числу «измерений» $\xi(t)$) может быть косвенно определена.

Изложим возможность оценивания дисперсии σ_ξ^2 помехи $\xi(t)$, учитывающей «собственные» флуктуации переменной $y(t)$ и независимую нормальную случайную погрешность её измерения, прибором класса Кл и шкалой Ш. При таких допущениях справедлива формула

$$\hat{\sigma}_\xi^2 = \hat{\sigma}_B^2 + \hat{\sigma}_{\text{изм}}^2,$$

где $\hat{\sigma}_B^2$ и $\hat{\sigma}_{\text{изм}}^2$ – дисперсии воспроизводимости БТП и измерения $y(t)$, причём символ сверху означает оценки дисперсий, полученные по конечному числу измерений.

Для определения $\hat{\sigma}_B^2$ установим статический режим работы БТП, когда $x(t) = \text{const}$, а $y(t)$ будет изменяться только за счёт действия «собственной» помехи $\xi(t)$. Далее выполним ряд точных измерений $\xi(t_\gamma)$, $\gamma = \overline{1, d}$, используя при этом прецизионные средства контроля с $\hat{\sigma}_{\text{изм}}^2 \approx 0$. Вычислим несмещённые оценки $\hat{y} = \sum_{\gamma=1}^d y_\gamma / d$, $\hat{\sigma}_B^2 = \sum_{\gamma=1}^d (y_\gamma - \hat{y})^2 / (d-1)$, которые будут сходиться по вероятности к $M\{y\}$ и $\hat{\sigma}_B^2$ при $d \rightarrow \infty$.

Для определения $\hat{\sigma}_{\text{изм}}^2$ воспользуемся известными формулами $Kл = y_{\text{max}} \cdot 100\% / Ш$ и $y_{\text{max}} \approx 3\sigma_{\text{изм}}$, из которых следует $\sigma_{\text{изм}} \approx Kл \times Ш / (3 \cdot 100\%)$ и $\hat{\sigma}^2 \approx \sigma_{\text{изм}}^2$ (здесь принято, что с вероятностью 0,997 величина y_{max} равна $3\sigma_{\text{изм}}$ для случайной нормальной ошибки измерения, а y_{max} и шкала Ш имеют одинаковую физическую размерность).

Знание модели измерения и величины σ_{ξ}^2 используется при построении математической модели БТП и решении многих задач математического моделирования технологических процессов/объектов.

Физико-биохимическая модель технологического процесса

Приведённое сжатое описание БТП содержит минимальную информацию о характеристиках наблюдаемых входных/выходных переменных и не раскрывает внутренние свойства исследуемого объекта, необходимые для его содержательного математического описания. В таблице 3.1 приводится универсальный список общезначимых допущений/гипотез или аксиом о возможных режимах функционирования и основных свойствах БТП, который будем в дальнейшем использовать при построении математической модели объекта.

3.1. Физико-биохимическая модель БТП [9]

Режим работы, свойства БТП	
Статика	Динамика
Линейность	Нелинейность
Сосредоточенность координат	Распределённость координат
Стационарность	Нестационарность
Незначимые физико-биохимические явления	Значимые физико-биохимические явления
Диффузия	Диффузия
Теплообмен	Теплообмен
Теплопроводность	Теплопроводность
Массо-, теплообмен	Массо-, теплообмен
Биохимические превращения	Биохимические превращения
Экстракция	Экстракция
Адсорбция/десорбция	Адсорбция/десорбция
Испарение/конденсация и т.д.	Испарение/конденсация и т.д.

Покажем назначение и применимость каждого свойства/режима БТП.

Статика. Статический режим работы БТП имеет место при $x(t) \equiv \text{const}$, $y(t) \equiv \text{const}$, где символ \equiv означает тождественное равенство в каждый момент времени. Тогда $dx/dt \equiv 0$, $dy/dt \equiv 0$ и, следовательно, математическая модель статики будет включать в себя только конечные зависимости или производные по пространственным переменным.

Динамика. Этот режим возникает при изменении $x(t) = \text{var}$ и, следовательно, $y(t) = \text{var}$, поэтому скорости изменения $x(t)$ и $y(t)$ не равны нулю тождественно, и в уравнениях модели имеются производные по времени. Такая модель более информативна, чем модель статики, но одновременно более сложна.

Линейность БТП. Пусть дана некоторая функция $y = f(x)$, определённая на ограниченном множестве $X = \{x : |x| \leq N\}$, N – действительное число. Эта функция будет линейной на множестве X , если для любых x_1, x_2 из X и любого числа c справедливы условия аддитивности (суммируемости) и однородности:

$$f(x_1 + x_2) = f(x_1) + f(x_2), \quad f(cx) = cf(x). \quad (3.2)$$

С помощью этих условий можно проверять линейность на множестве X конечных, дифференциальных и интегральных уравнений. Анализ линейности БТП можно проводить экспериментально.

Нелинейность БТП. Свойство нелинейности характерно для подавляющего большинства БТП и проявляется, как правило, при больших амплитудах входных/выходных переменных. Учёт нелинейности в модели позволяет получить более информативную модель (по сравнению с линейной моделью), однако при этом заметно усложняется структурная идентификация уравнений динамики/статики, затрудняется численное решение многих типовых задач моделирования.

Сосредоточенность координат БТП. Это допущение/гипотеза означает, что во всех точках l трёхмерного БТП координаты $x(t) \equiv \text{const}$, $y(t) \equiv \text{const}$, а векторные частные производные $\partial x(l)/\partial l \equiv 0$, $\partial y(l)/\partial l \equiv 0$, модель будет содержать в себе только конечные уравнения (статика) или обыкновенные дифференциальные уравнения с производной dy/dt (динамика).

В реальных БТП всегда имеется распределённость координат $y(l)$, $x(l)$, но если $|\text{grad}_y(l)|$ невелик или слабо влияет на выходную переменную $y(t)$, то ТП часто рассматривается как сосредоточенный.

Распределённость координат БТП. Все реальные БТП обладают распределёнными в пространстве координатами $x(l) = \text{var}$, $y(l) = \text{var}$ и тогда соответствующая модель статики/динамики будет содержать в себе некоторые частные производные $\partial x(l)/\partial l \neq 0$, $\partial y(l)/\partial l \neq 0$. Использование уравнений с частными производными вызывает значительные

сложности и поэтому их (уравнения) часто стремятся аппроксимировать обыкновенными дифференциальными уравнениями с запаздывающим аргументом или одномерными обыкновенными дифференциальными уравнениями с производной dy/dl (в задачах расчёта статических режимов).

Стационарность БТП. Технологический процесс/объект называется стационарным, если основные характеристики/свойства неизменны во времени. Понятно, что стационарность является идеализированным свойством БТП (все объекты «стареют», при этом ухудшаются их показатели качества и характеристики), но на небольших отрезках времени соответствующие изменения качества/свойств объекта могут быть малыми или несущественными для его использования по назначению. В таком случае можно говорить о стационарности БТП на определённом отрезке времени и описывать его (БТП) поведение автономными дифференциальными уравнениями типа $\dot{y}(t) = f(y(t), x(t), a)$, где a – параметр, правая часть которых не зависит явно от времени t . Использование автономных моделей с постоянными параметрами a существенно упрощает построение дифференциальных уравнений и их использование в типовых задачах математического моделирования.

Нестационарность БТП. Свойства/характеристики нестационарного ТП изменяются во времени, причём скорость этих изменений может зависеть от режима работы объекта, в частности, от вектор-функций $x(t)$, $y(t)$; конструкции и материала аппаратуры; действия внешних факторов типа помехи ξ , температуры окружающей среды и т.д.

Для описания поведения нестационарного объекта применяют три типа моделей:

$$\text{I. } \dot{y}(t) = f_1(x(t), y(t), a, t).$$

$$\text{II. } \dot{y}(t) = f_{II}(x(t), y(t), a(t)).$$

$$\text{III. } \dot{y}(t) = f_{III}(x(t), y(t), a(t)), \dot{a}(t) = \varphi(x(t), y(t), a(t)).$$

Уравнение I широко используется в математических задачах, где не требуется анализ физико-биохимических явлений и установление причин изменения структуры f_1 во времени. Нестационарное уравнение II содержит в правой части f_{II} изменяющийся параметр $a(t)$, но причины изменения последнего остаются неизвестными и физически непонятными. Модели типа II применимы для описания периодических объектов, для которых можно найти временную программу $a(t)$, $0 \leq t \leq T$, по накопленным данным $x^3(t)$, $y^3(t)$, $t \in [0, T]$. Модель III содержит в себе уравнение II и дополнительное дифференциальное уравнение $\dot{a}(t) = \varphi(x(t), y(t), a(t))$, описывающее скорость изменения физико-

химического параметра $a(x^2(t), y^2(t)) = a(t)$ в зависимости от режима работы БТП, т.е. от $x(t), y(t)$. Система уравнений III является стационарной или автономной (не зависящей от времени t явно), но при этом возросла размерность системы (число неизвестных функций) $y(t)$ и $a(t)$, потребовалось дополнительное уравнение $\dot{a}(t) = \varphi(x(t), y(t), a(t))$ по трудноизмеримым данным $a^3(t)$ (примерами параметра $a(t)$ служат активность катализатора, толщина слоя накипи на теплопередающей поверхности и т.п.).

Незначимость физико-биохимических явлений. Если объявить все физико-биохимические явления незначимыми и учесть остальные допущения, то будем иметь линейную стационарную модель статики с сосредоточенными координатами. При неучёте всех физико-биохимических явлений получим линейную нестационарную модель динамики с сосредоточенными координатами типа

$$a_2(t)\ddot{y}(t) + a_1(t)\dot{y}(t) + y(t) = f(x) \text{ и т.д.}$$

Значимые физико-биохимические явления. Все или некоторые значимые/важные физико-биохимические явления могут или должны учитываться при исследовании БТП и в первую очередь построении математической модели. При этом следует иметь в виду, что каждое физико-биохимическое явление описывается соответствующими уравнениями, рассматриваемыми в учебных дисциплинах естественно-научного цикла (физика, химия, физическая химия, теоретическая механика, термодинамика и др.).

Назовём *математической моделью* (ММ) исследуемого объекта функциональную (операторную) зависимость векторов y, x, a , описывающую с определённой точностью поведение БТП во времени и пространстве:

$$F(x, y, a) = 0, \tag{3.3}$$

где $F(\cdot)$ – оператор.

Параметр a отличается принципиально от координат x, y тем, что он неизмерим, но его можно вычислить по известным экспериментальным данным x^2, y^2 , т.е. в векторной форме: $a = a(x^2, y^2)$.

Если параметр a входит линейно в $F(\cdot)$, то он называется *коэффициентом*.

Конечные уравнения

В эти уравнения не входят операции дифференцирования/интегрирования. К таким уравнениям относятся алгебраические и трансцендентные равенства или уравнения.

Алгебраические уравнения конструируются из чисел, символов-букв и символов алгебраического действия: сложения, вычитания, ум-

ножения, деления, возведения в степень, извлечения корня, где всюду n натуральное число, т.е. $n = 1, 2, 3, \dots$. Алгебраическое уравнение с одним неизвестным x имеет такой вид: $a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0$, где коэффициент $a_n \neq 0$. Аналитическое решение (нахождение n «корней») этого уравнения возможно при $n \leq 4$, при $n > 4$ «корни» или «нули» определяются численно. Определённая трудность – неустойчивость численного решения – возникает в случае так называемой *плохой обусловленности* системы. Нелинейные алгебраические уравнения и особенно их системы решают преимущественно численными методами, сталкиваясь при этом с явлениями неустойчивости и множественности решений или их отсутствия.

Алгебраические линейные и нелинейные уравнения находят широкое применение при построении математических моделей статистики и даже динамики (например, при исследовании характеристических уравнений динамических систем), планировании и оптимизации функционирования технологических объектов, проектировании и конструировании технологической аппаратуры, обработке экспериментальных данных.

Трансцендентные уравнения. Эти уравнения включают в себя трансцендентные (не алгебраические) функции: логарифмические – $\lg x$; тригонометрические – $\sin x$ или $\operatorname{tg} x$; показательные – $\exp\{-x\}$ или c^x ; обратные тригонометрические – $\arcsin x$. Примеры трансцендентных уравнений: $\sin x + \lg x = x$, $2^x - \lg x = \arccos x$. Для узкого ряда простейших уравнений известны аналитические решения, но для большинства трансцендентных уравнений их решения находят приближёнными или численными методами. Трансцендентные уравнения находят ограниченное применение в задачах математического моделирования и обработке результатов эксперимента.

Обыкновенные дифференциальные уравнения

В структуру дифференциальных уравнений входит неизвестная функция y и её производные любого порядка по времени t и пространственным переменным l_1, l_2, l_3 . Если уравнение зависит только от y и её производных по t , то оно называется обыкновенным дифференциальным уравнением, если оно зависит от $y(t, l_i)$ и производных по t, l_i , $i = 1, 2, 3, \dots$, то имеем дифференциальное уравнение с частными производными.

Простейшее обыкновенное дифференциальное уравнение первого порядка можно записать в следующем виде:

$$F(y(t), \dot{y}(t), a, t) = 0, \quad \text{или} \quad \dot{y}(t) = f(y(t), a, t). \quad (3.4)$$

здесь первая запись есть общее уравнение, а вторая – разрешённое (относительно производной $\dot{y}(t)$); a – параметр.

Решение уравнения (3.4) обращает уравнение в тождество, т.е. левая часть уравнения будет равна правой в любой момент времени t . Число таких решений оказывается бесконечно большим, поэтому для получения единственного решения $y(t)$ надо задавать дополнительное условие типа $y(t=0) = y(0)$, называемое начальным. По аналогии дифференциальные уравнения высшего порядка n записываются в таком виде:

$$F(y, \dot{y}, \dots, y^{(n-1)}, y^{(n)}, a) = 0, \quad (3.4a)$$

или в формате системы из n уравнений:

$$\dot{y}_1 = y_2, \dot{y}_2 = y_3, \dots, \dot{y}_{n-1} = y_n, \dot{y}_n = f_n(y_1, y_2, \dots, y_n, a, t), \quad (3.4б)$$

где $y_1 = y$.

Для единственного решения $y(t)$ следует задать n начальных условий $y_i(0)$ $i = \overline{1, n}$ или конечных $y_i(T)$ $i = \overline{1, n}$ условий (чисел), где $[0, T]$ – интервал определения $y(t)$. Решение уравнения (3.4a) или системы (3.4б) с соответствующими начальными условиями существует и единственно, если непрерывно дифференцируемая (гладкая) функция $F(\cdot)$ удовлетворяет условию Липшица $F(\cdot) < L$, где L – трудно определяемая константа Липшица, т.е. $F(\cdot)$ – ограниченная функция, не обращающаяся в бесконечность при любых значениях аргументов.

Дифференциальное уравнение называется *автономным*, если независимая переменная t не входит явно в функцию $F(\cdot)$ или $f_n(\cdot)$:

$$F(y, \dot{y}, \dots, y^{(n-1)}, y^{(n)}, a) = 0, \quad \text{или}$$

$$\dot{y}_1 = y_{i+1}, \dot{y}_n = f_n(y_1, y_2, \dots, y_n, a), \quad i = \overline{1, n-1}. \quad (3.4в)$$

Автономные уравнения описывают стационарные физико-биохимические или технологические процессы, свойства (механизма), которые не изменяются во времени. Многие стационарные динамические явления или процессы характеризуются несколькими переменными $y_i(t)$, $i = 1, 2, \dots, n$ от одной и той же независимой переменной t . Тогда такие процессы описываются системой автономных дифференциальных уравнений:

$$\dot{y}_i(t) = f_i(y_1(t), y_2(t), \dots, y_n(t), a), \quad i = \overline{1, n} \quad (3.5)$$

с начальными условиями $y_i(t=0) = y_i(0)$, $i = \overline{1, n}$.

Система (3.5) называется *нормальной* системой, а задача определения её решения при заданных начальных условиях и известном параметре a получила название задачи Коши (в честь великого французского математика О. Коши, доказавшего в 20 – 30 гг. XIX в. теоремы существования и единственности решения систем (3.5) и (3.4б), в последствии названные условием Липшица).

Среди большого числа обыкновенных дифференциальных уравнений особое место занимают *линейные дифференциальные уравнения*, удовлетворяющие условиям аддитивности и однородности и обладающие единственной структурой:

$$a_n y^{(n)}(t) + a_{n-1} y^{(n-1)}(t) + \dots + a_1 y^{(1)}(t) + a_0 y(t) = x(t), \quad t \in [0, T], \quad (3.6)$$

где $a_n \neq 0$, $a = \{a_0, a_1, \dots, a_n\}$ – вектор числовых параметров; $x(t)$ – известная на отрезке $[0, T]$ функция.

Для решения уравнения (3.6) должны быть заданы начальные условия $y(0)$, $y^{(1)}(0)$, ..., $y^{(n)}(0)$, числовой вектор a и натуральное число n , называемое порядком уравнения.

Линейные однородные (при $x(t) = 0$) уравнения типа (3.6) имеют аналитическое решение следующего вида:

$$y(t) = \sum_{i=1}^n c_i \exp\{\lambda_i t\}, \quad (3.7)$$

где λ_i , $i = \overline{1, n}$ – некрратные (разные) корни характеристического уравнения

$$a_n \lambda^n + a_{n-1} \lambda^{n-1} + \dots + a_1 \lambda^1 + a_0 = 0; \quad (3.7a)$$

c_i , $i = \overline{1, n}$ – постоянные интегрирования, вычисляемые через λ_i и начальные условия $y(0)$, $y^{(1)}(0)$, ..., $y^{(n)}(0)$. Решение уравнения (3.7) устойчиво по t (т.е. не обращается в бесконечность при $t \rightarrow \infty$, если все λ_i отрицательны, кроме того, оно непрерывно зависит от начальных условий, коэффициентов a_0, a_1, \dots, a_n и параметров $\lambda_1, \lambda_2, \dots, \lambda_n$).

Из анализа формулы (3.7) следует, что решение $y(t)$ линейного дифференциального уравнения – нелинейно по параметрам или корням $\lambda_1, \lambda_2, \dots, \lambda_n$ и, следовательно, по коэффициентам a_0, a_1, \dots, a_n уравнения (3.6) (нелинейная зависимость $y(t)$ от a_i , $i = \overline{0, n}$ существенно затрудняет параметрическую идентификацию модели).

Необходимо указать, что фактическое использование зависимостей (3.7) для построения точного (не приближённого!) аналитического решения уравнения (3.6) возможно только при $n \leq 3 - 4$, так как корни или нули многочлена (3.7a) можно вычислять по формулам при $n \leq 4$.

Среди линейных дифференциальных уравнений типа (3.6) встречаются так называемые «жесткие» уравнения, для которых характерно существенное различие максимального λ^+ и минимального λ^- корней, что приводит к значительным изменениям $|\dot{y}(t)|$ на разных отрезках времени t . Степень «жесткости» линейного уравнения (3.6) характеризуется критерием Годта $T = \lambda^+/\lambda^-$, оценивание которого затруднено сложностью определения минимального корня λ^- , близкого к нулю.

Жесткость линейных дифференциальных уравнений заметно затрудняет (замедляет) их численное решение в области погранслоя, что вынуждает использовать вычислительные алгоритмы с трудоёмкими неявными схемами интегрирования.

В задачах математического моделирования БТП линейные обыкновенные дифференциальные уравнения (как автономные, так и особенно неавтономные) чаще всего решают методами Рунге–Кутты 3-го, 4-го порядка, реже методом Адамса с явными схемами интегрирования (для решения жестких уравнений используют неявные схемы интегрирования). Решение нелинейных дифференциальных уравнений, удовлетворяющих условию Липшица, также выполняется численными методами Рунге–Кутты, Адамса. При решении краевых задач для обыкновенных дифференциальных уравнений, когда известна часть начальных условий $y(0)$ и часть конечных условий $y(T)$, используют метод «стрельбы» или «прострела», требующий неоднократного интегрирования уравнений, или метод прогонки (прямая и обратная прогонка). Отметим возможность неустойчивости в методе «стрельбы».

Последняя четверть XX в. ознаменовалась массовым созданием пакетов прикладных программ, предназначенных для математического моделирования, оптимизации и проектирования физико-химических и технологических процессов и объектов. Такие пакеты впоследствии стали именоваться пакетами моделирующих программ. В состав каждого пакета входят: программные реализации типовых математических моделей основных и вспомогательных объектов исследуемого процесса; программные средства решения конечных/дифференциальных уравнений статики/динамики объектов; программы решения конечномерных задач оптимизации; программные средства вычисления физико-химических параметров уравнений математических моделей; базы данных/знаний о свойствах используемых веществ, конструкционных/режимных параметрах объектов; программные средства представления результатов моделирования и оптимизации биотехнологического процесса/объекта.

Ключевые элементы пакетов моделирующих программ – математические модели типовых объектов практически всегда нуждаются в

серьёзном анализе адекватности по экспериментальным данным, полученным на функционирующем биотехнологическом объекте. Последующее обеспечение адекватности модели основано на уточнении физико-химических параметров уравнений типовой модели, для чего также потребуются экспериментальные данные, полученные на специализированных лабораторных установках. Указанные операции весьма трудоёмки, но отказ от их выполнения не гарантирует получения достаточно точных и надёжных результатов математического моделирования и оптимизации биотехнологического процесса/объекта.

3.2. ПОСТРОЕНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

После обзорного знакомства с классическими конечными и дифференциальными уравнениями вернёмся к задаче построения математической модели (ММ) биотехнологического процесса.

Искомую ММ, как и раньше, будем обозначать так:

$$F(x, y, a) = 0 \quad (3.8)$$

и полагать, что она включает в себя те или иные конечные и дифференциальные уравнения в векторной форме:

$$\dot{y} = f(x, y, a); \quad (3.8a)$$

$$f(x, y, a) = 0, \quad (3.8б)$$

здесь вектор a пока неизвестен и в дальнейшем подлежит определению по экспериментальным данным $x^3, y^3 = \bar{y} + \xi$. Неизвестна пока и структура $F(x, y, a)$ (или, точнее, $f(x, y, a)$) уравнений ММ, которую придётся выбирать, опираясь на гипотетические знания, аккумулированные в физико-биохимической модели БТП.

Итак, построение ММ биотехнологического процесса заключается в определении/выборе подходящих в некотором смысле структур уравнений типа (3.8a), (3.8б) и идентификации их параметров a^* по данным $x^3, y^3 = \bar{y} + \xi$, таких, что решения этих уравнений с известными a^* будут в достаточной мере близки к \bar{y} , свободным от помехи ξ .

Качество ещё не построенной ММ будем характеризовать следующими критериями:

- 1) трудоёмкость построения ММ;
- 2) удобство/простота использования модели;
- 3) адекватность/точность ММ.

Введённые критерии качества ММ будут полезны при принятии субъективных решений на различных этапах построения модели.

Обобщённая процедура построения абстрактной математической модели $F(x, y, a) = 0$ включает пять этапов.

1. Выбор структуры $F(x, y, a)$ модели.

2. Получение экспериментальных данных x^3, y^3 об исследуемом биотехнологическом процессе/объекте.

3. Определение вектора параметров $a^* = a(x^3, y^3)$ математической модели (или параметрический синтез ММ).

4. Анализ адекватности математической модели $F(x, y, a^*) = 0$ биотехнологическому процессу/объекту по экспериментальным данным \tilde{x}, \tilde{y} независимых экспериментов, т.е. отличных от экспериментальных данных, по которым проводилась параметрическая идентификация.

5. Обеспечение адекватности математической модели $F(x, y, a^*) = 0$.

Выбор структуры $F(x, y, a) = 0$ математической модели. Данный этап имеет и другое название: структурная идентификация математической модели $F(x, y, a) = 0$. Соответствующая задача структурной идентификации ММ заключается в определении структуры $F(x, y, a)$, такой, что будущая математическая модель $F(x, y, a) = 0$ будет описывать поведение исследуемого биотехнологического процесса с допустимой или желаемой погрешностью $\delta_{ж}$.

Можно предложить эвристический алгоритм выбора подходящей структуры $F(x, y, a)$ и самой математической модели с желаемой погрешностью $\delta_{ж}$. Этот алгоритм базируется на принципе «от простого – к сложному» и понятии сложности структуры $F(x, y, a)$. Сконструируем ряд приемлемых структур $F_1(x, y, a), F_2(x, y, a), \dots, F_i(x, y, a), \dots, F_m(x, y, a)$, таких, что в определённом смысле каждая i -я структура сложнее $(i + 1)$ -й. В качестве количественного показателя сложности структуры можно принять, например, порядок обыкновенной производной функции y , размерность k вектора $a = \{a_1, a_2, \dots, a_k\}$, число частных производных и т.д.

Затем проведём на исследуемом биотехнологическом процессе/объекте нужный эксперимент, состоящий из N опытов, и получим данные x^3, y^3 , где $x^3 = \{x_j^3\}, y^3 = \{y_j^3\}, j = \overline{1, N}$.

Далее возможен следующий алгоритм.

Шаг 1. Возьмём самую простую (минимально сложную) структуру $F_1(x, y, a)$, где a – неизвестный вектор-параметр, построим на базе $F(\cdot)$ векторное уравнение $F_1(x, y, a) = 0$ и, задавшись произвольным параметром \tilde{a} , найдём решение $y_1 = y_1(x, \tilde{a})$.

Шаг 2. Решим некоторым регулярным методом задачу параметрической идентификации с использованием экспериментальных дан-

ных x^3, y^3 и решения $y_1 = y_1(x, \tilde{a})$, в результате чего найдём параметр $a_1^* = a(x^3, y^3, F_1)$, существенно зависящий от структуры $F(\cdot)$.

Шаг 3. Подставим параметр a_1^* в уравнение $F_1(x^3, y, a_1^*) = 0$ и найдём решение $y_1(x^3, a_1^*)$, где x^3 – неиспользованные ранее (на шаге 2) данные.

Шаг 4. Вычислим «невязку» (рассогласование) между экспериментальными данными y^3 (не использованными на шаге 2) и расчётными $y_1(x^3, a_1^*)$:

$$\|y^3 - y_1(x^3, a_1^*)\|^2 = \rho_1.$$

Шаг 5. Сравним ρ_1 с $\delta_{ж}$: если $\rho_1 < \delta_{ж}$, то задача структурной и попутно параметрической идентификации решены, т.е. построена подходящая ММ, удовлетворяющая требованиям пользователя; в противном случае, т.е. если $\rho_1 \geq \delta_{ж}$, то возврат к шагу 1 с использованием более сложной структуры $F_2(\cdot)$, и т.д., до тех пор пока для некоторой структуры $F_\lambda(\cdot)$, $\lambda \leq m$ не выполнится условие $\rho_\lambda < \delta_{ж}$.

Предложенный эвристический алгоритм весьма трудоёмок и отнюдь не гарантирует нахождение подходящей структуры и самой ММ по конечному списку структур $F_\lambda(\cdot)$, $\lambda = \overline{1, m}$. Кроме того, результативность алгоритма существенно зависит от понятия «сложность структуры» и требования существования и единственности решения a_i^* , $i = \overline{1, k}$ задачи параметрической идентификации.

Вернёмся к исходной задаче выбора структуры ММ и преобразуем её в более «узкую» и более определённую задачу *эвристического выбора* подходящей структуры $F(x, y, a)$, наиболее полно учитывающей наши знания о качестве будущей модели. В качестве объектов эвристического выбора $F(x, y, a)$ можно рекомендовать следующие структуры: формальная, аналитическая, комбинированная.

Формальную структуру формирует сам исследователь исходя из требований наибольшей простоты построения и использования ММ, а также удобства обеспечения желаемой точности/погрешности модели. Если для решения задачи моделирования нужна ММ типа «вход–выход» (по терминологии кибернетики – «чёрный ящик»), не учитывающая внутренние физико-биохимические явления в БТП, то формальная структура состоит из тех или иных известных конечных (обычно алгебраических) и обыкновенных дифференциальных уравнений:

$$y = a_0 + a_1x, \quad y = a_0 + a_1x + a_2x^2 + \dots + a_kx^k,$$

$$\sum_{i=1}^k a_i y^{(i)} + a_0 y = x, \quad k = 1, 2, 3, \dots, \quad y(t) = \int_0^t a(t-\tau)x(\tau)d\tau,$$

где $a(t-\tau)$ – ядро уравнения Винера–Хопфа; коэффициенты a_i , $i = \overline{0, k}$, и ядро $a(t-\tau)$ – неизвестные пока величины/функции; параметр k чаще всего задаёт исследователь, базируясь на соображении «чем меньше k , тем проще строить математическую модель».

Формальная структура малоинформативна, она не содержит каких-либо непосредственных знаний о конструкции и принципе функционирования БТП, режиме работы, происходящих в биотехнологическом объекте физико-биохимических явлений и т.д. ММ на базе формальной структуры применяется в задачах анализа/синтеза систем автоматизации, оптимизации БТП с экстремальными статическими характеристиками, среднесрочного планирования/распределения нагрузок биотехнологических объектов, обработки и агрегирования экспериментальной информации и т.д. Такая модель неприменима для задач конструирования/проектирования БТП, проектных и конструкторских расчётов оборудования, оптимального конструирования или управления БТП.

Аналитическая структура $F(x, y, a) = 0$ в том случае, когда исследователь исходя из целевого назначения технологического объекта выбирает из физико-биохимической модели нужный режим функционирования и те или иные свойства, в том числе обязательно физико-биохимические явления (см. табл. 3.1).

Непосредственное формирование аналитической структуры $F(x, y, a)$ и соответствующей аналитической модели $F(x, y, a) = 0$ (где a – символ параметра) заключается в использовании известных уравнений, описывающих выбранные физико-биохимические явления, и составлении/выводе дифференциальных (реже – конечных) уравнений баланса субстанции (массы, энергии, импульса) с учётом конструкций технологического оборудования (в форме граничных условий) и свойств перерабатываемых веществ. Получаемая аналитическая структура $F(x, y, a)$ включает в себя не только алгебраические операции, но и обыкновенные/частные производные, операторы Лапласа, Пуассона и др., а соответствующая аналитическая ММ состоит из всевозможных нелинейных/линейных обыкновенных дифференциальных уравнений, конечных соотношений, параболических и реже – эллиптических/гиперболических уравнений. Укажем, что во все эти уравнения входят неизвестные коэффициенты/параметры, имеющие, как правило, физическую или биохимическую трактовку, вдобавок зависящие иногда от переменной y и всегда нуждающиеся в определении по экспериментальным данным x^2, y^2 .

Аналитическая структура содержит много информации о конструкции и принципах работы БТП, характеристиках, происходящих в нём физико-биохимических явлений, и в целом обо всём, что включил исследователь в оператор $F(x, y, a)$. Это обстоятельство позволяет использовать аналитические ММ для решения разнообразных задач конструирования (оптимального конструирования/проектирования) технологической аппаратуры, оптимизации и оптимального управления БТП, проектных и конструкторских расчётов, планирования и автоматизации биотехнологического производства.

Комбинированная структура $F(x, y, a)$ содержит в себе математические элементы из формальной и аналитической структур. При формировании комбинированной структуры исследователь стремится упростить/ускорить процедуру построения модели и поэтому принимает те или иные допущения: статика или динамика, линейность или нелинейность, сосредоточенность координат, стационарность БТП; но одновременно он должен учитывать и целевое назначение будущей модели путём введения в структуру $F(x, y, a)$ уравнений значимых физико-биохимических явлений (табл. 3.1). Таким образом, комбинированная структура и соответствующая ей модель $F(x, y, a) = 0$ содержат в себе «формальные» равенства, формулы с абстрактными параметрами a и математические зависимости/уравнения с физически трактуемыми параметрами \bar{a} , характеризующими значимые физико-биохимические явления.

Использование комбинированной структуры $F(x, y, a)$ позволяет по сравнению с аналитической структурой уменьшить трудоёмкость построения ММ, упростить процедуру обеспечения адекватности модели, но одновременно сузить область её применимости.

Известные способы получения экспериментальных данных x^3, y^3 можно разделить по месту проведения опытов на две группы: производственные/технологические эксперименты и лабораторные эксперименты.

Производственный/технологический эксперимент проводится на функционирующем (работающем) БТП с использованием активного или пассивного способа получения нужной информации x^3, y^3 . Выбор того или иного способа осуществляется исследователем с учётом степени безопасности и организованности/стабильности БТП, необходимой длительности наблюдения и бюджета времени на проведение опытов. Отметим, что получаемые в производственных условиях данные $y^3(t)$ зависят от случайной неизмеримой помехи ξ и влияния неизвестных исследователю факторов.

Лабораторный эксперимент проводится активным способом на специализированных лабораторных стендах, в каждом из которых имеет место один, а хуже – два физико-биохимических процесса (типа

диффузия, теплопроводность, биохимическая реакция, массообмен и т.п.) из числа значимых явлений, указанных в табл. 3.1 физико-биохимической модели БТП. Обозначим через x_p^3, y_p^3 – результаты исследования моноявления на каждом p -м лабораторном стенде, $p = 1, 2, \dots, M$. Независимость каждого p -го явления от других физико-биохимических процессов обеспечивается конструкцией лабораторного стенда и методикой экспериментирования. В лабораторных условиях удаётся, как правило, устранить или уменьшить случайную помеху и погрешность измерения $x(t), y(t)$ за счёт использования прецизионных датчиков и приборов. Вследствие этого получаемые результаты опытов $x_p^3, y_p^3, p = \overline{1, M}$ оказываются независимыми от помехи ξ и влияния других неизвестных факторов/явлений, т.е. «чистыми», и представляют определённую научную ценность, так как могут использоваться для конструирования/проектирования различных БТП, в которых имеют место аналогичные явления.

Итак, результатом трудоёмкого активного/пассивного или производственного/лабораторного эксперимента являются экспериментально-числовые данные x^3, y^3 или функциональные данные $x^3(t), y^3(t), t \in [0, T]$, используемые в дальнейшем для решения задач параметрической идентификации и обеспечения адекватности ММ.

Внимание! В этих задачах используются независимые экспериментальные данные, поэтому разобьём полученную выборку x^3, y^3 на две примерно равные группы с объёмами N_1 и $N_2, N_1 + N_2 = N$ и обозначим их так: x^3, y^3 и \tilde{x}, \tilde{y} . Данные $x_i^3, y_i^3, i = \overline{1, N_1}$ будем использовать в задаче параметрической идентификации ММ, а $\tilde{x}_i, \tilde{y}_i, i = \overline{1, N_2}$ – в задачах анализа и обеспечения адекватности модели.

Более подробно и доказательно методы построения ММ технологических процессов изложены в книге под редакцией В. С. Балакирева [9].

3.3. ПАРАМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ МАТЕМАТИЧЕСКОЙ МОДЕЛИ

Конкретизируем наши знания о биотехнологическом процессе/объекте. Нам известны экспериментальные данные x^3, y^3 , характеризующие статический или динамический режим работы БТП. В первом случае имеем числовые векторы $x_i^3, y_i^3, i = \overline{1, N_1}$, во втором случае имеем вектор-функции $x_i^3(t), y_i^3(t), i = \overline{1, N_1}, t \in [0, T]$, T – длительность опыта. Напомним, что x^3 измерен точно, а $y^3 = \bar{y} + \xi$, где \bar{y} –

точная реакция БТП на воздействие x^3 , ξ – случайная помеха с нулевым средним и конечной дисперсией σ^2 .

Кроме того, известна обобщённая математическая модель $F(x, y, a) = 0$, которую в дальнейшем будем представлять в форме обыкновенного дифференциального уравнения (для описания динамики БТП) или конечного уравнения (для описания статики):

$$\dot{y} = f(x(t), y(t), a); \quad (3.8в)$$

$$f(x, y, a) = 0, \quad (3.8г)$$

где y и x – векторы n и m переменных; параметр $a = \{a_1, a_2, \dots, a_j, \dots, a_k\}$, k – известное натуральное число, $k < N_1$, или иногда $k = N_1$; компоненты вектора a ограничены сверху и снизу; вектор-функция $f(\cdot)$ удовлетворяет условию Липшица и решение $y(t, a) = y(t, a, x(t), y(0))$ уравнения (3.8в) существует и единственно.

Аппроксимационная параметрическая идентификация ММ. Подставим в дифференциальное уравнение (3.8в) известную вектор-функцию $x^3(t)$ и найдём его численное решение $y(t, a) = y(t, a, x^3(t), y^3(0))$ при начальных условиях $y^3(0)$. Сформулируем задачу параметрической идентификации уравнения (3.8в): найти параметр a^* , такой, что решение $y(t, a^*)$ дифференциального уравнения (3.8в) будет максимально близко в среднем на интервале времени $[0, T]$ к функции $\bar{y}(t)$, т.е.

$$\int_0^T (y(t, a^*) - \bar{y}(t))^2 dt \approx 0.$$

Однако функция $\bar{y}(t)$, являющаяся реакцией БТП на воздействие $x^3(t)$ при отсутствии помехи $\xi(t)$, неизвестна, поэтому мы вынуждены в дальнейшем использовать экспериментально полученную аддитивную смесь $y^3(t) = \bar{y}(t) + \xi(t)$. Запишем задачу параметрической идентификации в следующей форме: найти $a^* \in \Lambda$, такое, что

$$\Phi(a^*) = \int_0^T (y(t, a^*) - y^3(t))^2 dt = \min_{a \in \Lambda} \Phi(a), \quad (3.9)$$

здесь a принадлежит некоторому открытому (незамкнутому) ограниченному и выпуклому множеству Λ , задаваемому неравенствами $a_i^- < a < a_i^+$, где a_i^- и a_i^+ – нижняя и верхняя грани Λ , известные из физических соображений для каждого конкретного БТП.

Задача параметрической идентификации дифференциального уравнения (3.8в) сводится к известной аргументной оптимизационной задаче нахождения параметра a^* , доставляющей безусловный минимум функции многих переменных $\Phi(a)$. Аргументная оптимизационная

задача (3.9) непрерывной гладкой функции $\Phi(a)$ всегда разрешима при замкнутом множестве Λ (что возможно при известных a_i^-, a_i^+) и устойчива относительно малых возмущений $x^3(t), y^3(0)$ и ошибок расчёта. Однако общее число решений a^* может быть разным:

- 1) если $\Phi(a)$ невыпукла, то число точек a^* произвольное;
- 2) при выпуклой $\Phi(a)$ может быть сколь угодно много или одна точка a^* ;
- 3) если $\Phi(a)$ строго выпукла, то точка a^* единственна.

Численное исследование степени выпуклости функции многих переменных ($k \geq 2 \dots 3$), заданной алгоритмом вычисления, связано со значительными трудностями и фактически неприменимо в задачах математического моделирования.

Параметрическая идентификация векторного уравнения статистики. Пусть задано уравнение (3.8г) вида $f(x, y, a) = 0$, где $x = \{x_\gamma\}$, $\gamma = \overline{1, m}$, $y = \{y_\rho\}$, $\rho = \overline{1, n}$, $a = \{a_j\}$, $j = \overline{1, k}$ – числовые векторы с известными размерностями m, n, k и известны экспериментальные данные $x_i^3, y_i^3 = \bar{y} + \xi_i$, $i = \overline{1, N_1}$, i – номер опыта, причём $N_1 > k$ (и даже $N_1 \gg k$), а векторы y_i^3 независимые (линейно независимые). Требуется найти параметр $a^* \in \Lambda$, такой, чтобы решение $y_i(x_i^3, a^*)$ уравнения $f(x^3, y, a) = 0$ было максимально близко в среднем вектору $\bar{y}_i, i = \overline{1, N_1}$.

За неимением значений \bar{y}_i сформируем функцию $\Phi(a)k$ – переменных $a = \{a_1, a_2, \dots, a_k\}$:

$$\Phi(a) = \frac{1}{N_1} \sum_{i=1}^{N_1} (y_i(x_i^3, a) - \bar{y}_i)^2 = \frac{1}{N_1} \sum_{i=1}^{N_1} \sum_{\rho=1}^n (y_{\rho i}(x_i^3, a) - y_{\rho i}^3)^2, \quad (3.10)$$

где $y_i(x_i^3, a), i = \overline{1, N_1}$ – решение конечного векторного уравнения (3.8г) с $x = x_i^3$ и параметром $a \in \Lambda$ и, применяя какие-либо итерационные процедуры нулевого/первого порядка, будем искать $\min \Phi$ и, соответственно, одну или несколько точек $a^* \in \Lambda$, на которых, возможно, выполняется приближенное условие $y_i(x_i^3, a^*) \approx \bar{y}$. В данной задаче требуется на каждой итерации $k + 1$ раз численно решать N_1 -мерную систему конечных уравнений.

Более подробно и доказательно вопросы структурной и параметрической идентификации ММ изложены в книге под редакцией В. С. Балакирева [9].

3.4. АНАЛИЗ АДЕКВАТНОСТИ МОДЕЛИ БИОТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА

На данном этапе построения ММ нам известно векторное уравнение $\dot{y}(t) = f(x, y, a^*)$, параметр a^* которой определён по экспериментальным векторным данным $x_i^3(t), y_i^3(t), i = \overline{1, N_1}$, а также линейно независимые и неиспользованные ранее результаты измерений $\tilde{x}_i(t), \tilde{y}_i(t), i = \overline{1, N_2}$ (здесь i – номер опыта или измерения, $N_1 + N_2 = N$ – общее число измерений). Кроме того, известна желаемая/допустимая погрешность $\delta_{\text{ж}}$ будущей модели.

Для оценивания адекватности ММ необходимо определить её фактическую среднюю квадратическую ошибку (СКО) по экспериментальным данным $\tilde{x}_i(t), \tilde{y}_i(t), i = \overline{1, N_2}$. Пусть БТП описывается векторным дифференциальным уравнением (3.8в): $\dot{y}(t) = f(x, y, a^*)$, решение которого при $x_i(t) = \tilde{x}_i(t), y_i(0) = \tilde{y}_i(0)$ имеет вид $y_i(t, \tilde{x}_i, a^*)$, $t \in [0, T], i = \overline{1, N_2}$. Далее вычислим СКО математической модели динамики:

$$\delta_{\text{ММ}}^{\text{д}} = \left\| \tilde{y} - y(t, \tilde{x}, a^*) \right\|_{L_2} = \left[\sum_{i=1}^{N_2} \frac{1}{2} \int_0^T (\tilde{y}_i(t) - y_i(t, \tilde{x}, a^*))^2 dt / N_2 \right]^{\frac{1}{2}}, \quad (3.11)$$

размерность которой должна совпадать с физической размерностью вектор-координаты $y(t)$ (если компоненты вектора $y(t)$ имеют разные физические размерности, то следует преобразовать их в безразмерные функции путём деления каждой компоненты $y_p(t)$ на базовую величину $y_p^{\text{б}}$, $p = \overline{1, n}$, n – число компонент $y(t)$; L_2 – нормированное пространство).

СКО модели статики $\delta_{\text{ММ}}^{\text{с}}$ вычисляется по той же формуле, что и $\delta_{\text{ММ}}^{\text{д}}$, однако в последней снимается символ интегрирования и используется числовой вектор $y_i(t, \tilde{x}, a^*), i = \overline{1, N_2}$.

Математическая модель динамики $\dot{y}(t) = f(x, y, a^*)$ считается адекватной БТП на базе экспериментальных данных \tilde{x}, \tilde{y} , если

$$\delta_{\text{ММ}}^{\text{д}} = \left\| \tilde{y} - y(t, \tilde{x}, a^*) \right\|_{L_2} < \delta_{\text{ж}}. \quad (3.12)$$

Аналогично: математическая модель статики $f(x, y, a^*) = 0$ будет адекватной БТП на базе данных \tilde{x} , \tilde{y} , если

$$\delta_{MM}^c = \left\| \tilde{y} - y(t, \tilde{x}, a^*) \right\|_E < \delta_{ж}. \quad (3.12a)$$

Отметим, что $\delta_{MM}^d \neq \delta_{MM}^c$ в общем случае.

Более подробно и доказательно вопросы анализа и обеспечения адекватности математической модели изложены в книге под редакцией В. С. Балакирева [9].

3.5. ТИПОВЫЕ ЗАДАЧИ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ ПРОЕКТИРУЕМОГО БИОТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА

Миссия проектируемого биотехнологического процесса заключается в рациональном или экономически оптимальном и безопасном преобразовании определённого количества исходных веществ с использованием биологического агента (микроорганизмов, изолированных клеток, ферментов или клеточных органелл) в конечные продукты заданного качества. Проектирование БТП традиционно осуществляется в несколько последовательно реализуемых этапов, основными из которых являются предпроектные исследования, разработка технического задания и непосредственное техническое проектирование объекта. Далее будем считать, что техническое проектирование завершено и нам известна следующая информация о процессе, аккумулированная в технологическом регламенте:

- предварительное технико-экономическое обоснование способа получения конечного продукта;
- характеристики и свойства исходных (сырьевых) веществ, промежуточных и конечных продуктов;
- физико-биохимические механизмы/основы и принципиальная технологическая схема БТП;
- перечень технического оборудования/аппаратуры и их основные характеристики (технические паспорта);
- исходные данные для составления уравнений баланса (массы, энергии, импульса) по всему БТП;
- рекомендуемые режимы функционирования биотехнологического процесса и критерии их оптимальности;
- разработанная аналитическая модель динамики и статики стационарного технологического процесса;

- рекомендации для проектирования систем автоматизации и контроля (перечни координат, подлежащих автоматическому регулированию/контролю);
- желаемые показатели надёжности/восстановления систем автоматизации и технической аппаратуры;
- методы и параметры очистки/ликвидации вредных отходов и выбросов.

Вместе с тем в технологическом регламенте часто отсутствуют: рекомендации по пускам и плановым/аварийным остановкам БТП; результаты оптимизации и расчёта надёжности автоматизированного оборудования и систем автоматизации/стабилизации; не показана возможность оптимизации и определения технико-экономических показателей в режиме реального времени и др. Большинство указанных «неделок», однако, может быть выполнено при наличии математической модели БТП с помощью математического моделирования и вычислительного эксперимента на ЭВМ.

Математическое описание автоматизированного биотехнологического процесса

Изменения координат автоматизированного БТП описываются векторным дифференциальным уравнением

$$\dot{y}(t) = f(x(t), y(t), u(t), a) \quad (3.13a)$$

где $x(t) = \{x_j(t), j = \overline{1, m}\}$ – вектор входных (независимых) координат; $y(t) = \{y_i(t), i = \overline{1, n}\}$ – вектор выходных (зависимых) координат; $u(t) = \{u_j(t), j = \overline{1, n}\}$ – вектор управляющих координат (далее управлений); $a = \{a_\rho, \rho = \overline{1, r}\}$ – вектор-параметр; $f(\cdot) = \{f_i(\cdot), i = \overline{1, n}\}$ – вектор-функция, непрерывно дифференцируемая по всем аргументам x, y, u, a .

Уравнение (3.13a) разрешимо, если определены $x(t), a, u(t)$, а при задании начального условия $y(0) = \{y_i(0), i = \overline{1, n}\}$ оно будет иметь единственное решение $y(t, x, u, a)$. Векторы $x(t), y(0)$ определяются условиями задач моделирования, вектор-параметр a находят при построении математической модели, а управление $u(t)$ формирует автоматическая система регулирования (АСР), которая в нашем случае осуществляет стабилизацию выходных переменных объекта, описываемых вектор-функцией $y(t)$. Математическая модель линейной АСР такова:

$$u(t) = c_0 \varepsilon(t) + c_1 \int_0^t \varepsilon(t) dt + c_2 \dot{\varepsilon}(t); \quad \varepsilon(t) = y(t) - y_{\text{здн}}(t), \quad (3.13\text{б})$$

где c_0, c_1, c_2 – параметры настройки регуляторов; $y_{\text{здн}}$ – вектор независимых заданий регуляторам АСР, $y_{\text{здн}}(t) \equiv y_{\text{здн}} = \text{const}$ для задачи стабилизации $y(t)$.

Система уравнений (3.13а), (3.13б) замкнута, разрешима и имеет единственное решение $y(t, x(t), y(0), y_{\text{здн}})$. Запишем систему (3.13а), (3.13б) в нормальной форме:

$$\begin{aligned} \dot{y}_i(t) &= f_i(x_j(t), y_i(t), u_i(t), a_\rho), \quad i = \overline{1, n}, \quad j = \overline{1, m}, \quad \rho = \overline{1, r}; \\ u_i(t) &= c_{0i} \varepsilon_i(t) + c_{1i} \int_0^t \varepsilon_i(t) dt + c_{2i} \dot{\varepsilon}_i(t); \\ \varepsilon_i(t) &= y_i(t) - y_{\text{здн},i}(t), \quad i = \overline{1, n}, \end{aligned} \quad (3.13)$$

и назовем выражение (3.13) математической моделью автоматизированного биотехнологического процесса/объекта.

Рассмотрим более подробно аргументы правых частей системы (3.13), опуская по возможности термин «вектор». Входная координата $x(t)$ принадлежит ограниченному открытому множеству X кусочно-непрерывных функций и состоит в общем случае из трёх компонент: $x_n(t)$, $\tilde{x}(t)$ и $\xi(t)$. Функция $x_n(t)$, называемая далее «нагрузкой», включает в свой состав переменные $x_j(t), j = 1, 2, 3, \dots$, определяющие производительность ТП и качество получаемого продукта; нагрузка $x_n(t)$ измерима и управляема оператором или автоматизированной системой управления ТП. Функция $\tilde{x}(t)$ оказывает слабое влияние на производительность ТП, чаще всего она линейно зависима от $x_n(t)$ и тогда считается измеримой (иногда $\tilde{x}(t)$ называют возмущением режима работы ТП). Третья функция $\xi(t)$ неизмерима и случайна; гипотетично будем рассматривать её как низкочастотный стационарный случайный процесс с постоянным средним значением и конечной дисперсией. Помеха $\xi(t)$, так же как и $\tilde{x}(t)$, оказывает слабое влияние (возмущение) на $y(t)$, которое частично компенсируется управлением $u(t)$.

Управление $u(t)$ принадлежит ограниченному замкнутому множеству U кусочно-непрерывных функций; функция $u(t)$ измерима, может достигать граничных значений множества U ; $u(t)$ явно зависит от $y(t)$, $y_{\text{здн}}$ и вектора настроек регулятора $c = \{c_0, c_1, c_2\}$, которые устанавливаются оператором ТП такими, что замкнутая система (3.13) была устойчива и $\varepsilon(t) = y(t) - y_{\text{здн}} \rightarrow 0$ при $t \rightarrow \infty$.

Вектор-параметр $a = \{a_\rho, \rho = 1, 2, 3, \dots\}$ принадлежит ограниченному замкнутому множеству Λ из числового евклидова пространства (граничные значения a_ρ обычно неизвестны и поэтому не используются в задачах моделирования). Отдельные компоненты a_ρ характеризуют те или иные физико-химические явления и технологические/технические решения (операции), включённые в структуру математической модели (3.13). Эти параметры определяют по лабораторным экспериментальным данным или заимствуют из различных литературных источников, что не гарантирует их точность и надёжность (чаще всего a зависит от $y(t)$ или $x(t)$, но далее принимаем $a_\rho = \text{const}$, $\rho = 1, 2, 3, \dots$).

Функция $y(t)$ принадлежит ограниченному открытому множеству Y из нормированного пространства $L_2(0, T)$ функций, интегрируемых с квадратом, однако она дифференцируема всюду (т.е. «гладкая»); $y(t)$ измерима, причём величина или норма функции такова:

$$\|y\|_{L_2} = \left(T^{-1} \int_0^T y^2(t) dt \right)^{1/2},$$

где $[0, T]$ – отрезок наблюдения $y(t)$.

При наличии помехи $\xi(t)$ в $x(t)$ функция $y(t)$ может быть искажена аддитивным случайным процессом, соизмеримым с погрешностью вычисления и измерения выходной координаты ТП.

Вернёмся к анализу вектор-функции $f(x(t), y(t), u(t), a) = f(\cdot)$. Обычно $f(\cdot)$ содержит несущественные нелинейные зависимости, которые многократно дифференцируемы по переменной y и, следовательно, могут разлагаться в ряд Тейлора, что позволяет впоследствии проводить линейризацию «в малом» уравнений (3.13). При этом надо помнить, что любая линейризация означает снижение точности линейного дифференциального уравнения. Основная слабость исходного уравнения (3.13а) – низкая точность, связанная с его происхождением. Это уравнение построено аналитическим методом, базирующимся на фундаментальных законах сохранения субстанции (массы, энергии, импульса), учёте ряда физико-биохимических явлений и технологических операций, осуществляющихся в известных конструктивных форматах. При выводе уравнений баланса субстанции всегда принимается ряд допущений и гипотез об учёте/неучёте тех или иных физико-биохимических явлений, направленных на упрощение структуры функции $f(\cdot)$ и определения параметров a_ρ , $\rho = 1, 2, 3, \dots$, что приводит к заметному снижению точности уравнения (3.13а). Анализ адекватности системы (3.13а) невозможен из-за отсутствия экспериментальных данных о динамике проектируемого БТП. По экспертным оценкам относительная средняя квадратическая ошибка описания динамики

функционирующего БТП аналитическим уравнением (3.13а) не ниже 10...15%. Эту оценку погрешности следует помнить при численном решении системы (3.13) и решении прикладных задач математического моделирования проектируемого БТП.

Рассмотрим далее вопросы нахождения единственного решения системы (3.13а), полагая в ней для простоты анализа постоянными векторы x , u , a . Известно, что для единственности решения $y(t)$, $0 \leq t \leq T$, исходной системы $\dot{y}(t) = f(x(t), y(t), u(t), a)$ необходимо знание вектора начальных условий $y(0) = \{y_i(0), i = \overline{1, n}\}$, значения $y_i(0)$ которого определяются экспертами при постановке задач математического моделирования. Однако в ряде случаев (чаще всего в задачах синтеза) задаются n конечных (краевых) условий $y_i(T)$, $i = \overline{1, n}$, или K начальных $y_i(0)$, $i = \overline{1, K}$, $K < n$, и $n - K$ краевых условий $y_i(T)$, $i = \overline{K+1, n}$, что заметно затрудняет численное решение системы (3.13а). Покажем некоторые способы преобразования таких краевых задач к исходной системе (3.13а) с известными начальными условиями $y_i(0)$, $i = \overline{1, n}$.

Пусть заданы только краевые условия $y_i(T)$, $i = \overline{1, n}$. Преобразуем систему (3.13а) к такому виду, что её решение $y(\beta)$ начиналось в нулевой момент «обратного» времени $\beta = 0$ и заканчивалось при $\beta = T$. Примем для этого $t = T - \beta$ и после подстановки t в уравнение (3.13а) получим «обращённое» уравнение $\dot{y}(\beta) = -f(x, y(T - \beta), u, a)$ с начальным условием $y(\beta = 0)$ (или $y_i(\beta = 0)$, $i = \overline{1, n}$). Интегрировать такое уравнение надо осторожно и с повышенной точностью (так, например, для ли- неаризованных исходных уравнений «обращённые» равенства будут неустойчивыми из-за смены знака их собственных чисел).

Рассмотрим более сложный случай, когда задано одно краевое условие, например $y_1(T)$ и $n - 1$ условий $y_i(0)$, $i = \overline{2, n}$. Зададим произвольное условие $y_1(0)$ и, проинтегрировав на отрезке $[0, T]$ уравнение (3.13а) с начальными условиями $y_1(0)$, $y_i(0)$, $i = \overline{2, n}$, получим вектор- решение $y(t) = \{y_1(t, y_1(0)), y_i(t, y_i(0)), i = \overline{2, n}\}$. Далее найдём значение $y_1^*(0)$, такое, что функция $\Phi(y_1^*(0)) = \min_{y_1(0)} |y_1(T) - y_1(T, y_1(0))|^2$ при связи в форме равенства (3.13а). Данная оптимизационная аргументная задача разрешима, но при невыпуклой/выпуклой функции

$\Phi(y_1(0))$ число решений $y_1^*(0)$ может быть произвольным: поиск значения $y_1^*(0)$, обеспечивающего выполнение условия $\Phi(y_1^*(0)) \approx 0$, может быть выполнен методом сканирования.

Если задано K , $2 \leq K \leq n - 1$, краевых $y_i(T)$, $i = \overline{1, K}$, и $n - K$ начальных $y_i(0)$, $i = \overline{K + 1, n}$, условий, то соответствующая аргументная задача поиска недостающих условий $y_i(0)$, $i = \overline{1, K}$, будет такой: найти вектор $y_n^*(0) = \{y_i^*(0), i = \overline{1, K}\}$, такой, что функция K переменных

$$\Phi(y_n^*(0)) = \min_{y_i(0)} \sum_{i=1}^K |y_i(T) - y_i(T, y_i(0))|^2 \approx 0$$

при связи в форме равенства (3.13а). Здесь $y_i(T, y_i(0))$, $i = \overline{1, K}$ – значения при $t = T$ решения уравнения (3.13а) с начальными условиями $y_i(0)$, $i = \overline{1, n}$. Данная задача разрешима, но из-за невыпуклости $\Phi(y_n(0))$ (и, возможно, её «овражности») имеет сколь угодно много (множество) стационарных точек, что затрудняет применение градиентных методов поиска $y_n^*(0)$, обеспечивающих близость $\Phi(y_n^*(0))$ к нулю. Если имеется линейная версия уравнения (3.13а), то с помощью известного метода «прогонки» можно найти для неё недостающие начальные условия $y_i(0)$, $i = \overline{1, K}$, которые затем использовать как «хорошие» начальные приближения для градиентного поиска начальных условий $y_i^*(0)$ для нелинейной системы (3.13а), что заметно уменьшит трудоёмкость аргументной задачи.

Непосредственное решение уравнения (3.13а) с начальными условиями $y_i(0)$, $i = \overline{1, n}$, осуществляется известными методами Эйлера, Рунге–Кутты 4-го порядка и иногда Адамса, программные реализации которых имеются во всех математических пакетах прикладных программ (ППП). При этом процесс решения можно разделить на два этапа: на первом этапе интегрирование уравнения (3.13а) осуществляют быстрым, но не очень точным явным методом Эйлера или более точным модифицированным методом Эйлера, подбирая экспериментально схему, шаг h и отрезок $[0, T]$ интегрирования, а также погрешность счета δ . Если при постоянном δ наблюдается резкое уменьшение h (или при фиксированном h существенно возрастает δ), то уравнение (3.13а) жёсткое и надо использовать неявную схему интегрирования, что значительно увеличивает затраты процессорного времени. На вто-

ром этапе при решении конкретных задач моделирования следует использовать в основном явный метод Рунге–Кутты 4-го порядка с постоянным шагом h и переменной погрешностью δ , меньшей допустимой ошибкой счета ε (для интегрирования жёстких уравнений надо использовать неявный метод Рунге–Кутты 4-го порядка).

Анализ динамических свойств проектируемого процесса

Динамические свойства функционирующего БТП традиционно оценивают по экспериментальным переходным процессам, возникающим в технологическом объекте при резком изменении нагрузки, проведении операций пуска/останова, случайных полных/частичных отказах технологического оборудования и/или автоматических систем регулирования (стабилизации). Знание основных характеристик экспериментальных переходных процессов позволяет повышать ремонтно-пригодность БТП, оптимизировать его технико-экономические показатели (в частности, себестоимость и КПД), улучшать эффективность работы АСР [9].

На этапе проектирования такие характеристики динамики можно приближённо определять при имитационном моделировании переходных процессов. Для анализа переходных процессов используем векторное нелинейное уравнение динамики (3.13):

$$\dot{y}(t) = f(x_n(t), y(t), u(t), a); u(t) = c_0(y(t) - y_{\text{здн}}),$$

здесь нагрузка $x_n(t)$ принимает минимальное x_n^- , оптимальное (рабочее) x_n^p и максимальное x_n^+ значения; управление $u(t)$ также принимает значения u^-, u^p, u^+ ; выходная координата $y(t)$ в установившемся состоянии объекта принимает установившиеся (статические) значения y^-, y^p, y^+ ; аналогично вектор $y_{\text{здн}} = \{y_{\text{здн}}^-, y_{\text{здн}}^p, y_{\text{здн}}^+\}$; c_0 – вектор-параметр П-регуляторов, обеспечивающих быструю неколебательную стабилизацию $y(t)$ на уровне $y_{\text{здн}}$.

Выполним математическое моделирование «малых» (по амплитуде) переходных процессов $y(t)$ в проектируемом процессе, обусловленных действиями многочисленных АСР. Пусть БТП находится в статическом оптимальном/рабочем состоянии, когда $y(0) = y^p = \text{const}$, $u(0) = u^p = \text{const}$. Нанесём в момент $t = 0$ малое ступенчатое возмущение Δx_n , осуществим численное интегрирование уравнения (3.13) с параметрами $x_n + \Delta x_n$, $y(0) = y^p$, a , u^p на отрезке времени $[0, T_i]$, где T_i – время «затухания» переходного процесса i -й координаты y_i , $i = 1, n$ (напомним, что y_{T_i} определяется из условия $\dot{y}(T_i) \approx 0$). Кривую раз-

гона $y_i(t) - y^p$ по каналу $x_n \rightarrow y_i$ разделим на Δx_n и назовём единичной переходной функцией $h_i(t)$, $i = \overline{1, n}$ (при $\Delta x_n = (5 \dots 10)\%$ от x_n канал $x_n \rightarrow y_i$ обычно линеен в малом).

Функции $h_i(t)$, $i = \overline{1, n}$, чаще всего нужны для выбора наиболее эффективного канала. Для этого определим размерные чувствительности $\Psi^i(t, \Delta x_n) = \partial y_i(t) / \partial \Delta x_n(t)$, $i = \overline{1, n}$, затем найдём логарифмические (безразмерные) функции чувствительности $\Phi^i(t, \Delta x_n) = \Psi^i(t, \Delta x_n) \times \Delta x_n / y_i(t)$, $i = \overline{1, n}$, и, наконец, вычислим величины этих чувствительностей или нормы $\|\Phi^i(t, \Delta x_n)\|_{L_2} = \left(T_i^{-1} \int_0^{T_i} (\Phi^i(t, \Delta x_n))^2 dt \right)^{1/2}$, $i = \overline{1, n}$.

Здесь верхний индекс означает номер канала, а Δx_n в аргументах Ψ и Φ показывает переменную, по которой дифференцируют $y_i(t)$. Канал с наибольшей нормой считается самым чувствительным или эффективным (отметим, что для строго линейных каналов вместо норм можно использовать коэффициенты усиления $k_i(t) = h_i(T_i)$, $i = \overline{1, n}$).

Аналогично находятся кривые разгона $h_i(t, \Delta u_j)$ по каналам $u_j \rightarrow y_i$, $j = \overline{1, n}$, $i = \overline{1, n}$. В этом случае уравнение (3.13) интегрируется с параметрами $x_n = \text{const}$, $y(0) = y^p = \text{const}$, $u(0) = u^p = \text{const}$; малое возмущение Δu_j наносится по каналу $u_j \rightarrow y_i$, $j = \overline{1, n}$, $i = \overline{1, n}$; затем регистрируются функции $h_i(t, \Delta u_j)$ на отрезке $[0, T_{ij}]$, $i = \overline{1, n}$ и вычисляются функции чувствительности $\Psi^i(t, \Delta u_j) = \partial y_i(t) / \partial u_j$, $\Phi^i(t, \Delta u_j) = \Psi^i(t, \Delta u_j) / y_i(t)$ и их нормы $\|\Phi^i(t, \Delta u_j)\|_{L_2} = \left(T_{ij}^{-1} \int_0^{T_{ij}} (\Phi^i(t, \Delta u_j))^2 dt \right)^{1/2}$, $i = \overline{1, n}$, $j = \overline{1, n}$ (если размерность n вектора $u(t)$ велика, то можно использовать меньшее число каналов, например, такие, у которых диапазон изменения u_j наибольший и «статика» будет малой).

При численном определении функции $h_i(t, \Delta u_j)$ по каналу $u_j \rightarrow y_i$ полезно одновременно регистрировать и другие кривые разгона $u_\lambda(t)$ или $h_\lambda(t, \Delta u_j)$ по каналам $u_j \rightarrow y_\lambda$, где $\lambda \neq i$, $\lambda = \overline{1, n}$. Если функция $h_\lambda(t, \Delta u_j)$ мала относительно $h_i(t, \Delta u_j)$ (в смысле $\|h_\lambda\| < \|h_i\|$ или $\|h_\lambda\| \ll \|h_i\|$), то канал $u_j \rightarrow y_i$ автономный (независимый), иначе канал $u_j \rightarrow y_\lambda$ и переходную функцию $h_\lambda(t, \Delta u_j)$ следует учитывать при анализе системы автоматизации БТП.

Ранее для линейных или линеаризованных в малом статических характеристик получены при $\Delta u_j = 1$ единичные переходные функции $h_i(t, \Delta u_j)$, не равные нулю при $i = j$ и пренебрежимо малые (в смысле $\|h_i\| \approx 0$) или равные нулю при $i \neq j$, $j = \overline{1, n}$. Переобозначим первые функции так: $h_{ii}(t, 1) = h_{ii}(t)$, а соответствующие им каналы $u_j \rightarrow y_i$ назовём автономными. Если для некоторых каналов возмущение Δu_j влияет на y_i и y_{i+1} , то имеют место две переходные функции: $h_{ii}(t)$ и $h_{i+1,i}(t, \Delta u_j) = h_{i+1,i}(t)$, а соответствующие каналы $u_i \rightarrow y_i$ и $u_i \rightarrow y_{i+1}$ назовём связанными. Для БТП со сложными физико-биохимическими процессами возможен следующий вариант взаимодействия каналов: $u_i \rightarrow y_i$ и $u_i \rightarrow y_{i+1}$, $u_{i+1} \rightarrow y_{i+1}$ и $u_{i+1} \rightarrow y_i$; такие каналы называют взаимосвязанными, а соответствующие расчётные динамические характеристики обозначаются так: $h_{i+1,i}(t)$ и $h_{i,i+1}(t)$.

Далее нам понадобятся динамические характеристики по каналу $x_n \rightarrow y_i$, $i = \overline{1, n}$, получаемые при малых ступенчатых изменениях $\Delta x_n \in X_n$ и регистрации кривых разгона, а затем единичных переходных функций $h_i(t, \Delta x_n)$, где $\Delta x_n = 1$, $i = \overline{1, n}$. Некоторые из переходных функций могут иметь «ёмкостное» (или реже – «транспортное») запаздывание τ , которое можно определить при визуальном анализе кривых разгона и выделить (удалить) из $h(t)$ путём переноса начала координат в точку $t = \tau$. Впоследствии при моделировании АСР запаздывание необходимо вводить в управление $u(t - \tau)$.

Перейдём далее от временных динамических характеристик $h(t)$ к линейным дифференциальным уравнениям (ДУ_I, ДУ_{II}, ДУ_{III}), соответственно:

$$a_1 \dot{y}(t) + a_0 y(t) = 1(t), \quad a_2 \ddot{y}(t) + a_1 \dot{y}(t) + a_0 y(t) = 1(t), \\ a_3 \ddot{y}(t) + a_2 \dot{y}(t) + a_1 \dot{y}(t) + a_0 y(t) = 1(t),$$

где $1(t) = 1$ при $t \geq 0$, $1(t) = 0$ при $t < 0$; $a_I = \{a_0, a_1\}$; $a_{II} = \{a_0, a_1, a_2\}$; $a_{III} = \{a_0, a_1, a_2, a_3\}$ – вектор-коэффициенты ДУ.

При наличии единичной переходной функции $h(t)$, $0 \leq t \leq T$ – коэффициент $a_0 = h(T)$, где T определяется из условия $\dot{h}(T) \approx 0$, когда все производные равны нулю. Остальные коэффициенты a_1, a_2, a_3 найдём следующим образом.

Построим три функции $\Phi_\gamma(a_\gamma) = T^{-1} \int_0^T [h(t) - y(t, a_\gamma)]^2 dt$, $\gamma = I, II, III$, где $h(t)$ – единичная переходная функция; $y(t, a_\gamma)$ – решение ДУ _{γ} , $\gamma = I, II, III$, при известном a_0 и начальном условии $y_\gamma(0) = 0$, $\gamma = I, II, III$.

Найдём a_1^* , такое, что функция $\Phi_I(a_1^*) = \min_{a_1} \Phi_I(a_1)$, затем вычислим относительную среднюю квадратическую ошибку (СКО) аппроксимации $\sigma_I = (\Phi_I(a_1^*))^{1/2}$ и сравним её с достаточно большой допустимой СКО $\sigma_{ж}$: если $\sigma_I < \sigma_{ж}$, то задача решена и динамика канала описывается уравнением ДУ: $a_1^* \dot{y}(t) + a_0 y(t) = 1(t)$; иначе при $\sigma_I \geq \sigma_{ж}$ – переход к следующему этапу. Найдём a_{II}^* , такое, что функция $\Phi_{II}(a_{II}^*) = \min_{a_1, a_2} \Phi_{II}(a_{II})$. Затем вычислим $\sigma_{II} = (\Phi_{II}(a_{II}^*))^{1/2}$ и сравним с $\sigma_{ж}$: если $\sigma_{II} < \sigma_{ж}$, то задача решена и найдено аппроксимирующее уравнение ДУ с коэффициентами a_2^* , a_1^* и a_0^* ; иначе при $\sigma_{II} \geq \sigma_{ж}$ осуществляется переход к минимизации $\Phi_{III}(a_{III})$ и т.д. до определения a_{III}^* и $\sigma_{III} < \sigma_{ж}$; если $\sigma_{III} \geq \sigma_{ж}$, то целесообразно увеличить $\sigma_{ж}$ и повторить третий этап. Здесь всюду $\sigma_{ж} = (0,05 \dots 0,10 \text{ или } 0,15) h_{\max}$.

Задачи минимизации $\Phi_\gamma(a_\gamma)$, $\gamma = I, II, III$ разрешимы, всегда устойчивы и при выпуклых $\Phi_\gamma(a_\gamma)$ почти всегда имеют единственный минимум $\Phi^*(a)$, достигаемый при разных a_γ^* , что не имеет особого значения для данной задачи из-за низких требований по точности ДУ; при невыпуклых $\Phi_\gamma(a_\gamma)$ за решение задачи надо выбирать a_γ^* , такое, что $\Phi_\gamma(a_\gamma^*)$ минимально.

Для минимизации $\Phi_I(a_I)$ следует применять метод сканирования, для поиска минимума $\Phi_{II}(a_{II})$ и $\Phi_{III}(a_{III})$ можно использовать метод градиентов с начальными точками a_I^* , a_{II}^* .

Анализ статики биотехнологического процесса

В рамках используемой нами математической модели (3.13) идеальный статический режим БТП возникает при $x_n(t) = x_n = \text{const}$, $\tilde{x}(t) = \tilde{x} = \text{const}$, $z(t) \equiv 0$, $u(t) \equiv u = \text{const}$, $a = \text{const}$, $c_0 = 0$ и, как следствие, $y(t) \equiv y = \text{const}$, $\dot{y}(t) \equiv 0$. При этом векторное уравнение статики ТП будет таким:

$$f(x_n, y, u, a) = 0, \quad (3.14)$$

или

$$f(x_n, y_1, y_2, \dots, y_n, u, a) = 0 \quad i = \overline{1, n}, \quad (3.14a)$$

где $x_n = \{x_{nj}, j = \overline{1, m}\} \in X$; $y = \{y_i, i = \overline{1, n}\} \in Y$; $u = \{u_i, i = \overline{1, n}\} \in U$; X, U – ограниченные замкнутые множества из числового пространства E^n , $m \leq n$; Y – ограниченное открытое множество из E^n ; $f(\cdot)$ – непрерывная и дважды непрерывно дифференцируемая по x_n, u , y вектор-функция, в общем случае нелинейная по u, y, a .

Система конечных уравнений (3.14) или (3.14а) разрешима. Для определения численного решения $y(x_n)$ при фиксированных u, a требуется задать «грубое» начальное приближение y^0 и медленно сходящимся методом простой итерации найти «хорошее» приближение y^r , где верхние индексы «0» и «r» означают номера итерации. Далее этот вектор y^r принимается в качестве начального условия для быстро сходящегося итерационного метода Ньютона, который в случае неособенности якобиана $f(\cdot)$ на некоторой итерации l числовой вектор y^l , такой, что при подстановке его в $f(\cdot)$ получаем $f(x_n, y^l, u, a) < \varepsilon$, где ε – допустимая погрешность решения системы уравнений (3.14). Если из анализа переходных процессов при пуске ТП известно «хорошее» начальное приближение y^0 , то решение $y(x_n)$ системы (3.14) можно находить сразу методом Ньютона.

Определение численного решения $y(u)$ при фиксированных x_n, a осуществляется аналогично рассмотренному.

Численный анализ свойств $(m + n)$ -мерной функции $f(\cdot)$ или решений $y(x_n, u)$ крайне сложен, а возможные результаты этого анализа пока не находят применения в задачах математического моделирования проектируемого БТП. Ограничимся поэтому анализом статических свойств ТП по отдельным каналам $u_j \rightarrow y_i$ и $x_{nj} \rightarrow y_i$, $j = \overline{1, m}$, $m \leq n$, $i = \overline{1, n}$.

Перспективные (для автоматизации) каналы $u_j \rightarrow y_i$ определены при анализе динамических свойств БТП. Численное определение скалярной статической характеристики $y_i = \varphi_{ij}(u_j)$ сводится к заданию эквидистантных значений $u_{j\lambda}$, $\lambda = \overline{1, d}$, из отрезка $[0, u_j^+]$, решению уравнения (3.14) с фиксированными значениями $x_n = x_n^p; u_1, u_2, \dots, u_{j-1}, u_{j+1}, \dots, u_n, a$; переменными $u_{j\lambda}$, $\lambda = \overline{1, d}$, и получению величин $u_{i\lambda}$, $\lambda = \overline{1, d}$; построению графика $u_{j\lambda}, y_{i\lambda}$ для визуального анализа его топологии. Для получения математического описания канала $u_j \rightarrow y_i$ зададим линейную по c_l структуру φ_{ij} в следующем виде: $\varphi_{ij}(u_j) = \sum_{l=1}^k c_l u_j^l$, где k и c_l , $l = \overline{0, k}$, $k < d$, подлежат определению извест-

ным методом наименьших квадратов (МНК) по данным $u_{j\lambda}$, $y_{i\lambda}$, $\lambda = \overline{1, d}$.

Здесь возможны два варианта решения задачи аппроксимации.

Первый вариант: задано значение k , $k \in \{1, 2, 3 \text{ или } 4\}$, найти c_l , $l = \overline{0, k}$, и среднюю квадратическую ошибку (СКО) σ_k ; задана желаемая величина $\sigma_{ж}$, найти параметр k , коэффициенты c_l , $l = \overline{0, k}$, и СКО σ_k , такие, что $\sigma_k < \sigma_{ж}$.

Рассмотрим вначале более простой первый вариант задачи аппроксимации.

Пусть из анализа графика $u_{j\lambda}$, $y_{i\lambda}$ определён параметр k , $k < d$. Построим квадратичную функцию $k + 1$ переменных $c = \{c_0, c_1, \dots, c_k\}$:

$$\Phi(c) = \sum_{\lambda=1}^d \left(y_{i\lambda} - \sum_{l=0}^k c_l u_{j\lambda}^l \right)^2,$$

затем минимизируем её по c_l и из решения линейных уравнений $\partial\Phi(c)/\partial c_l = 0$ найдём МНК-значения c_l^* , $l = \overline{0, k}$. Зависимость $y_i = c_0^* + c_1^* u_j + \dots + c_k^* u_j^k$ описывает в среднем расчётные данные $u_{j\lambda}$, $y_{i\lambda}$ с погрешностью σ_k . Заметим, что программная реализация МНК имеется во всех математических пакетах прикладных программ.

Во втором варианте известно $\sigma_{ж}$, для определения c_l зададим минимально возможное k_1 , например $k_1 = 1$, построим функцию $\Phi(c_0, c_1)$ и найдём МНК-значения c_0^* , c_1^* и СКО σ_{k_1} . Если $\sigma_{k_1} < \sigma_{ж}$, то задача аппроксимации расчётных данных $u_{j\lambda}$, $y_{i\lambda}$ решена и полученная зависимость $y_i = c_0^* + c_1^* u_j$ описывает в среднем эти данные с погрешностью σ_{k_1} ; иначе надо задать $k_2 = k_1 + 1$ и, повторив МНК-процедуру, найти значения c_0^* , c_1^* , c_2^* и σ_{k_2} , а затем проверить выполнение условия $\sigma_{k_2} < \sigma_{ж}$, если оно выполнено, то задача решена и $y_i = c_0^* + c_1^* u_j + c_2^* u_j^2$, иначе при $\sigma_{k_2} \geq \sigma_{ж}$ следует задать $k_3 = k_2 + 1$ и т.д. до тех пор, пока при некотором k^* впервые не выполнится условие $\sigma_{k^*} < \sigma_{ж}$ и будет получена зависимость $y_i(u_j) = \sum_{l=0}^{k^*} c_l^* u_j^l$.

Графические $u_{j\lambda}$, $y_{i\lambda}$ и аналитические $y_i = \sum_{l=0}^k c_l u_j^l$ статические характеристики БТП определяются для всех (или некоторых) каналов

$u_j \rightarrow y_i, j = \overline{1, m}, i = \overline{1, n}$, при фиксированной нагрузке x_n^p , а затем при минимальной x_n^- и максимальной x_n^+ нагрузках.

Статические данные $x_n^p, y^p, x_n^+, y^+, x_n^-, y^-$ широко используются для расчёта технико-экономических показателей типа производительности $G = \varphi_1(x_n, y)$, себестоимости $C = \varphi_2(x_n, y)/\varphi_1(x_n, y)$, КПД $K = \varphi_3(x_n, y)/\varphi_4(x_n)$ и др., а результаты анализа статических свойств БТП используются при проектировании систем контроля и автоматизации.

Оптимизация проектируемого биотехнологического процесса

Напомним, что миссия проектируемого БТП заключается в экономически оптимальной переработке заданного количества исходных веществ в конечные продукты определённого качества. Следовательно, за критерий оптимальности БТП нужно принять некоторый широко распространённый технико-экономический показатель, например, себестоимость продукции $I(\cdot)$, зависящий от входных координат x_n, u и выходной переменной $y = y(x_n, u, a)$, являющейся решением векторного уравнения статики $f(x_n, y, u, a) = 0$. Тогда при постоянных x_n и a получаем запись для себестоимости $I(x_n, y(x_n, u, a), u) = I(u, y(u)) = I(u)$, где u – вектор управлений из выпуклого замкнутого множества $U = \{u_j, j = \overline{1, n}\}$. Себестоимость $I(u)$ не учитывает так называемые «расходы на собственные нужды» производства $\varphi_{сн}(x_n, y, u)$ и выбросы (потери) полупродуктов и веществ в окружающую среду $\varphi_{в}(x_n, y, u)$, которые, согласно технологическому регламенту, не должны превышать заданных нормативов $g_{сн}$ и $g_{в}$. Условия $\varphi_{сн}(\cdot) \leq g_{сн}$ и $\varphi_{в}(\cdot) \ll g_{в}$ можно записать так: $\varphi_{сн}(\cdot) - g_{сн} \leq 0$, $\varphi_{в}(\cdot) - g_{в} \leq 0$ или $g_1(x_n, y, u) \leq 0$, $g_2(x_n, y, u) \leq 0$. Если $g_{сн}, g_{в}$ достаточно велики (относительно $I(u)$), то указанные ограничения следует учитывать при оптимизации БТП, вычисляя $g_j(x_n, y, u)$ через решение $y(x_n, u)$ уравнения статики (поэтому произведение множеств $U \times g_1(\cdot) \times g_2(\cdot)$ будет непусто).

Предваряя постановки задач оптимизации проектируемого БТП, укажем ещё раз, что используемое при этом уравнение статики возможно будет неадекватно реальному функционирующему объекту, и полученные результаты оптимизации следует рассматривать как предварительные и уточнять их впоследствии с использованием адекватной математической модели.

Рассмотрим ряд задач оптимизации проектируемого БТП. Пусть управление u – скаляр (это возможно при синхронном изменении всех управлений $u_j, j = \overline{1, n}$, на отрезке $[u_j^-, u_j^+]$ одним регулирующим

органом); ограничения $g_i(\cdot)$, $i = 1, 2$, отсутствуют; функция $I(u)$ – гладкая (непрерывно дифференцируемая). Требуется найти минимальную себестоимость продукции $I^*(u)$, такую, что

$$I^*(u) = \min_{u \in U} I(u) \quad (3.15)$$

при связях $f(x_n^p, y, u, a) = 0$.

Задача минимизации (3.15) разрешима и устойчива. Если $I(u)$ выпуклая вниз функция, то решение I^* – единственное; при невыпуклой $I(u)$ число стационарных точек $I^c(u^c)$ может быть сколь угодно большим, тогда для выбора наименьшего значения I^* следует сканировать $I(u)$ при $u \in [u^-, u^+]$ и определять $I^* = \min_{\rho} (I_1, I_2, \dots, I_{\rho}, \dots, I_K)$, где $I_{\rho} = I(u_{\rho})$ – сканированная ордината $I(u_{\rho})$, $\rho = \overline{1, K}$. Если значение I^* достигается в точке u^* , расположенной внутри открытого множества (u^-, u^+) , т.е. слева и справа от $I^*(u^*)$ имеются ординаты I_{ρ} , $\rho = \overline{1, K}$, то эти ординаты можно аппроксимировать параболой $\tilde{I}(u) = c_0 + c_1 u + c_2 u^2$, коэффициенты c_0, c_1, c_2 которой определяются методом наименьших квадратов. Функция $\tilde{I}(u)$ сильно выпуклая и имеет единственный минимум $\tilde{I}^* (-c_1/2c_2)$ в точке $\tilde{u}^* = -c_1/2c_2$.

Найденные значения I^* и/или \tilde{I}^* , а также аппроксимирующая функция $\tilde{I}(u)$ и значение \tilde{u}^* используются в регламенте для характеристики оптимального рабочего режима функционирования БТП. Оптимизационная задача (3.15) решается аналогично для $x_n = x_n^-$ и $x_n = x_n^+$, соответствующие результаты решения также используются в технологическом регламенте.

В некоторых случаях более актуальной является аргументная задача оптимизации БТП, когда требуется найти управление u^* , такое, что

$$I(u^*) = \min_{u \in U} I(u) \quad (3.15a)$$

при связях $f(x_n^p, y, u, a) = 0$. При выпуклой и/или невыпуклой функции $I(u)$ аргументная задача (3.15a) разрешима и устойчива, но число решений u^* может быть сколь угодно большим. Лучшим способом определения u^* является сканирование функции $I(u)$ на отрезке $[u^-, u^+]$ и выявление минимальной ординаты $I^* = \min_{\rho} (I_1, I_2, \dots, I_{\rho}, \dots, I_K)$, где $I_{\rho} = I(u_{\rho})$, $\rho = \overline{1, K}$. Аргумент u^* , на котором ордината I^* минимальна, объявляется решением задачи (3.15a). Если $u^- < u^* < u^+$ и слева, и справа от u^* имеются несколько сканированных ординат I_{ρ} , то их можно

аппроксимировать параболой $\tilde{I}(u) = c_0 + c_1u + c_2u^2$, где c_0, c_1, c_2 – МНК-оценки, и определить $\tilde{u}^* = -c_1/2c_2$ и $\tilde{I}(\tilde{u}^*)$. Значения u^*, I^* и \tilde{u}^*, \tilde{I}^* используются при оптимизации функционирующего БТП с адекватной моделью статики.

Пусть управление $u = \{u_1, u_2\}$; остальные параметры задачи оптимизации аналогичны задаче (3.15). Требуется найти минимальную себестоимость $I^* \{u_1, u_2\}$, такую, что

$$I^*(u_1, u_2) = \min_{u \in U} I(u_1, u_2) \quad (3.15б)$$

при связях $f(x_n^p, y, u_1, u_2, a) = 0$ и условии $u_j^- \leq u_j \leq u_j^+, j = 1, 2$. При выпуклой функции $I(u_1, u_2)$ задача (3.15б) разрешима и устойчива, а единственное решение может достигаться в точках u^* , расположенных внутри или на границах замкнутого множества $U = \{u: u_j^- \leq u_j \leq u_j^+, j = 1, 2\}$.

Для исследования топологии $I(u_1, u_2)$ и определения I^* можно использовать известную итерационную градиентную процедуру: $u_j^{r+1} = u_j^r - h \cdot \partial I(u^r) / \partial u_j, j = 1, 2; r = 0, 1, 2, \dots, r^+$, где r – номер итерации, r^+ – номер окончания итерационного процесса при $|u_j^r - u_j^{r-1}| \leq \varepsilon_1$ или при $u_j^{r+1} \leq u_j^-$, или $u_j^{r+1} \geq u_j^+, j = 1, 2; h$ – параметр, подбираемый при счёте. Зададим несколько начальных точек $u_\lambda^{r=0} = u_\lambda^0, \lambda = \overline{1, d}, d \geq 15 \dots 20$, равномерно распределённых по множеству U ; затем запустим итерационные процедуры и зафиксируем точки их останова $u_\lambda^{r^*} = u_\lambda^*, \lambda = \overline{1, d}$. Если все u_λ^* примерно равны, то $I(u_1, u_2)$ – строго или просто выпукла и имеет единственную точку минимума u^* и $I^*(u^*)$, где u^* – среднее значение u_λ^* ; иначе $I(u_1, u_2)$ – невыпуклая функция, имеющая «плато» с $\|\text{grad} I(u)\| \approx 0$. В любом случае полезно аппроксимировать ординаты $I(u_\lambda^0), \lambda = \overline{0, d}$, и некоторые промежуточные значения $I(u_\lambda^r), 2 < r < r^+, \lambda = \overline{1, d}$, квадратичной параболой $\tilde{I}(u) = c_0 + c_1u_1 + c_2u_1^2 + c_3u_2 + c_4u_2^2$, где $c_0 \dots c_4$ – МНК-оценки, определяемые из условий $\partial I(u) / \partial c_j = 0, j = \overline{1, 4}$. Единственное значение $\min \tilde{I}(u)$ является приближённым решением задачи (3.15б).

Пусть управление $u = \{u_1, u_2, \dots, u_K\}, K \leq 5 \dots 7$; остальные компоненты задачи аналогичны задаче (3.15б). Требуется найти минимальную себестоимость $I^*(u_1, u_2, \dots, u_K) = I^*(u)$, такую, что

$$I^*(u) = \min_{u \in U} I(u)$$

при связях $f(x_{\text{н}}^{\text{р}}, y, u_1, u_2, a) = 0$ и условиях $u_j^- \leq u_j \leq u_j^+$, $j = \overline{1, K}$.

Анализ данной оптимизационной задачи и его результаты аналогичны исследованию топологии функции $I(u_1, u_2)$, однако аппроксимация ординат $I(u_{\lambda}^0)$, $\lambda = \overline{1, d}$, и $I(u_{\lambda}^r)$, $2 < r < r^+$, квадратичной параболой в данном случае крайне затруднительна (практически невозможна).

Для приближённого преобразования выпуклой или невыпуклой с «плато» функции $I(u)$ в сильно выпуклую функцию построим следующую зависимость:

$$\Phi_{\alpha}(u) = I(u) + \alpha(u - u_0)^2,$$

где u_0 – ранее найденная точка u_{λ}^{r*} с минимальной $I(u_{\lambda}^{r*})$ ординатой $I(u_{\lambda}^r)$; α – положительный параметр, подбираемый при минимизации $\Phi_{\alpha}(u)$ из условия её сильной выпуклости, что проявляется в быстрой сходимости итерационной процедуры к единственной «минимальной» точке u_{α}^* . За приближённое решение оптимизационной задачи принимается $I^* = I(u_{\alpha}^*)$, где α^* – минимальное значение α , при котором $\Phi_{\alpha^*}(u)$ – сильно выпуклая функция.

Пусть далее $u = \{u_1, u_2\} \in U = \{u_j : u_j^- \leq u_j \leq u_j^+, j = \overline{1, 2}\}$; имеется выпуклое ограничение $g_1(x_{\text{н}}^{\text{р}}, y, u) = g_1(\cdot) \leq 0$, функция $I(u_1, u_2)$ – гладкая, выпуклая. Требуется найти минимальную себестоимость $I^*(u)$, такую, что

$$I^*(u) = \min_u I(u) \quad (3.15\text{в})$$

при связи $f(x_{\text{н}}^{\text{р}}, y, u, a) = 0$ и выпуклом ограничении $g_1(x_{\text{н}}^{\text{р}}, y, u) \leq 0$.

Преобразуем эту условную оптимизационную задачу (3.15в) к задаче безусловной оптимизации (не учитывая множество U):

$$\Phi_{\alpha_i}(u) = I(u) + \alpha_i \langle g_1(\cdot) \rangle^2,$$

где α_i – положительный параметр, $i = 1, 2, \dots$, подбираемый в процессе минимизации $\Phi_{\alpha_i}(u)$; «штрафная» функция $\langle g_1(\cdot) \rangle^2$ типа квадрата «срезки»:

$$\langle g_1(\cdot) \rangle = 0 \text{ при } g_1(\cdot) \leq 0; \langle g_1(\cdot) \rangle = g_1(\cdot) > 0.$$

Поиск минимума $\Phi_{\alpha_i}(u)$ при заданном малом α_i осуществляется методом градиентов из произвольной начальной точки $u^{r=0} = u^0$, лежащей вне области допустимых управлений $U \times g_1(\cdot)$ (это – метод «внешней точки»). Последовательно увеличивая α_i , будем получать последовательность точек минимизирующей последовательности $u_{\alpha_i}^{r*}$, сходящейся к некоторой точке минимума \tilde{u}^* , на которой $I(\tilde{u}^*) \approx I^*$.

Если начальная точка $u^{r=0}$ принадлежит области $U \times g_1(\cdot)$, то следует применять метод «внутренней точки», в котором используется «штрафная» функция $\alpha_i g_1^{-1}(\cdot)$, стремящаяся к бесконечности при $g_1(\cdot) \rightarrow 0$. При минимизации $\Phi_{\alpha_i}(u)$ методом градиентов параметр штрафа α_i постепенно уменьшают, что способствует сходимости точек останова $u_{\alpha_i}^{r*}$ к некоторой точке \tilde{u}^* , на которой $I(\tilde{u}^*) \approx I^*$. Отметим, что все варианты метода «штрафных» функций не позволяют находить точное решение задачи условной оптимизации из-за образования «овражных» поверхностей в окрестности $g_1(x_n^p, y, u)$.

Пусть $u = \{u_1, u_2, u_3\} \in U = \{u_j : u_j^- \leq u_j \leq u_j^+, j = 1, 2, 3\}$; имеются выпуклые ограничения $g_\lambda(x_n^p, y, u) = g_\lambda(\cdot) \leq 0$, остальные компоненты аналогичны задаче (3.15в). Требуется найти минимальную себестоимость $I^* \{u_1, u_2, u_3\}$, такую, что

$$I^*(u_1, u_2, u_3) = \min_{u \in U} I(u_1, u_2, u_3)$$

при связях $f(x_n^p, y, u_1, u_2, u_3, a) = 0$ и ограничениях $g_1(\cdot) \leq 0, g_2(\cdot) \leq 0$. Преобразуем эту задачу условной оптимизации в задачу безусловной оптимизации методом «штрафной» функции:

$$\Phi_{\alpha_i}(u) = I(u) + \alpha_i \left(\langle g_1(\cdot) \rangle^2 + \langle g_2(\cdot) \rangle^2 \right).$$

Поиск минимума $\Phi_{\alpha_i}(u)$ при заданном α_i выполняется методом градиентов из произвольной начальной точки $u^{r=0} = u^0$, лежащей вне области $U \times g_1(\cdot) \times g_2(\cdot)$. Если постепенно увеличивать α_i , то последовательность точек останова градиентной процедуры $u_{\alpha_i}^{r*}$ будет сдвигаться к некоторой точке \tilde{u}^* , на которой достигается приближённое решение $I(\tilde{u}^*) \approx I^*$.

Если начальная точка $u^r = 0$ лежит внутри области $U \times g_1(\cdot) \times g_2(\cdot)$, то следует применять описанный метод внутренней точки со «штрафной» функцией $\alpha_i(g_1^{-1}(\cdot) + g_2^{-1}(\cdot))$, где α_i – убывающий параметр.

Примечание. Если в задаче оптимизации число ограничений $g_\lambda(\cdot)$, $\lambda = 3, 4, \dots$, то в уравнении статики необходимо использовать вектор управлений размерности $\lambda + 1$, т.е. $u = \{u_{\lambda+1}, \lambda = 3, 4, \dots\}$.

Построение областей допустимых режимов функционирования процесса

При постановке и решении задач статической оптимизации БТП необходима информация о существовании и размерах областей $U_{нд}$ допустимых нагрузок $x_n \in U_{нд}$ и допустимых управлений $u \in U$ проектируемого БТП. Предположим, что на стадии предпроектных исследований определены точные граничные значения вектора нагрузки x_n^- и x_n^+ , а минимальные и максимальные значения управления заданы априори и уточняются в ходе построения областей допустимых управлений U_d для каждого значения нагрузки по сырью $x_n = \{x_n^-, x_n^p, x_n^+\}$.

Анализ построенных областей допустимых управлений U_d позволит убедиться в их существовании для всего набора нагрузок и обеспечить возможность выбора номинального (рабочего) режима функционирования БТП с заданным ресурсом по управлению.

Допустимыми управлениями назовём такие переменные $u \in U_d$, которые обеспечивают выполнение регламентных ограничений на производительность БТП, себестоимость, КПД и т.п. Зададим эти ограничения в форме неравенств:

$$g_j(x_n, y, u) \leq 0, \quad j = \overline{1, m}, \quad (3.16)$$

значения которых будем вычислять с помощью уравнения статики ТП: $f(x_n, y, u) = 0$.

Подставляя значения (x_n, u) в уравнение (3.14) и решая его численно, определяем вектор y и затем вводим его в $g_j(x_n, y, u)$, $j = \overline{1, m}$. После этого проверяем выполнимость ограничений (3.16) в заданном наборе (x_n, u) . Если все ограничения выполняются, то точка (x_n, u) является допустимой, в противном случае – недопустимой.

Если управление u – скаляр, то задача построения области U_d сводится к определению двух граничных значений u^-, u^+ , а в случае вектора u и выпуклости функций $g_j(x_n, y, u)$ по переменной u – к опреде-

лению вершин параллелепипеда U_d , в которых должны выполняться ограничения (3.16).

Область U_d допустимых управлений можно построить методом сканирования функций $g_j(x_n, y, u)$, $j = \overline{1, m}$, для фиксированных нагрузок $x_n = \{x_n^-, x_n^p, x_n^+\}$. Покажем алгоритм сканирования на примере двухмерного вектора управлений $u = [u_1, u_2]$.

На первом шаге для заданной нагрузки $x_n = x_n^-$ подбираем допустимые нижние значения $u^- = [u_1^-, u_2^-]$. Процедура подбора проста: задаёмся произвольными значениями $u_1^{-(v)}$, $u_2^{-(v)}$, где v – номер итерации, подставляем набор $(x_n, u_1^{-(v)}, u_2^{-(v)})$ в уравнение (3.14) и определяем вектор $y^{(v)}$. Далее $y^{(v)}$ подставляем в $g_j(x_n, y, u)$ и проверяем выполнимость ограничений (3.16) при наборе $(x_n, u_1^{-(v)}, u_2^{-(v)})$. В случае выполнимости ограничений точка $(x_n, u_1^{-(v)}, u_2^{-(v)})$ является допустимой и можно уменьшить u^- по формуле $u^{-(v+1)} = u^{-(v)} - \Delta u$ до тех пор, пока $u^{-(v+1)}$ не станет недопустимым. В противном случае поочерёдно увеличиваются нижние границы $u^{-(v+1)} = u^{-(v)} + \Delta u$ до тех пор, пока $u^{-(v+1)}$ не станет допустимым.

После уточнения нижних границ u^- выполняем сканирование $g_j(x_n^-, y, u)$, $j = \overline{1, m}$, по u_2 при фиксированном значении u_1^- : изменяем $u_2^{(v)} = u_2^- + \Delta u_2$, подставляем переменные $x_n = x_n^-$, $u^{(v)} = [u_1^-, u_2^{(v)}]$ в уравнение (3.14) и, решая его, получаем значение $y^{(v)} = y(x_n^-, u^{(v)})$. Далее вводим в ограничения переменные $x_n = x_n^-$, $u^{(v)} = [u_1^-, u_2^{(v)}]$, $y^{(v)}$, вычисляем значения функций-ограничений $g_j(x_n^-, y^{(v)}, u^{(v)})$, $j = \overline{1, m}$, и анализируем выполнимость ограничений (3.16). Если они выполняются, то включаем значения $u^{(v)} = [u_1^-, u_2^{(v)}]$ в допустимую область U_d и переходим к следующему шагу. В противном случае вектор управлений $u^{(v)} = [u_1^-, u_2^{(v)}]$ не включаем в область U_d и переходим к следующему шагу, и т.д. при изменении $u_2^{(v)}$ до максимального значения u_2^+ .

Затем выполняем сканирование функций-ограничений $g_j(x_n^-, y, u)$ по u_2 при новом значении $u_1^{(k)} = u_1^- + \Delta u_1$, где k – номер

итерации по переменной u_1 . Сканирование продолжается до тех пор, пока не закончим построение области допустимых управлений в пространстве переменных u_1, u_2 для заданного значения $x_n = x_n^-$.

Далее зафиксируем нагрузки $x_n = x_n^p, x_n = x_n^+$ и повторяем все действия, описанные выше при изменении $u = [u_1, u_2]$ от минимальных до максимальных значений с шагами $\Delta u = [\Delta u_1, \Delta u_2]$, величины которых подбираем экспериментально.

Если размерность вектора u более 2, а число ограничений менее 3, то построение областей допустимых режимов U_d методом сканирования становится проблематичным. Необходимы другие, менее трудоёмкие подходы к анализу существования допустимых управлений, не требующих построения области U_d .

Введём функцию $\eta(x_n)$ выполнимости ограничений при фиксированном x_n с учётом связи между переменными x_n, y, u в форме (3.14):

$$\eta(x_n) = \min_u \max_j g_j(x_n, y, u). \quad (3.17)$$

Знак $\eta(x_n)$ указывает на соответствие или несоответствие функционирования БТП технологическому регламенту: если $\eta(x_n) < 0$ для всех $x_n = \{x_n^-, x_n^p, x_n^+\}$, то БТП функционирует в соответствии с технологическим регламентом; в противном случае имеет место его нарушение. Решая задачу (3.17) для фиксированного $x_n = \{x_n^-, x_n^p, x_n^+\}$, можно определить вектор управлений u , обеспечивающий выполнение ограничений (3.16).

Задачу (3.17) можно упростить, переписав её в форме оптимизационной задачи на условный минимум: определить скалярную величину $\alpha < 0$, такую, что $\eta(x_n) = \min_{u, \alpha} \alpha$, при выполнении связей в форме уравнений (3.14) и ограничений $g_j(x_n, y, u) \leq \alpha, j = \overline{1, m}$.

Глава 4

ОПТИМИЗАЦИЯ БИОТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ И ДЕТЕРМИНИРОВАННЫЕ ОЦЕНКИ АЛЬТЕРНАТИВ

4.1. ПОСТАНОВКА И АНАЛИЗ ОСНОВНЫХ КОМПОНЕНТОВ ОПТИМИЗАЦИОННЫХ ЗАДАЧ

Общая постановка задачи оптимизации – *поиска минимума функций* содержит:

1) целевую функцию $I(x)$, определённую на n -мерном евклидовом пространстве R^n , где $x = (x_1, \dots, x_n)^T$; её значения характеризуют степень достижения цели, во имя которой поставлена или решается задача оптимизации;

2) множество допустимых решений $X \subseteq R^n$, среди элементов которого осуществляется поиск.

Требуется найти такой вектор x^* из множества допустимых решений, которому соответствует минимальное значение целевой функции $I(x)$ на этом множестве X , т.е. $I(x^*) = \min_{x \in X} I(x)$.

Задача поиска минимума и максимума целевой функции $I(x)$ называется задачей поиска *экстремума*: $I(x^*) = \text{ext}_{x \in X} I(x)$, а задача поиска максимума функции $I(x)$ сводится к задаче поиска минимума путём замены знака перед функцией на противоположный:

$$I(x^*) = \max_{x \in X} I(x) = -\min_{x \in X} [-I(x)].$$

Если множество допустимых решений X задаётся ограничениями/условиями, накладываемыми на вектор x , то решается задача отыскания *условного экстремума*. Если $X = R^n$, т.е. ограничения/условия на вектор x отсутствуют, решается задача поиска *безусловного экстремума*.

Решением задачи поиска экстремума является пара $(x^*, I(x^*))$, включающая точку x^* и значение целевой функции в этой точке.

Множество точек минимума (максимума) целевой функции $I(x)$ на множестве X обозначим X^* . Оно может содержать конечное число точек (в том числе одну), бесконечное число точек или быть пустым.

Точка $x^* \in X$ называется точкой *глобального* минимума функции $I(x)$ на множестве X , если функция достигает в этой точке своего наименьшего значения, т.е.

$$I(x^*) \leq I(x) \quad \forall x \in X.$$

Точка $x^* \in X$ называется точкой *локального* минимума функции $I(x)$ на множестве X , если существует $\varepsilon > 0$, такое, что если $x \in X$ и

$\|x - x^*\| < \varepsilon$, то $I(x^*) \leq I(x)$. Здесь $\|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ – евклидова норма вектора x .

Поверхностью уровня $I(x)$ называется множество точек, в которых функция принимает постоянное значение, т.е. $I(x) = \text{const}$. Если $n = 2$, поверхности уровня изображаются линиями уровней на плоскости R^2 .

Градиентом $\nabla I(x)$ непрерывной дифференцируемой функции $I(x)$ в точке x называется вектор-столбец, элементами которого являются частные производные первого порядка, вычисленные в данной точке:

$$\nabla I(x) = \left(\frac{\partial I(x)}{\partial x_1}, \frac{\partial I(x)}{\partial x_2}, \dots, \frac{\partial I(x)}{\partial x_n} \right)^T.$$

Градиент функции направлен по нормали к поверхности уровня, т.е. перпендикулярно к касательной плоскости, проведённой в точке x , в сторону наибольшего возрастания функции в данной точке.

Матрицей Гессе $H(x)$ дважды непрерывно дифференцируемой в точке x функции $I(x)$ называется матрица частных производных второго порядка, вычисленных в данной точке:

$$H(x) = \begin{pmatrix} \frac{\partial^2 I(x)}{\partial x_1^2}, & \frac{\partial^2 I(x)}{\partial x_1 \partial x_2}, & \dots, & \frac{\partial^2 I(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 I(x)}{\partial x_2 \partial x_1}, & \frac{\partial^2 I(x)}{\partial x_2^2}, & \dots, & \frac{\partial^2 I(x)}{\partial x_2 \partial x_n} \\ \dots, & \dots, & \dots, & \dots \\ \frac{\partial^2 I(x)}{\partial x_n \partial x_1}, & \frac{\partial^2 I(x)}{\partial x_n \partial x_2}, & \dots, & \frac{\partial^2 I(x)}{\partial x_n^2} \end{pmatrix}.$$

Матрица Гессе является симметрической ($n \times n$).

Вместе с градиентом можно определить вектор *антиградиента*, равный по модулю вектору градиента, но противоположный по направлению. Он указывает в сторону наибольшего убывания функции в данной точке. С помощью градиента и матрицы Гессе, используя разложение в ряд Тейлора, приращение функции $I(x)$ в точке x может быть записано в форме

$$\Delta I(x) = I(x + \Delta x) - I(x) = \nabla I(x)^T \Delta x + \frac{1}{2} \Delta x^T H(x) \Delta x + O(\|\Delta x\|^2),$$

где $O(\|\Delta x\|^2)$ – сумма всех членов разложения, имеющих порядок выше второго; $\Delta x^T H(x) \Delta x$ – квадратичная форма.

Квадратичная форма $\Delta x^T H(x) \Delta x$ (а также соответствующая матрица Гессе $H(x)$) называется:

– *положительно определённой* ($H(x) > 0$), если для любого ненулевого Δx выполнится неравенство $\Delta x^T H(x) \Delta x > 0$;

– *отрицательно определённой* ($H(x) < 0$), если для любого ненулевого Δx выполнится неравенство $\Delta x^T H(x) \Delta x < 0$;

– *положительно полуопределённой* ($H(x) \geq 0$), если для любого ненулевого Δx выполнится неравенство $\Delta x^T H(x) \Delta x \geq 0$ и имеется отличный от нуля вектор Δx , для которого $\Delta x^T H(x) \Delta x = 0$;

– *отрицательно полуопределённой* ($H(x) \leq 0$), если для любого ненулевого Δx выполнится неравенство $\Delta x^T H(x) \Delta x \leq 0$ и имеется отличный от нуля вектор Δx , для которого $\Delta x^T H(x) \Delta x = 0$.

Множество $X \subseteq R^n$ называется *выпуклым*, если оно содержит всякий отрезок, концы которого принадлежат X , т.е. если для любых $x^1, x^2 \in X$ и $0 \leq \lambda \leq 1$ справедливо $\lambda x^1 + (1 - \lambda)x^2 \in X$.

Образно говоря, выпуклыми являются множества, которые не содержат «вмятин», «дырок» и состоят из одного «куска». Примерами выпуклых множеств служат само пространство R^n , отрезок, прямая, шар.

Функция $I(x)$, определённая на выпуклом множестве X , называется *выпуклой*, если

$$I(\lambda x^1 + (1 - \lambda)x^2) \leq \lambda I(x^1) + (1 - \lambda)I(x^2) \quad \forall x^1, x^2 \in X, \quad 0 \leq \lambda \leq 1.$$

Функция $I(x)$, определённая на выпуклом множестве X , называется *строго выпуклой*, если

$$I(\lambda x^1 + (1 - \lambda)x^2) < \lambda I(x^1) + (1 - \lambda)I(x^2) \quad \forall x^1, x^2 \in X, \quad x^1 \neq x^2, \\ 0 < \lambda < 1.$$

Функция $I(x)$, определённая на выпуклом множестве X , называется *сильно выпуклой*, если

$$I(\lambda x^1 + (1 - \lambda)x^2) \leq \lambda I(x^1) + (1 - \lambda)I(x^2) - \frac{1}{2} \lambda(1 - \lambda) \|x^1 - x^2\|^2 \\ \forall x^1, x^2 \in X, \quad x^1 \neq x^2, \quad 0 \leq \lambda \leq 1.$$

Далее при решении задач оптимизации будем использовать следующие свойства выпуклых функций:

1) если $I(x)$ выпуклая функция на выпуклом множестве X , то всякая точка локального минимума является точкой её глобального минимума на X ;

2) если выпуклая функция достигает своего минимума в двух различных точках, то она достигает минимума во всех точках отрезка, соединяющего эти две точки;

3) если $I(x)$ строго выпуклая функция на выпуклом множестве X , то она достигает своего глобального минимума на X не более чем в одной точке.

4.2. НЕОБХОДИМЫЕ И ДОСТАТОЧНЫЕ УСЛОВИЯ БЕЗУСЛОВНОГО ЭКСТРЕМУМА

Постановка задачи: дана дважды непрерывно дифференцируемая функция $I(x)$, определённая на множестве $X \in R^n$. Требуется исследовать функцию $I(x)$ на экстремум, т.е. определить точки $x^* \in R^n$ её локальных минимумов и максимумов на R^n :

$$I(x^*) = \min_{x \in R^n} I(x); \quad I(x^*) = \max_{x \in R^n} I(x).$$

Стратегия решения задачи: находятся точки x^* локальных экстремумов с помощью необходимых условий первого и второго порядка (порядок условий определяется порядком используемых производных), а также достаточных условий безусловного локального экстремума. Вычисляются значения $I(x^*)$ функции в найденных точках локальных экстремумов.

Необходимые условия экстремума первого порядка [10]: пусть $x^* \in R^n$ есть точка локального минимума (максимума) функции $I(x)$ на множестве R^n и $I(x)$ дифференцируема в точке x^* ; тогда градиент функции $I(x)$ в точке x^* равен нулю, т.е. $\nabla I(x^*) = 0$ или $\frac{\partial I(x^*)}{\partial x_i} = 0$, $i = 1,$

2, ..., n. Точки x^* , удовлетворяющие необходимым условиям экстремума первого порядка, называются *стационарными*.

Необходимые условия экстремума второго порядка [10]: пусть точка x^* есть точка локального минимума (максимума) функции $I(x)$ на множестве R^n и функция $I(x)$ дважды дифференцируема в этой точке. Тогда матрица Гессе $H(x^*)$ функции $I(x)$, вычисленная в точке x^* , является положительно полуопределённой (отрицательно полуопределённой), т.е. $H(x^*) \geq 0$, ($H(x^*) \leq 0$).

Достаточные условия экстремума [10]: пусть функция $I(x)$ в точке $x^* \in R^n$ дважды дифференцируема, её градиент равен нулю, а матрица

Гессе является положительно определённой (отрицательно определённой), т.е. $\nabla I(x^*) = 0$ и $H(x^*) > 0$, ($H(x^*) < 0$), тогда точка x^* есть точка локального минимума (максимума) функции $I(x)$ на множестве R^n .

Рассмотрим определитель матрицы Гессе $\det H(x^*) =$

$$= \begin{vmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \dots & \dots & \dots & \dots \\ h_{n1} & h_{n2} & \dots & h_{nn} \end{vmatrix},$$

вычисленной в точке x^* . Определители $\Delta_1 = h_{11}$,

$\Delta_2 = \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix}$, $\Delta_3 = \begin{vmatrix} h_{11} & \dots & h_{1n} \\ \dots & \dots & \dots \\ h_{n1} & \dots & h_{nn} \end{vmatrix}$ называются *угловыми минорами*.

Определители m -го порядка ($m \leq n$), получающиеся из определителя матрицы $H(x^*)$ вычеркиванием каких-либо $(n - m)$ строк и $(n - m)$ столбцов с одними и теми же номерами, называются *главными минорами*.

Для проверки выполнения достаточных условий экстремума и необходимых условий второго порядка используют два способа.

Первый способ (с помощью угловых и главных миноров) – *критерий Сильвестра*:

1) для того чтобы матрица Гессе $H(x^*)$ была положительно определённой ($H(x^*) > 0$) и точка x^* являлась точкой локального минимума, необходимо и достаточно, чтобы знаки угловых миноров были строго положительны: $\Delta_1 > 0$, $\Delta_2 > 0$, ..., $\Delta_n > 0$;

2) для того чтобы матрица Гессе $H(x^*)$ была отрицательно определённой ($H(x^*) < 0$) и точка x^* являлась точкой локального максимума, необходимо и достаточно, чтобы знаки угловых миноров чередовались, начиная с отрицательного: $\Delta_1 < 0$, $\Delta_2 > 0$, $\Delta_3 < 0$, ..., $(-1)^n \Delta_n > 0$.

Критерий проверки необходимых условий экстремума второго порядка:

1) для того чтобы матрица Гессе $H(x^*)$ была положительно полуопределённой ($H(x^*) \geq 0$) и точка x^* являлась точкой локального минимума, необходимо и достаточно, чтобы все главные миноры определителя матрицы Гессе были неотрицательны;

2) для того чтобы матрица Гессе $H(x^*)$ была отрицательно полуопределённой ($H(x^*) \leq 0$) и точка x^* являлась точкой локального максимума, необходимо и достаточно, чтобы все главные миноры чётного порядка определителя матрицы Гессе были неотрицательны, а все главные миноры нечётного порядка – неположительны.

Второй способ (с помощью собственных значений матрицы Гессе).

Собственные значения λ_i , $i = 1, 2, \dots, n$ матрицы $H(x^*)$ размера $(n \times n)$ находятся как корни характеристического уравнения (алгебраического уравнения n -й степени):

$$\left| H(x^*) - \lambda E \right| = \begin{vmatrix} h_{11} - \lambda & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} - \lambda & \dots & h_{2n} \\ \dots & \dots & \dots & \dots \\ h_{n1} & h_{n2} & \dots & h_{nn} - \lambda \end{vmatrix} = 0.$$

Собственные значения вещественной симметрической матрицы $H(x^*)$ вещественны.

Для случая функции $I(x)$ одной переменной ($n = 1$) можно сформулировать следующее правило – если $I(x)$ и её производные непрерывны, то точка x^* является точкой экстремума тогда и только тогда, когда число m – чётное, где m – порядок первой не обращающейся в нуль в точке x^* производной:

- 1) если $I^{(m)}(x^*) > 0$, то в точке x^* – локальный минимум;
- 2) если $I^{(m)}(x^*) < 0$, то в точке x^* – локальный максимум;
- 3) если m нечётное, в точке x^* нет экстремума.

Часто на практике, особенно при применении численных методов, требуется проверить, выполняются ли необходимые и достаточные условия экстремума в некоторой точке. Такой анализ необходим, так как многие численные методы позволяют найти лишь стационарную точку, тип которой требует уточнения. Большое количество примеров на эту тему разобрано в книге [10].

4.3. НЕОБХОДИМЫЕ И ДОСТАТОЧНЫЕ УСЛОВИЯ УСЛОВНОГО ЭКСТРЕМУМА

Рассмотрим случаи, когда множество допустимых решений задётся равенствами и неравенствами, т.е.

$$I(x^*) = \min_{x \in X} I(x),$$

где $X = \left\{ x \left| \begin{array}{l} g_j(x) = 0, \quad j = 1, 2, \dots, m; \quad m < n \\ g_j(x) \leq 0, \quad j = m+1, \dots, p \end{array} \right. \right\}$, m и p – числа; $I(x)$ – целевая функция; $g_j(x)$, $j = 1, 2, \dots, p$ – функции, задающие ограничения (условия).

Будем считать функции $I(x)$, $g_j(x)$, $j = 1, 2, \dots, p$, дважды непрерывно дифференцируемыми на множестве R^n . При $p = m$ задача со

смешанными ограничениями преобразуется в задачу с ограничениями типа равенств, а при $m = 0$ – в задачу с ограничениями типа неравенств.

Функция $L(x, \lambda_0, \lambda) = \lambda_0 I(x) + \sum_{j=1}^P \lambda_j g_j(x)$ называется *обобщённой функцией Лагранжа*, числа $\lambda_0, \lambda_1, \dots, \lambda_p$ – множителями Лагранжа; *классической функцией Лагранжа* называется функция $L(x, \lambda) = I(x) + \sum_{j=1}^P \lambda_j g_j(x)$.

Условный экстремум при ограничениях типа равенств. Постановка задачи: даны дважды непрерывно дифференцируемые целевая функция $I(x) = (x_1, x_2, \dots, x_n)$ и функции ограничений $g_j(x) = g_j(x_1, x_2, \dots, x_n) = 0, j = 1, 2, \dots, m$, определяющие множество допустимых решений X ; требуется исследовать функцию $I(x)$ на экстремум, т.е. определить точки $x^* \in X$ её локальных минимумов и максимумов на множестве X :

$$I(x^*) = \min_{x \in X} I(x); \quad I(x^*) = \max_{x \in X} I(x), \quad (4.1)$$

где $X = \{x | g_j(x) = 0, j = 1, 2, \dots, m; m < n\}$.

Стратегия решения задачи: находятся точки x^* локальных экстремумов с помощью необходимых и достаточных условий минимума и максимума первого и второго порядка при ограничениях типа равенств, вычисляются значения $I(x^*)$ функции в найденных точках локального экстремума.

Необходимые условия экстремума первого порядка [10]: пусть x^* точка локального экстремума в задаче (4.1), тогда найдутся числа $\lambda_0^*, \lambda_1^*, \dots, \lambda_p^*$, не равные одновременно нулю, и такие, что выполняются следующие условия:

1) условие стационарности обобщённой функции Лагранжа по x :

$$\frac{\partial L(x^*, \lambda_0^*, \lambda^*)}{\partial x_i} = 0, \quad i = 1, 2, \dots, n; \quad (4.1a)$$

2) условие допустимости решения:

$$g_j(x^*) = 0, \quad j = 1, 2, \dots, m; \quad (4.1б)$$

3) если при этом градиенты $\nabla g_1(x^*), \nabla g_2(x^*), \dots, \nabla g_m(x^*)$ в точке x^* линейно независимы (выполняется условие регулярности), то $\lambda_0^* \neq 0$.

Условие (4.1a) можно записать в векторной форме $\nabla_x L(x^*, \lambda_0^*, \lambda^*) = 0$. Система (4.1) содержит $n + m$ уравнений с $n + m + 1$ неизвест-

ными λ_0^* , $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)^T$, $x^* = (x_1^*, x_2^*, \dots, x_n^*)^T$. Точки x^* , удовлетворяющие системе при некоторых λ_0^* , λ^* , называются условно-стационарными.

При решении задач проверка условия регулярности затруднена, так как точка x^* заранее не известна. Поэтому, как правило, рассматриваются два случая: $\lambda_0^* = 0$ и $\lambda_0^* \neq 0$. Если $\lambda_0^* \neq 0$, в системе (4.1а) полагают $\lambda_0^* = 1$, что эквивалентно делению системы (4.1а) на λ_0^* и замене $\frac{\lambda_j^*}{\lambda_0^*}$ на λ_j^* . При этом обобщённая функция Лагранжа становится классической, а сама система (4.1а, 4.1б) имеет вид:

$$\begin{aligned} \frac{\partial L(x^*, \lambda^*)}{\partial x_i} &= 0, \quad i = 1, 2, \dots, n; \\ g_j(x^*) &= 0, \quad j = 1, 2, \dots, m. \end{aligned} \quad (4.2)$$

Здесь число уравнений равно числу неизвестных.

Система (4.2) отражает тот факт, что антиградиент целевой функции в регулярной точке экстремума x^* является линейной комбинацией градиентов ограничений. Действительно, с учётом $L(x, \lambda) = I(x) + \sum_{j=1}^p \lambda_j g_j(x)$ можно переписать условие (4.2а) в форме $\nabla I(x^*) + \sum_{j=1}^m \lambda_j^* g_j(x^*) = 0$, отсюда $-\nabla I(x^*) = \sum_{j=1}^m \lambda_j^* \nabla g_j(x^*)$.

Точка x^* условного экстремума является точкой касания линии уровня целевой функции и кривой, описывающей ограничение; при $\lambda_0^* \neq 0$ эта точка называется *регулярной*, а при $\lambda_0^* = 0$ – *нерегулярной*.

Случай $\lambda_0^* = 0$ отражает вырожденность ограничений; при этом в обобщённой функции Лагранжа исчезает член, содержащий целевую функцию, а в необходимых условиях экстремума не используется информация, представляемая градиентом целевой функции.

Необходимые условия экстремума второго порядка [10]. Пусть x^* – регулярная точка минимума (максимума) в задаче (4.1) и имеется решение (x^*, λ^*) системы (4.2). Тогда второй дифференциал классической функции Лагранжа, вычисленный в точке (x^*, λ^*) , неотрицателен (неположителен):

$$d^2 L(x^*, \lambda^*) \geq 0 \quad (d^2 L(x^*, \lambda^*) \leq 0) \quad (4.3)$$

для всех $dx \in R^n$, таких, что $dg_j(x^*) = \sum_{i=1}^n \frac{\partial g_j(x^*)}{\partial x_i} dx_i = 0$, $j = 1, 2, \dots, m$.

Достаточные условия экстремума [10]. Пусть имеется точка (x^*, λ^*) , удовлетворяющая системе (4.2). Если в этой точке $d^2L(x^*, \lambda^*) > 0$ ($d^2L(x^*, \lambda^*) < 0$) для всех ненулевых $dx \in R^n$, таких, что

$$dg_j(x^*) = \sum_{i=1}^n \frac{\partial g_j(x^*)}{\partial x_i} dx_i = 0, \quad j = 1, 2, \dots, m,$$

то точка x^* является точкой локального минимума (максимума) в задаче (4.1). Достаточные и необходимые условия экстремума второго порядка проверяются в условно-стационарных точках, которые удовлетворяют системе (4.1) при $\lambda_0^* \neq 0$ или системе (4.2), так как для практики безусловно представляет интерес случай, когда в функции Лагранжа присутствует целевая функция, экстремум которой ищется.

Алгоритм решения задачи.

Шаг 1. Составить обобщённую функцию Лагранжа:

$$L(x, \lambda_0, \lambda) = \lambda_0 I(x) + \sum_{j=1}^m \lambda_j g_j(x).$$

Шаг 2. Записать необходимые условия экстремума первого порядка:

а) $\frac{\partial L(x^*, \lambda_0^*, \lambda^*)}{\partial x_i} = 0, \quad i = 1, 2, \dots, n;$

б) $g_j(x^*) = 0, \quad j = 1, 2, \dots, m.$

Шаг 3. Решить систему для двух случаев:

1) $\lambda_0^* = 0;$

2) $\lambda_0^* \neq 0$ (при этом поделить условие «а» на λ_0^* и заменить $\frac{\lambda_j^*}{\lambda_0^*}$

на λ_j^*). В результате найти условно-стационарные точки x^* , выделив из них полученные при $\lambda_0^* \neq 0$ (они могут быть регулярными точками экстремума).

Шаг 4. Для выделенных на шаге 3 точек проверить достаточные условия экстремума:

а) записать выражение для второго дифференциала классической функции Лагранжа в точке (x^*, λ^*)

$$d^2L(x^*, \lambda^*) = \sum_{i=1}^n \sum_{j=1}^n \frac{d^2L(x^*, \lambda^*)}{\partial x_i \partial x_j} dx_i dx_j.$$

б) записать систему $dg_j(x^*) = \sum_{i=1}^n \frac{\partial g_j(x^*)}{\partial x_i} dx_i = 0, j = 1, 2, \dots, m$

в точке x^* ;

в) из предыдущей системы выразить любые m дифференциалов dx_i через остальные $(n - m)$ и подставить в $d^2L(x^*, \lambda^*)$;

г) если $d^2L(x^*, \lambda^*) > 0$ при ненулевых dx , то в точке x^* – условный локальный минимум; если $d^2L(x^*, \lambda^*) < 0$ при ненулевых dx , то в точке x^* – условный локальный максимум. Если достаточные условия экстремума не выполняются, следует проверить выполнение необходимых условий второго порядка и в случае их невыполнения сделать вывод, что в точке x^* нет условного экстремума.

Шаг 5. Вычислить значения целевой функции в точках условного экстремума.

Примеры решения экстремальных задач при ограничениях типа равенств достаточно подробно разобраны в книге [10].

Условный экстремум при ограничениях типа неравенств. Постановка задачи: даны дважды непрерывно дифференцируемые целевая функция $I(x) = I(x_1, x_2, \dots, x_n)$ и функции ограничений $g_j(x) = g_j(x_1, x_2, \dots, x_n) \leq 0, j = 1, 2, \dots, m$, определяющие множество допустимых решений X ; требуется исследовать функцию $I(x)$ на экстремум, т.е. определить точки $x^* \in X$ её локальных минимумов и максимумов на множестве X (4.1):

$$I(x^*) = \min_{x \in X} I(x); \quad I(x^*) = \max_{x \in X} I(x),$$

где $X = \{x | g_j(x) \leq 0, j = 1, 2, \dots, m\}$.

Стратегия решения задачи: находятся точки x^* локального экстремума с помощью необходимых и достаточных условий минимума и максимума первого и второго порядка при ограничениях типа неравенств; затем вычисляются значения $I(x^*)$ функции в найденных точках локального экстремума.

Необходимые условия минимума (максимума) первого порядка (были доказаны Ф. Джоном (F. John), а при $\lambda_0^* \neq 0$ Куном и Таккером (H. W. Kuhn, A. W. Tucker)): пусть x^* – точка локального минимума (максимума) в задаче (4.1), тогда найдётся такое число $\lambda_0^* \geq 0$ и вектор

$\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)^T$, не равные одновременно нулю, и такие, что выполняются следующие условия:

1) стационарности обобщённой функции Лагранжа по x :

$$\frac{\partial L(x^*, \lambda_0^*, \lambda^*)}{\partial x_i} = 0, \quad i = 1, 2, \dots, n; \quad (4.4a)$$

2) допустимости решения:

$$g_j(x^*) \leq 0, \quad j = 1, 2, \dots, m; \quad (4.4б)$$

3) неотрицательности для условного минимума:

$$\lambda_j^* \geq 0, \quad j = 1, 2, \dots, m \text{ (условие неположительности)} \\ \text{для условного максимума} - \lambda_j^* \leq 0, \quad j = 1, 2, \dots, m); \quad (4.4в)$$

4) дополняющей нежесткости:

$$\lambda_j^* g_j(x^*) = 0, \quad j = 1, 2, \dots, m. \quad (4.4г)$$

Если при этом градиенты активных в точке x^* ограничений линейно независимы (выполняется условие регулярности), то $\lambda_0^* \neq 0$.

Точки x^* , удовлетворяющие системе (4.4а – 4.4г), называются *условно-стационарными*. Точка экстремума, удовлетворяющая предыдущей системе при $\lambda_0^* \neq 0$, называется регулярной, а при $\lambda_0^* = 0$ – нерегулярной (случай $\lambda_0^* = 0$ отражает вырожденность ограничений).

Условие $\frac{\partial L(x^*, \lambda_0^*, \lambda^*)}{\partial x_i} = 0, \quad i = 1, 2, \dots, n$ в регулярной точке экстремума x^* отражает тот факт, что антиградиент целевой функции является неотрицательной (неположительной в случае максимума) линейной комбинацией градиентов функций, образующих активные ограничения в точке x^* . Действительно, условие

$\frac{\partial L(x^*, \lambda_0^*, \lambda^*)}{\partial x_i} = 0, \quad i = 1, 2, \dots, n,$

с учётом $\lambda_j^* g_j(x^*) = 0, \quad j = 1, 2, \dots, m,$ можно переписать в форме

$$-\nabla I(x^*) = \sum_{j=1}^m \lambda_j^* \nabla g_j(x^*) = \sum_{j \in J_a} \lambda_j^* \nabla g_j(x^*).$$

Если функции $I(x), g_j(x), j = 1, 2, \dots, m,$ выпуклые, то необходимые условия являются одновременно и достаточными условиями гло-

бального минимума, при этом множество допустимых решений X выпукло. Из условия дополняющей нежесткости следует, что если ограничение в точке x^* пассивное, т.е. $g_j(x) < 0$, то $\lambda_j^* = 0$, а если – активное, т.е. $g_j(x) = 0$, то $\lambda_j^* \geq 0$ (для минимума) и $\lambda_j^* \leq 0$ (для максимума).

Достаточные условия минимума (максимума) первого порядка. Пусть имеется точка (x^*, λ^*) , удовлетворяющая системе (4.4а – 4.4г) при $\lambda_0^* \neq 0$, число активных ограничений в точке x^* совпадает с числом n переменных (при этом условие регулярности выполняется). Если $\lambda_j^* > 0$ для всех $j \in J_a$, то точка x^* – точка условного локального минимума; если $\lambda_j^* < 0$ для всех $j \in J_a$, то точка x^* – точка условного локального максимума в задаче (4.3).

Необходимое условие минимума (максимума) второго порядка. Пусть x^* – регулярная точка минимума (максимума) в задаче (4.3) и имеется решение (x^*, λ^*) системы (4.4а – 4.4г). Тогда второй дифференциал классической функции Лагранжа, вычисленный в точке (x^*, λ^*) , неотрицателен (неположителен):

$$d^2L(x^*, \lambda^*) \geq 0 \quad (d^2L(x^*, \lambda^*) \leq 0) \quad (4.5a)$$

для всех $dx \in R^n$, таких, что:

$$dg_j(x^*) = 0, \quad j \in J_a, \quad \lambda_j^* > 0 \quad (\lambda_j^* < 0);$$

$$dg_j(x^*) \leq 0, \quad j \in J_a, \quad \lambda_j^* = 0. \quad (4.5б)$$

Достаточные условия экстремума второго порядка: пусть имеется точка (x^*, λ^*) , удовлетворяющая системе (4.4а – 4.4г) при $\lambda_0^* \neq 0$. Если в этой точке $d^2L(x^*, \lambda^*) \geq 0$ ($d^2L(x^*, \lambda^*) \leq 0$) для всех ненулевых $dx \in R^n$, таких, что $dg_j(x^*) = 0, j \in J_a, \lambda_j^* > 0$ ($\lambda_j^* < 0$); $dg_j(x^*) \leq 0, j \in J_a, \lambda_j^* = 0$, то точка x^* является точкой локального минимума (максимума) в задаче (4.3).

Стратегия решения задачи: находятся точки x^* локального экстремума с помощью необходимых и достаточных условий минимума и максимума первого и второго порядка при ограничениях типа неравенств (порядок условий определяется порядком используемых производных). Вычисляются значения $I(x^*)$ функции в найденных точках.

Алгоритм решения задачи.

Шаг 1. Составить обобщённую функцию Лагранжа:

$$L(x, \lambda_0, \lambda) = \lambda_0 I(x) + \sum_{j=1}^m \lambda_j g_j(x).$$

Шаг 2. Записать необходимые условия минимума (максимума) первого порядка:

а) $\frac{\partial L(x^*, \lambda_0^*, \lambda^*)}{\partial x_i} = 0, i = 1, 2, \dots, n;$

б) $g_j(x^*) \leq 0, j = 1, 2, \dots, m;$

в) $\lambda_j^* \geq 0, j = 1, 2, \dots, m$ (для минимума), $\lambda_j^* \leq 0, j = 1, 2, \dots, m$

(для максимума);

г) $\lambda_j^* g_j(x^*) = 0, j = 1, 2, \dots, m.$

Шаг 3. Решить систему для двух случаев:

1) $\lambda_0^* = 0;$

2) $\lambda_0^* \neq 0$ (при этом поделить условия, записанные на шаге 2, на

λ_0^* и заменить $\frac{\hat{\lambda}_j^*}{\lambda_0^*}$ на λ_j^*).

В результате найти условно-стационарные точки x^* , выделив из них полученные при $\lambda_0^* \neq 0$ (они могут быть регулярными точками экстремума). В каждом из двух случаев следует начинать с рассмотрения 2^m вариантов удовлетворения условия «г» дополняющей нежёсткости.

Шаг 4. Для выделенных на шаге 3 точек проверить достаточные условия экстремума первого или второго порядка:

а) определить число l активных в точке x^* ограничений;

б) если $l = n$ и $\lambda_j^* > 0$ для всех $j \in J_a$, то в точке x^* – локальный минимум; если $l = n$ и $\lambda_j^* < 0$ для всех $j \in J_a$, то в точке x^* – локальный максимум.

Для проверки достаточных условий второго порядка следует:

а) записать выражение для второго дифференциала классической функции Лагранжа в точке (x^*, λ^*) :

$$d^2 L(x^*, \lambda^*) = \sum_{i=1}^n \sum_{j=1}^n \frac{d^2 L(x^*, \lambda^*)}{\partial x_i \partial x_j} dx_i dx_j;$$

б) записать условия, накладываемые на первые дифференциалы активных ограничений: $dg_j(x^*) = \sum_{i=1}^n \frac{\partial g_j(x^*)}{\partial x_i} dx_i = 0, \quad j \in J_a, \lambda_j^* > 0$

$$(\lambda_j^* < 0); \quad dg_j(x^*) = \sum_{i=1}^n \frac{\partial g_j(x^*)}{\partial x_i} dx_i \leq 0, \quad j \in J_a, \lambda_j^* = 0;$$

в) исследовать знак второго дифференциала функции Лагранжа для ненулевых dx , удовлетворяющих предыдущей системе; если $d^2L(x^*, \lambda^*) > 0$, то в точке x^* – условный локальный минимум, в случае $d^2L(x^*, \lambda^*) < 0$ в точке x^* – условный локальный максимум.

Если достаточные условия первого и второго порядка не выполняются, следует проверить выполнение необходимых условий второго порядка, следуя аналогичной процедуре. Если они выполняются, то требуется дополнительное исследование, а если нет, то в точке x^* нет условного экстремума.

Шаг 5. Вычислить значения функции в точках условного экстремума.

Исследование функции $I(x)$ на экстремум при ограничениях типа равенств, неравенств, а также смешанных ограничениях с помощью необходимых и достаточных условий минимума и максимума первого и второго порядка и большое количество разобранных примеров приведено в книге [10].

4.4. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ БЕЗУСЛОВНЫХ ОПТИМИЗАЦИОННЫХ ЗАДАЧ

Принципы построения численных методов поиска безусловного экстремума

Применение необходимых и достаточных условий безусловного экстремума, изложенных в предыдущем параграфе, эффективно для решения ограниченного числа задач, в которых вытекающие из условий соотношения имеют аналитическое решение. Для решения большинства практических задач они не могут быть рекомендованы по следующим причинам:

- 1) целевая функция $I(x)$ может не иметь непрерывные производных до второго порядка включительно;
- 2) использование необходимого условия первого порядка ($\nabla I(x^*) = 0$) связано с решением системы n в общем случае нелинейных алгебраических уравнений, что представляет собой самостоятельную задачу, трудоёмкость решения которой сравнима с трудоёмкостью численного решения поставленной задачи поиска экстремума;

3) возможны случаи, когда о целевой функции известно лишь то, что её значение может быть вычислено нужной точностью, а сама функция задана неявно.

Подвляющее большинство численных методов оптимизации относится к классу *итерационных*, т.е. порождающих последовательность точек в соответствии с предписанным набором правил, включающим критерий окончания поиска экстремума. В заданной начальной точке x^0 методы генерируют последовательность x^0, x^1, x^2, \dots . Преобразование точки x^k в x^{k+1} представляет собой *итерацию*.

Для определённости рассмотрим задачу поиска безусловного локального минимума: $I(x^*) = \min_{x \in R^n} I(x)$. Численное решение этой задачи связано с построением последовательности $\{x^k\}$ точек, обладающих свойством $I(x^{k+1}) < I(x^k), k = 0, 1, \dots$.

Общее правило построения последовательности $\{x^k\}$ имеет вид

$$x^{k+1} = x^k + \alpha_k s^k, k = 0, 1, \dots,$$

где x^0 – начальная точка поиска; s^k – приемлемое направление перехода из точки x^k в точку x^{k+1} , обеспечивающее выполнение условия $I(x^{k+1}) < I(x^k), k = 0, 1, \dots$, и называемое *направлением спуска*; α_k – *величина шага*.

Начальная точка поиска x^0 задаётся исходя из физического содержания решаемой задачи и наличия априорной информации о положении точек экстремума.

Приемлемое направление спуска s^k должно удовлетворять условию

$$(\nabla I(x^k), s^k) < 0, k = 0, 1, \dots,$$

обеспечивающему убывание функции $I(x)$. Примером приемлемого направления является направление вектора антиградиента $s^k = -\nabla I(x^k)$.

Величина шага $\alpha_k > 0$ выбирается либо из условия $I(x^{k+1}) < I(x^k), k = 0, 1, \dots$, либо из условия минимума функции вдоль направления спуска

$$I(x^k + \alpha_k s^k) \rightarrow \min_{\alpha_k}.$$

Выбор шага из последнего условия делает спуск в направлении s^k *наискорейшим*.

Последовательность $\{x^k\}$ называется *минимизирующей*, если $\lim_{k \rightarrow \infty} I(x^k) = I^*$, т.е. последовательность сходится к нижней грани $I^* = \inf_{x \in R^n} I(x)$.

Последовательность $\{x^k\}$ называется *сходящейся к точке минимума* x^* , если $\|x^k - x^*\| \rightarrow 0$ при $k \rightarrow \infty$.

В зависимости от наивысшего порядка частных производных целевой функции $I(x)$, используемых для формирования s^k и α_k , численные методы решения задачи безусловной оптимизации принято делить на три группы:

1) *методы нулевого порядка*, использующие только информацию о значении функции $I(x)$;

2) *методы первого порядка*, использующие информацию о первых производных целевой функции $I(x)$;

3) *методы второго порядка*, требующие для своей реализации знания вторых производных целевой функции $I(x)$.

Работоспособность метода ещё не гарантирована доказательством сходимости соответствующей последовательности – нужна определённая скорость сходимости. Рассмотрим последовательность $\{x^k\}$, сходящуюся к x^* . Наиболее эффективный способ оценивания скорости сходимости состоит в сопоставлении расстояния между x^{k+1} и x^* и расстояния между x^k и x^* .

Последовательность $\{x^k\}$ называется сходящейся с порядком r , если r – максимальное число, для которого

$$0 \leq \lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^r} < \infty.$$

Поскольку величина r определяется предельными свойствами $\{x^k\}$, она называется *асимптотической скоростью сходимости*.

Если последовательность $\{x^k\}$ сходится с порядком r , то число $c = \lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^r}$ называется *асимптотическим параметром ошибки*.

Если $r = 1$, $c < 1$, то сходимость *линейная*; если $r = 2$, то сходимость *квадратическая*; если $r = 1$, $c = 0$, то сходимость *сверхлинейная*. Линейная сходимость является синонимом сходимости со скоростью геометрической прогрессии.

На практике линейная сходимость может быть медленной, в то время как квадратичная или сверхлинейная сходимость является довольно быстрой. Реальное поведение итерационного процесса зависит от константы c , например, линейная сходимость с константой $c = 0,001$, вероятно, является вполне удовлетворительной, а с константой $c = 0,9$ – нет.

Одним из критериев сходимости, часто используемым при сравнении алгоритмов, является их способность эффективно минимизиро-

вать квадратичные функции. Это объясняется тем, что вблизи минимума квадратичная функция может быть достаточно хорошей аппроксимацией целевой функции. Таким образом, алгоритм, который не даёт хороших результатов при минимизации квадратичной функции, вряд ли может быть с успехом использован в случае общей нелинейной функции, когда текущая точка находится в окрестности минимума.

Методы нулевого порядка. Методы одномерной оптимизации

Постановка задачи: требуется найти безусловный минимум функции $I(x)$ одной переменной, т.е. такую точку $x^* \in R$, что $I(x^*) = \min_{x \in R} I(x)$.

Поставленная задача одномерной минимизации может быть решена с помощью необходимых и достаточных условий безусловного экстремума. Однако проблема получения решения уравнений $\frac{dI(x)}{dx} = 0$ может оказаться весьма сложной. Более того, в практических задачах функция $I(x)$ может быть не задана в аналитическом виде или часто неизвестно, является ли она дифференцируемой. Поэтому получение численного решения поставленной задачи является актуальным.

Для методов одномерной минимизации типично:

а) задание априорной информации о положении точки минимума с помощью начального интервала неопределённости $L_0 = [a_0, b_0]$; предполагается, что точка минимума x^* принадлежит интервалу L_0 , но её точное значение неизвестно;

б) большинство известных методов одномерной минимизации применяется для класса унимодальных функций.

Функция $I(x)$ называется *унимодальной на интервале* $L_0 = [a_0, b_0]$, если она достигает глобального минимума на $[a_0, b_0]$ в единственной точке x^* , причём слева от x^* эта функция строго убывает, а справа от x^* строго возрастёт. Если $a_0 \leq y < z < x^*$, то $I(y) > I(z)$, а если $x^* < y < z \leq b_0$, то $I(y) < I(z)$.

Методы одномерной минимизации широко применяются в методах первого и второго порядков для нахождения оптимальной величины шага.

Стратегии поиска

Существуют две принципиально различные стратегии выбора точек, в которых производится вычисление значений целевой функции. Если все точки задаются заранее, до начала вычислений, – это *пассив-*

ная (параллельная) стратегия. Если эти точки выбираются последовательно в процессе поиска с учётом результатов предыдущих вычислений, – это последовательная стратегия. Примером реализации пассивной стратегии является метод *сканирования целевой функции*.

Последовательную стратегию можно реализовать следующими способами:

1) построением последовательности вложенных друг в друга интервалов, каждый из которых содержит точку минимума;

2) применением квадратичной и кубической интерполяции, где по нескольким вычисленным значениям функции строится интерполяционный полином, а его минимум указывает на очередное приближение искомой точки экстремума.

Стратегия поиска включает в себя три этапа.

Шаг 1. Выбор начального интервала неопределённости. Границы a_0, b_0 интервала должны быть такими, чтобы функция $I(x)$ была унимодальной.

Шаг 2. Уменьшение интервала неопределённости.

Шаг 3. Проверка условия окончания. Поиск заканчивается, когда длина текущего интервала неопределённости $[a_k, b_k]$ оказывается меньше установленной величины.

Ответом является множество точек, принадлежащих последнему интервалу неопределённости, среди которых каким-либо образом выбирается решение задачи x^* .

Для эвристического выбора начального интервала неопределённости можно применить алгоритм Свенна [Swann W. H.]:

1) задать произвольно следующие параметры: x^0 – некоторую точку, $t > 0$ – величину шага, положить $k = 0$;

2) вычислить значение функции в трёх точках: $x^0 - t, x^0, x^0 + t$;

3) проверить условие окончания:

а) если $I(x^0 - \Delta x) \geq I(x^0) \leq I(x^0 + t)$, то начальный интервал неопределённости найден: $[a_0, b_0] = [x^0 - t, x^0 + t]$;

б) если $I(x^0 - t) \leq I(x^0) \geq I(x^0 + t)$, то функция не является унимодальной, а требуемый интервал неопределённости не может быть найден (вычисления прекращаются и рекомендуется задать другую начальную точку x^0);

в) если условие окончания не выполняется, то перейти к шагу 4;

4) определить величину Δx :

а) если $I(x^0 - \Delta x) \geq I(x^0) \geq I(x^0 + t)$, то $\Delta x = t; a_0 = x^0; x^1 = x^0 + t; k = 1$;

б) если $I(x^0 - \Delta x) \leq I(x^0) \leq I(x^0 + t)$, то $\Delta x = -t; b_0 = x^0; x^1 = x^0 - t; k = 1$;

5) найти следующую точку $x^{k+1} = x^k + 2^k \Delta x$;

б) проверить условие убывания функции:

а) если $I(x^{k+1}) < I(x^k)$ и $\Delta x = t$, то $a_0 = x^k$; если $I(x^{k+1}) < I(x^k)$ и $\Delta x = -t$, то $b_0 = x^k$; в обоих случаях положить $k := k + 1$ и перейти к шагу 5;

б) если $I(x^{k+1}) \geq I(x^k)$, процедура завершается; при $\Delta x = t$ положить $b_0 = x^{k+1}$, а при $\Delta x = -t$ положить $a_0 = x^{k+1}$. В результате имеем $[a_0, b_0]$ – искомый начальный интервал неопределённости.

Уменьшение интервала неопределённости, осуществляемое при использовании последовательной стратегии, производится на основании вычисления в двух точках текущего интервала. Свойство унимодальности позволяет определить, в каком из возможных подынтервалов точка минимума отсутствует.

Для оценки эффективности алгоритмов уменьшения интервала неопределённости при заданном числе N вычислений функции введём критерий: характеристикой $R(N)$ относительного уменьшения начального интервала неопределённости называется отношение длины интервала, получаемого в результате N вычислений функции, к длине начального интервала неопределённости – $R(N) = \frac{|L_N|}{|L_0|}$.

Метод сканирования целевой функции: задаётся начальный интервал неопределённости $L_0 = [a_0, b_0]$ и количество вычислений целевой функции N ; вычисления производятся в N равноотстоящих друг от друга точках (при этом интервал L_0 делится на $N + 1$ равных интервалов); путём сравнения величин $I(x_i)$, $i = 1, \dots, N$, находится точка x_k^* , в которой значение функции наименьшее. Искомая точка минимума x^* считается заключённой в интервале $[x_{k-1}, x_{k+1}]$.

Для метода сканирования характеристика относительного уменьшения начального интервала неопределённости находится по формуле

$R(N) = \frac{2}{N+1}$, где N – количество вычислений функции. Если задана

величина $R(N)$, то требуемое для достижения желаемой точности количество вычислений функции определяется как наименьшее целое

число, удовлетворяющее условию $N \geq \frac{2}{R(N)} - 1$.

Метод деления интервала пополам: метод относится к последовательным стратегиям и позволяет исключить из дальнейшего рассмотрения на каждой итерации в точности половину текущего интервала неопределённости; алгоритм уменьшения интервала основан на анализе величин целевой функции в трёх точках, равномерно распределённых на текущем интервале (делящих его на четыре равные части); поиск заканчивается, когда длина текущего интервала неопределённости оказывается меньше установленной величины (рис. 4.1).



Рис. 4.1. Уменьшение интервала неопределённости при использовании метода деления интервала пополам

Для метода деления интервала пополам характеристика относительного уменьшения начального интервала неопределённости находится по формуле $R(N) = \frac{1}{2^{N/2}}$, где N – количество вычислений

функции. Средняя точка последовательно получаемых интервалов всегда совпадает с одной из трёх пробных точек, найденных на предыдущей итерации, следовательно, на каждой итерации требуются два новых вычисления функции. Если задана величина $R(N)$, то требуемое для достижения желаемой точности количество вычислений целевой функции находится как наименьшее целое, удовлетворяющее условию

$$N \geq \frac{2 \ln R(N)}{\ln 0,5}.$$

Метод дихотомии: задаются (находится методом Свенна) начальный интервал неопределённости и требуемая точность; алгоритм опирается на анализ значений функции в двух точках: для их нахождения текущий интервал неопределённости делится пополам и в обе стороны от середины откладывается по $\varepsilon/2$, где ε – малое положительное число; поиск заканчивается, когда длина текущего интервала неопределённости оказывается меньше установленной величины (рис. 4.2).

Для метода дихотомии характеристика относительного уменьшения начального интервала неопределённости находится по формуле

$$R(N) = \frac{1}{2^{N/2}},$$

можно считать эффективность методов дихотомии и деления отрезка пополам при мадых ε одинаковой.

Метод золотого сечения: в этом методе в качестве двух внутренних точек выбираются точки золотого сечения (рис. 4.3).

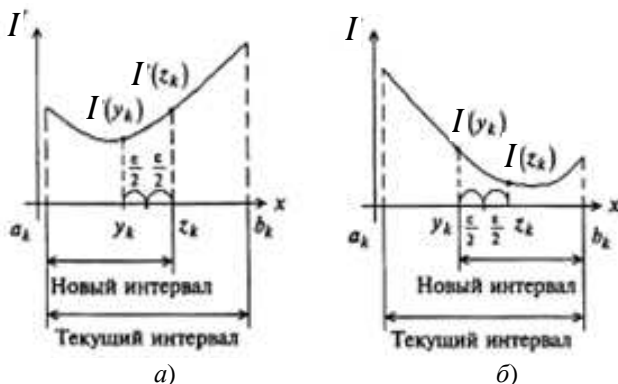


Рис. 4.2. Уменьшение интервала неопределённости при использовании метода дихотомии

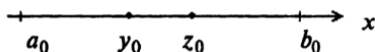


Рис. 4.3. Золотое сечение отрезка $[a_0, b_0]$

На отрезке $[a_0, b_0]$ имеются две симметричные относительно его концов точки y_0 и z_0 :

$$\frac{b_0 - a_0}{b_0 - y_0} = \frac{b_0 - y_0}{b_0 - a_0} = \frac{b_0 - a_0}{z_0 - a_0} = \frac{z_0 - a_0}{b_0 - z_0} = \frac{1 + \sqrt{5}}{2} \cong 1,618.$$

Точки y_0, z_0 производят золотое сечение отрезка, если отношение длины всего отрезка $[a_0, b_0]$ к большей части $[a_0, z_0]$ ($[y_0, b_0]$) равно отношению большей части к меньшей части. Кроме того, точка y_0 производит золотое сечение отрезка $[a_0, z_0]$, а точка z_0 – отрезка $[y_0, b_0]$.

Метод относится к последовательным стратегиям. Задаются начальный интервал неопределённости и требуемая точность. Алгоритм уменьшения интервала опирается на анализе значений функции в двух точках «золотого сечения» (рис. 4.4). С учётом свойств «золотого сечения» на каждой итерации, кроме первой, требуется только одно новое вычисление целевой функции. Условие окончания поиска стандартное: поиск заканчивается, когда длина текущего интервала неопределённости оказывается меньше установленной величины.

Для метода «золотого сечения» характеристика относительного уменьшения начального интервала неопределённости находится по формуле $R(N) = (0,618)^{N-1}$, где N – количество вычислений функции.

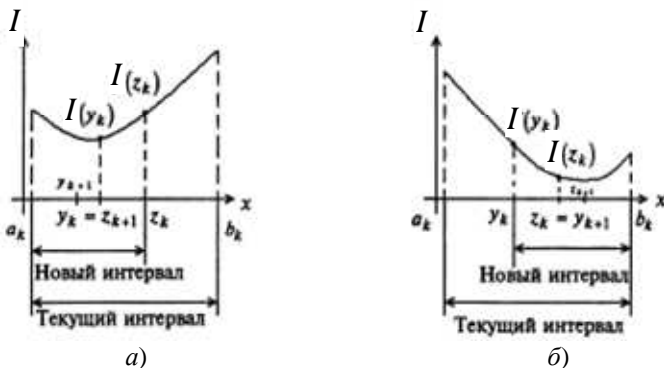


Рис. 4.4. Уменьшение интервала неопределённости при использовании метода золотого сечения

Если задана величина $R(N)$, то требуемое значение для достижения желаемой точности вычислений функции находится как наименьшее целое число, удовлетворяющее условию $N \gg 1 + \frac{\ln R(N)}{\ln 0,618}$.

Метод с использованием чисел Фибоначчи: обеспечивает максимальное сокращение интервала неопределённости при заданном количестве вычислений функции и опирается на числа Фибоначчи. Числа Фибоначчи определяются по формуле $F_0 = F_1 = 1$, $F_k = F_{k-1} + F_{k-2}$, $k = 2, 3, 4, \dots$. Последовательность чисел Фибоначчи имеет вид 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, ...

Стратегия поиска: точки вычисления находятся с использованием последовательностей $N + 1$ чисел Фибоначчи.

Алгоритм.

Шаг 1. Задать начальный интервал неопределённости $L_0 = [a_0, b_0]$; $\Delta l > 0$ – допустимую длину конечного интервала, $\varepsilon > 0$ – константу различимости.

Шаг 2. Найти количество вычислений функции как наименьшее целое число, при котором удовлетворяется условие $F_N \geq \frac{|L_0|}{\Delta l_0}$, и числа Фибоначчи F_0, F_1, \dots, F_N .

Шаг 3. Вычислить $y_0 = a_0 + \frac{F_{N-2}}{F_N}(b_0 - a_0)$; $z_0 = a_0 + \frac{F_{N-1}}{F_N}(b_0 - a_0)$.

Положить $k = 0$.

Шаг 4. Вычислить $I(y_k), I(z_k)$:

а) если $I(y_k) \leq I(z_k)$, положить $a_{k+1} = a_k$; $b_{k+1} = z_k$; $z_{k+1} = y_k$;
 $y_{k+1} = a_{k+1} + \frac{F_{N-k-3}}{F_{N-k-1}}(b_{k+1} - a_{k+1})$. Перейти к шагу 5;

б) если $I(y_k) > I(z_k)$, положить $a_{k+1} = y_k$; $b_{k+1} = b_k$; $y_{k+1} = z_k$;
 $z_{k+1} = a_{k+1} + \frac{F_{N-k-2}}{F_{N-k-1}}(b_{k+1} - a_{k+1})$.

Шаг 5. Проверить условие окончания и в случае необходимости сделать заключительное N -е вычисление функции для получения решения:

а) если $k \neq N - 3$, положить $k := k + 1$ и перейти к шагу 4;

б) если $k = N - 3$, то всегда $y_{N-1} = z_{N-2} = (a_{N-2} + b_{N-2})/2$, т.е. отсутствует точка нового вычисления функции.

Следует положить: $y_{N-1} = y_{N-2} = z_{N-2}$; $z_{N-1} = y_{N-1} + \varepsilon$. В точках y_{N-1} и z_{N-1} вычисляются значения функции и находятся границы конечного интервала неопределённости:

а) если $I(y_{N-1}) \leq I(z_{N-1})$, положить $a_{N-1} = a_{N-2}$, $b_{N-1} = z_{N-1}$;

б) если $I(y_{N-1}) > I(z_{N-1})$, положить $a_{N-1} = y_{N-1}$, $b_{N-1} = b_{N-2}$.

Процесс поиска завершается и $x^* \in [a_{N-1}, b_{N-1}]$. В качестве приближённого решения можно взять любую точку последнего интервала, например его середину $x^* \cong \frac{a_{N-1} + b_{N-1}}{2}$.

Для метода Фибоначчи характеристика относительного уменьшения начального интервала неопределённости интервала находится по формуле $R(N) = \frac{1}{F_n}$, где N – количество вычислений функции.

Метод квадратичной интерполяции (метод Пауэлла [Powell M. J. D.]): задаётся начальная точка и с помощью пробного шага находятся три точки так, чтобы они были как можно ближе к искомой точке минимума; в полученных точках вычисляются значения функции и строится полином второй степени, проходящий через имеющиеся точки; в качестве приближения точки минимума берётся точка минимума полинома; процесс поиска заканчивается, когда полученная точка отличается от наилучшей из трёх опорных точек не более чем на заданную величину.

Алгоритм.

Шаг 1. Задать начальную точку x_1 , величину шага $\Delta x > 0$, ε_1 , ε_2 – малые положительные числа, характеризующие точность.

Шаг 2. Вычислить $x_2 = x_1 + \Delta x$.

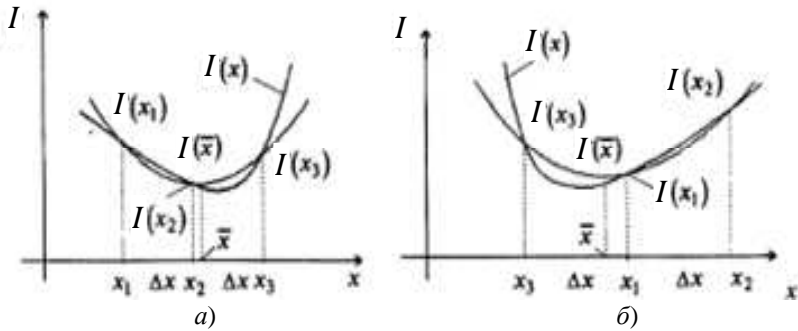


Рис 4.5. Иллюстрация к методу квадратичной интерполяции

Шаг 3. Вычислить $I(x_1) = I_1$ и $I(x_2) = I_2$.

Шаг 4. Сравнить $I(x_1)$ с $I(x_2)$:

а) если $I(x_1) > I(x_2)$, положить $x_3 = x_1 + 2\Delta x$ (рис. 4.5, а);

б) если $I(x_1) \leq I(x_2)$, положить $x_3 = x_1 - \Delta x$ (рис. 4.5, б).

Шаг 5. Вычислить $I(x_3) = I_3$.

Шаг 6. Найти $I_{\min} = \{\min I_1, I_2, I_3\}$, $x_{\min} = x_i$; $I(x_i) = I_{\min}$.

Шаг 7. Вычислить точку минимума \bar{x} интерполяционного полинома, построенного по трём точкам:

$$\bar{x} = \frac{1}{2} \frac{(x_2^2 - x_3^2)I_1 + (x_3^2 - x_1^2)I_2 + (x_1^2 - x_2^2)I_3}{(x_2 - x_3)I_1 + (x_3 - x_1)I_2 + (x_1 - x_2)I_3},$$

и значение функции $I(\bar{x})$ (рис. 4.5). Если знаменатель в формуле для \bar{x} на некоторой итерации обращается в нуль, то результатом интерполяции является прямая. В этом случае рекомендуется обозначить $x_1 = x_{\min}$ и перейти к шагу 2.

Переходим к рассмотрению численных методов многомерной минимизации: требуется найти безусловный минимум функции $I(x)$ многих переменных, т.е. найти такую точку $x^* \in R^n$, что $I(x^*) = \min_{x \in R^n} I(x)$.

Метод Хука и Дживса [R. Hoоke, N. A. Jeeves]: представляет собой комбинацию *исследующего поиска* с циклическим изменением переменных и *ускоряющего поиска по образцу*. Исследующий поиск ориентирован на выявление локального поведения целевой функции и определение направления её убывания вдоль «оврагов» [11].

Исследующий поиск начинается в некоторой начальной точке x^0 , называемой *старым базисом*. В качестве множества направлений поиска выбирается множество разных координатных направлений и пе-

ременной в процессе поиска. Фиксируется первое координатное направление и делается шаг в сторону увеличения соответствующей переменной. Если значение функции в пробной точке меньше значения функции в исходной точке, шаг считается удачным. В противном случае необходимо вернуться в предыдущую точку и сделать шаг в противоположном направлении с последующей проверкой поведения функции. После перебора всех координат исследующий поиск завершается. Полученная точка называется *новым базисом* (на рис. 4.6 в точке x^0 произведён исследующий поиск и получена точка x^1 – новый базис). Если исследующий поиск с данной величиной шага неудачен, то она уменьшается и процедура продолжается. Поиск заканчивается, когда текущая величина шага станет меньше некоторой величины.

Поиск по образцу заключается в движении по направлению от старого базиса к новому (от точки x^0 через точку x^1 через точку x^2 , из x^2 через x^3 на рис. 4.6). Величина ускоряющего шага задаётся ускоряющим множителем λ . Успех поиска по образцу определяется с помощью исследующего поиска из полученной точки (например, из точек 6, 11, 15 на рис. 4.6). Если при этом значение в наилучшей точке меньше, чем в точке предыдущего базиса, то поиск по образцу удачен (точки 6, 11 – результат удачного поиска по образцу, а точка 15 – неудачного). Если поиск по образцу неудачен, происходит возврат в новый базис, где продолжается исследующий поиск с уменьшенным шагом. На рисунке 4.6 удачный поиск отображается сплошными линиями, а неудачный – пунктирными. При поиске по i -му координатному направлению меняется только переменная x_i , а остальные переменные остаются зафиксированными.

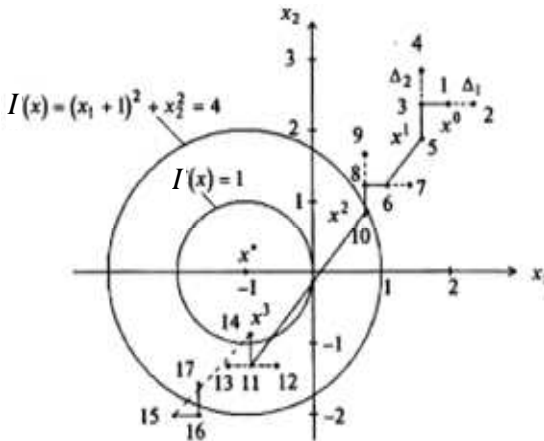


Рис. 4.6. Иллюстрация к методу Хука и Дживса

Метод деформируемого многогранника (метод Нелдера–Мида [J. A. Nelder, R. Mead]): в основу метода положено построение последовательности систем $n + 1$ точек $x^i(k)$, $i = 1, \dots, n + 1$, которые являются вершинами выпуклого многогранника. Точки системы $x^i(k + 1)$, $i = 1, \dots, n + 1$, на $k + 1$ итерации совпадают с точками системы $x^i(k)$, $i = 1, \dots, n + 1$, т.е. $I(x^h(k)) = \max_{1 \leq i \leq n+1} I(x^i(k))$. Точка $x^h(k)$ заменяется на другую точку по специальным правилам. В результате многогранники деформируются в зависимости от структуры линий уровня целевой функции, вытягиваясь вдоль длинных наклонных плоскостей, изменяя направление в изогнутых впадинах и сжимаясь в окрестности минимума. Построение последовательности многогранников заканчивается, когда значения функции в вершинах многогранника отличаются от значения функции в центре тяжести системы $x^i(k)$, $i = 1, \dots, n + 1$, $i \neq h$, не более чем на $\varepsilon > 0$.

Описание алгоритма реализации метода Нелдера–Мида и большое количество разобранных примеров приведено в книге [10].

Метод сопряжённых направлений (метод Пауэлла [M. J. D. Powell]): в этом методе используется тот факт, что минимум квадратичной функции может быть найден не более чем за n шагов при условии, что поиск ведётся вдоль сопряжённых направлений. Пусть H – симметрическая матрица размера $n \times n$; векторы d_1, d_2, \dots, d_n называются *H-сопряжёнными*, если $d_i^T H d_j = 0$, $i \neq j$. Задаётся начальная точка и направления d_1, d_2, \dots, d_n , совпадающие с координатными. Находится минимум $I(x)$ при последовательном движении по $(n + 1)$ направлениям с помощью одного из методов одномерной минимизации. При этом полученная ранее точка минимума берётся в качестве исходной для поиска по следующему направлению, а направление d_n используется как при первом ($d_0 = d_n$), так и последнем поиске. Находится новое направление поиска, сопряжённое с d_n . Оно проходит через точки, полученные при первом и последнем поиске: направление d_1 заменяется на d_2 , d_2 на d_3 и т.д. Направление d_n заменяется сопряжённым направлением, после чего повторяется поиск по $(n + 1)$ направлениям, уже не содержащим старого направления d_1 . Построение сопряжённого направления для квадратичной функции при $n = 2$ изображено на рис. 4.7, оно проходит через точки 1 и 3.

Алгоритм реализации метода сопряжённых направлений описан в работе [12].

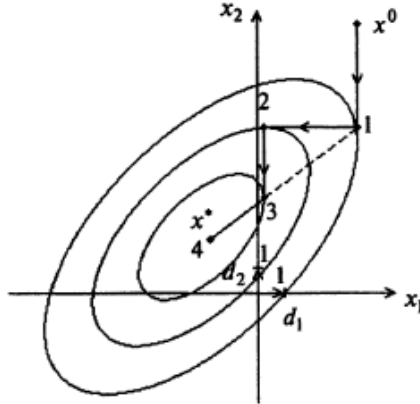


Рис. 4.7. К построению сопряжённых направлений

Методы первого порядка

Метод градиентного спуска с постоянным шагом. Точки последовательности $\{x^k\}$, $k = 0, 1, \dots$, вычисляются по правилу

$$x^{k+1} = x^k - \alpha_k \nabla I(x^k), \quad k = 0, 1, \dots,$$

где x^0 – точка, заданная пользователем; $\nabla I(x^k)$ – градиент функции $I(x)$, вычисленный в точке x^k ; α_k – величина шага, заданная пользователем и постоянная до тех пор, пока функция убывает в точках последовательности, что контролируется путём выполнения условия $I(x^{k+1}) - I(x^k) < 0$ или $I(x^{k+1}) - I(x^k) < -\varepsilon \|\nabla I(x^k)\|^2$, $0 < \varepsilon < 1$ [13]. Построение последовательности $\{x^k\}$ заканчивается в точке x^k , для которой $\|\nabla I(x^k)\| < \varepsilon_1$, где ε_1 – заданное малое положительное число, или $k \geq M$, где M – предельное число итераций, или при двукратном одновременном выполнении двух неравенств $\|x^{k+1} - x^k\| < \varepsilon_2$, $|I(x^{k+1}) - I(x^k)| < \varepsilon_2$, где ε_2 – малое положительное число.

Сходимость. Пусть функция $I(x)$ дифференцируема и ограничена снизу на R^n , а её градиент удовлетворяет условию Липшица $\|\nabla I(x) - \nabla I(y)\| \leq L \|x - y\| \quad \forall x, y \in R^n$, где $L > 0$. Тогда при произвольной начальной точке $x^0 \in R^n$ для метода градиентного спуска с постоянным шагом имеем

$$\lim_{k \rightarrow \infty} \|\nabla I(x^k)\| = 0,$$

что гарантирует сходимость последовательности $\{x^k\}$ к стационарной точке x^* , где $\nabla I(x^*) = 0$. Следовательно, найденная в результате применения метода точка x^* нуждается в дополнительном исследовании с целью её классификации. Метод градиентного спуска гарантирует сходимость последовательности $\{x^k\}$ к точке минимума для сильно-выпуклых функций [13].

Скорость сходимости. Оценки скорости сходимости получены только для сильновыпуклых функций, когда последовательность $\{x^k\}$ сходится к точке минимума $I(x)$ со скоростью геометрической прогрессии:

$$I(x^k) - I(x^*) \leq g^k(I(x^0) - I(x^*)), \quad \|x^k - x^*\| \leq C(\sqrt{g})^k,$$

где $g \in (0, 1)$, $C > 0$ – константы (рис. 4.8) [14].

Применяя алгоритм градиентного спуска с постоянным шагом, далее необходимо найти точку x^k , в которой выполнен хотя бы один из критериев окончания поиска, а затем провести анализ точки x^k с целью установить, является ли эта точка найденным приближением задачи. Процедура анализа определяется наличием у функции $I(x)$ непрерывных вторых производных; если да, то следует провести проверку выполнения достаточных условий минимума: если $H(x^*) > 0$, то точка x^k есть найденное приближение искомой точки x^* ; если $I(x)$ имеет непрерывные первые производные, то следует провести проверку функции $I(x)$ на выпуклость в окрестности x^k , используя критерий выпуклости для функций: функция $I(x)$ выпукла (строго выпукла) в том и только том случае, если $I(x + y) \geq I(x) + (\nabla I(x), y)$, если функция выпукла (строго выпукла), то x^k есть найденное приближение точки x^* . Если

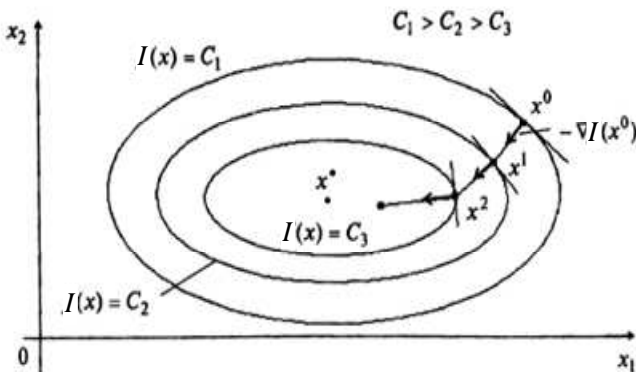


Рис. 4.8. Геометрическая интерпретация градиентного метода

требуется найти глобальный минимум функции $I(x)$, то для строго выпуклой $I(x)$ решение задачи аналогично поиску локального минимума функции, а если $I(x)$ имеет несколько локальных минимумов, поиск глобального минимума осуществляется в результате перебора всех локальных минимумов.

Метод наискорейшего градиентного спуска: стратегия поиска аналогична стратегии поиска градиентного метода за исключением выбора величины шага на каждой итерации k – она определяется из условия

$$\varphi(\alpha_k) = I(x^k - \alpha_k \nabla I(x^k)) \rightarrow \min_{\alpha_k}. \quad (4.6)$$

Решение задачи (4.6) может осуществляться с использованием необходимого условия минимума $\frac{d\varphi}{d\alpha_k} = 0$ с последующей проверкой

достаточного условия минимума $\frac{d^2\varphi}{d\alpha_k^2} > 0$. Такой путь может быть ис-

пользован либо для достаточно простой минимизируемой функции $\varphi(\alpha_k)$, либо при предварительной аппроксимации достаточно сложной функции $\varphi(\alpha_k) = I(x^k - \alpha_k \nabla I(x^k))$ полиномом $g(\alpha_k)$ (как правило, второй или третьей степени), и тогда условие $\frac{d\varphi}{d\alpha_k} = 0$ замещается усло-

вием $\frac{dg}{d\alpha_k} = 0$, а условие $\frac{d^2\varphi}{d\alpha_k^2} > 0$ – условием $\frac{d^2g}{d\alpha_k^2} > 0$.

Другой путь решения задачи (4.6) связан с использованием численных методов, когда ищется

$$\min_{\alpha_k \in [a, b]} \varphi(\alpha_k) = \min_{\alpha_k \in [a, b]} I(x^k - \alpha_k \nabla I(x^k)).$$

Границы интервала $[a, b]$ задаются пользователем. При этом степень близости найденного значения α_k к оптимальному значению α_k^* ,

удовлетворяющему условиям $\frac{d\varphi}{d\alpha_k} = 0$, $\frac{d^2\varphi}{d\alpha_k^2} > 0$, зависит от задания

интервала $[a, b]$ и точности методов одномерной минимизации [15].

Построение последовательности $\{x^k\}$, $k = 1, 2, \dots$, заканчивается как и в методе *градиентного спуска с постоянным шагом*, т.е. при $\|\nabla I(x^k)\| < \varepsilon_1$, где ε_1 – заданное малое положительное число, или $k \geq M$,

где M – предельное число итераций, или при двукратном одновременном выполнении двух неравенств $\|x^{k+1} - x^k\| < \varepsilon_2$, $|I(x^{k+1}) - I(x^k)| < \varepsilon_2$, где ε_2 – малое положительное число. Вопрос о том, может ли точка x^k рассматриваться как найденное приближение искомой точки локального минимума x^* , решается путём дополнительного исследования.

Метод наискорейшего спуска гарантирует сходимость последовательности $\{x^k\}$, $k = 1, 2, \dots$, к точке минимума для сильновыпуклых функций [13]. Оценки скорости сходимости получены только для сильновыпуклых функций, когда последовательность $\{x^k\}$ сходится к точке минимума функции $I(x)$ со скоростью геометрической прогрессии (линейная сходимость): $\|x^{k+1} - x^k\| \leq \frac{M-m}{M+m} \|x^k - x^*\|$, где M и m – оценки наибольшего и наименьшего собственных значений матрицы $H(x)$ функции $I(x)$ [13].

Метод покоординатного спуска (метод Гаусса–Зейделя [Gauss-Seidel]). Точки последовательности $\{x^k\}$, $k = 1, 2, \dots$, вычисляются по циклам в соответствии с правилом

$$x^{jk+1} = x^{jk} - \alpha_k \left(\frac{\partial I(x)}{\partial x_{k+1}} \right)_{x=x^{jk}} e_{k+1},$$

где j – номер цикла вычислений, $j = 0, 1, 2, \dots$; k – номер итерации внутри цикла, $k = 0, 1, 2, \dots, n-1$; e_{k+1} – единичный вектор; точка x^{00} задаётся пользователем, величина шага α_k выбирается из условия

$$I \left(x^{jk} - \alpha_k \left(\frac{\partial I(x)}{\partial x_{k+1}} \right)_{x=x^{jk}} e_{k+1} \right) - I(x^{jk}) < 0 \text{ или}$$

$$I(x^{jk+1}) - I(x^{jk}) < -\varepsilon \|\nabla I(x^{jk})\|^2.$$

Если выбранное условие при текущем α_k не выполняется, шаг уменьшается вдвое и точка $x^{jk} - \alpha_k \left(\frac{\partial I(x)}{\partial x_{k+1}} \right)_{x=x^{jk}} e_{k+1}$ вычисляется заново. Согласно стратегии метода Гаусса–Зейделя величина шага α_k выбирается из условия

$$\varphi(\alpha_k) = I \left(x^{jk} - \alpha_k \left(\frac{\partial I(x)}{\partial x_{k+1}} \right)_{x=x^{jk}} e_{k+1} \right) \rightarrow \min_{\alpha_k}.$$

Легко видеть, что при фиксированном j за одну итерацию с номером k изменяется только одна проекция точки x^{jk} , имеющая номер $k + 1$, а в течение всего цикла с номером j , т.е. начиная с $k = 0$ и кончая $k = n - 1$, изменяются все n проекций точки x^{j0} . После этого точке x^{jn} присваивается номер $x^{j+1,0}$ и она берётся за начальную точку для вычислений в $(j + 1)$ -м цикле. Расчёт заканчивается в точке x^{jk} при выполнении по крайней мере одного из трёх критериев окончания счёта: $\|\nabla I(x^{jk})\| < \varepsilon_1$, или $j \geq M$, или двукратного выполнения неравенств $\|x^{jk+1} - x^{jk}\| < \varepsilon_2$, $|I(x^{jk+1}) - I(x^{jk})| < \varepsilon_2$. Полученные в результате вычислений точки могут быть записаны как элементы последовательности x^l , где $l = nj + k$ – порядковый номер точки, т.е. $\{x^l\} = \{x^0 = x^{00}, x^1 = x^{01}, \dots, x^n = x^{0n} = x^{10}, x^{n+1} = x^{11}, x^{n+2} = x^{12}, \dots\}$.

Если функция $I(x)$ дифференцируема и ограничена снизу на R^n , а её градиент удовлетворяет условию Липшица, то построение последовательности $\{x^k\}$ по методу покоординатного спуска обеспечивает выполнение условия $\|\nabla I(x^k)\| \rightarrow 0$ при $k \rightarrow \infty$ [13]. Найденная в результате применения метода точка x^* нуждается в дополнительном исследовании с целью её классификации. Скорость сходимости оценивается как линейная.

Метод Флетчера–Ривса [R. Fletcher, C. M. Reeves]. Точки последовательности $\{x^k\}$ в методе вычисляются по правилу:

$$\begin{aligned} x^{k+1} &= x^k + \alpha_k s^k, \quad k = 0, 1, \dots; \\ s^k &= -\nabla I(x^k) + \beta_{k-1} s^{k-1}; \\ s^0 &= -\nabla I(x^0); \quad \beta_{k-1} = \frac{\|\nabla I(x^k)\|^2}{\|\nabla I(x^{k-1})\|^2}. \end{aligned}$$

Точка x^0 задаётся пользователем, величина шага α_k определяется для каждого значения k из условия $\varphi(\alpha_k) = I(x^k - \alpha_k s^k) \rightarrow \min_{\alpha_k}$.

Решение задачи одномерной минимизации может осуществляться либо из условия $\frac{d\varphi}{d\alpha_k} = 0$, $\frac{d^2\varphi}{d\alpha_k^2} > 0$, либо численно, с использованием методов одномерной минимизации, когда решается задача

$$\varphi(\alpha_k) \rightarrow \min_{\alpha_k \in [a, b]}.$$

Вычисление величины β_{k-1} по формуле $\beta_{k-1} = \frac{\|\nabla I(x^k)\|^2}{\|\nabla I(x^{k-1})\|^2}$ обеспе-

печивает для квадратичной формы $I(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$ построение последовательности H -сопряжённых направлений $s^0, s^1, \dots, s^k, \dots$, для которых $(s^j, Hs^i) = 0 \forall i, j = 0, 1, \dots, k; i \neq j$. При этом в точках последовательности $\{x^k\}$ градиенты функции $I(x)$ взаимно перпендикулярны, т.е. $(\nabla I(x^{k+1}), \nabla I(x^k)) = 0, k = 0, 1, \dots$.

Для квадратичных функций $I(x)$ с матрицей $H > 0$ метод Флетчера–Ривса является конечным и сходится за число шагов, не превышающее n – размерность вектора x .

Оценки скорости сходимости получены только для сильновыпуклых функций, когда последовательность $\{x^k\}$ сходится к точке минимума $I(x)$ со скоростью $\|x^{k+n} - x^*\| \leq C \|x^k - x^*\|^2, k \in 0, n, 2n, \dots$ [14].

Метод Дэвидона–Флетчера–Пауэлла [W. C. Davidon, R. Fletcher, M. J. D. Powell]. Точки последовательности $\{x^k\}$ в этом методе вычисляются по правилу

$$x^{k+1} = x^k - \alpha_k A^k \nabla I(x^k), \quad k = 0, 1, \dots,$$

где A^k – матрица размера $n \times n$, которая вычисляется по правилу:

$$A^{k+1} = A^k + A_c^k, \quad A^0 = E, \quad (4.7)$$

$$A_c^k = \frac{\Delta x^k (\Delta x^k)^T}{(\Delta x^k)^T \Delta g^k} - \frac{A^k \Delta g^k (\Delta g^k)^T A^k}{(\Delta g^k)^T A^k \Delta g^k}, \quad (4.8)$$

где $\Delta x^k = x^{k+1} - x^k, \Delta g^k = \nabla I(x^{k+1}) - \nabla I(x^k)$.

Точка x^0 задаётся пользователем, величина шага α_k определяется из условия

$$\varphi(\alpha_k) = I(x^k - \alpha_k A^k \nabla I(x^k)) \rightarrow \min_{\alpha_k}. \quad (4.9)$$

Формулы (4.7), (4.8) при аналитическом решении задачи (4.9) обеспечивают построение последовательности $\{A^k\}$ положительно определённых матриц, таких, что $A^k \rightarrow H^{-1}(x^*)$ при $k \rightarrow \infty$. Следствием этого для квадратичной функции $I(x) = \frac{1}{2}(Hx, x) + (b, x), H > 0$, является тот факт, что направления $s^k, k = 0, 1, \dots$, будут H -сопряжёнными, и, следовательно, алгоритм сойдётся не более чем за n шагов. Для не-

квадратичных функций $I(x)$ алгоритм перестаёт быть конечным, и его сходимость зависит от точности решения задачи (4.9). Глобальную сходимость алгоритма можно гарантировать лишь при его обновлении через каждые n шагов, т.е. когда в формуле $x^{k+1} = x^k - \alpha_k A^k \nabla I(x^k)$, $k = 0, 1, \dots$:

$$A^k = \begin{cases} E, & k \in j, \quad j = \{0, n, 2n, \dots\} \\ A^{k-1} + A_c^k, & k \notin j. \end{cases}$$

Если $I(x)$ дважды непрерывно дифференцируемая функция и $H(x^*) > 0$, то метод с обновлением сходится к точке локального минимума x^* со сверхлинейной скоростью [14], а если в дополнение выполняется условие $\|H(x)y\| \leq k\|y\| \quad \forall y \in R^n$ в окрестности точки x^* , то последовательность $\{x^k\}$ сходится к точке x^* с квадратичной скоростью [14].

Методы второго порядка

Метод Ньютона [I. Newton]. Направление спуска в методе Ньютона определяются для каждого значения k по формуле

$$s^k = -H^{-1}(x^k) \nabla I(x^k).$$

Вычисление s^k по предыдущей формуле гарантирует выполнение требования $I(x^{k+1}) < I(x^k)$ при условии, что $H(x^k) > 0$.

Пусть $I(x)$ дважды дифференцируемая сильновыпуклая функция с константой $l > 0$ на R^n и удовлетворяет условию $\|H(x) - H(y)\| \leq L\|x - y\| \quad \forall x, y \in R^n$, где $L > 0$, а начальная точка такова, что

$\|\nabla I(x^0)\| \leq \frac{8l^2}{L}$, т.е. $\|\nabla I(x^0)\| \leq \frac{8l^2 q}{L}$, где $q \in (0, 1)$. Тогда последовательность $\{x^k\}$ сходится к точке минимума с квадратичной скоростью $\|x^k - x^*\| \leq \frac{4lq^2}{L}$ [14].

Сходимость метода Ньютона доказана лишь для сильновыпуклых функций и для достаточно хорошего начального приближения, определяемого условием $\|\nabla I(x^0)\| \leq \frac{8l^2 q}{L}$, практическое использование которого крайне затруднено, так как постоянные l и L , как правило, неизвестны и требуют трудоёмкого исследования для их определения. Поэтому при практическом использовании метода Ньютона следует:

а) анализировать матрицу $H(x^k)$ на выполнение условия $H(x^k) > 0 \quad \forall k = 0, 1, \dots$ и заменять формулу $x^{k+1} = x^k - H^{-1}(x^k) \nabla I(x^k)$ на формулу $x^{k+1} = x^k - \alpha_k \nabla I(x^k)$ в случае его невыполнения;

б) проводить анализ точки x^k в целях выяснения, является ли она найденным приближением искомой точки x^* .

В методе Ньютона–Рафсона [Newton–Raphson] точки последовательности вычисляются по правилу

$$x^{k+1} = x^k - \alpha_k H^{-1}(x^k) \nabla I(x^k), \quad k = 0, 1, \dots,$$

здесь x^0 задаётся пользователем, а величина шага α_k определяется из условия

$$\varphi(\alpha_k) = I(x^k - \alpha_k H^{-1}(x^k) \nabla I(x^k)) \rightarrow \min_{\alpha_k}. \quad (4.10)$$

Задача (4.10) может решаться либо аналитически с использованием необходимого условия минимума $\frac{d\varphi}{d\alpha_k} = 0$ с последующей провер-

кой достаточного условия $\frac{d^2\varphi}{d\alpha_k^2} > 0$, либо численно как задача

$\varphi(\alpha_k) \rightarrow \min_{\alpha_k \in [a, b]}$, где интервал $[a, b]$ задаётся пользователем. Если

функция $\varphi(\alpha_k)$ достаточно сложна, то возможна её замена полиномом $P(\alpha_k)$ второй или третьей степени и тогда шаг α_k может быть опреде-

лён из условия $\frac{dP}{d\alpha_k} = 0$ при выполнении условия $\frac{d^2P}{d\alpha_k^2} > 0$.

4.5. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ УСЛОВНЫХ ОПТИМИЗАЦИОННЫХ ЗАДАЧ

Рассмотрим общую постановку задачи поиска условного экстремума со смешанными ограничениями. Даны дважды непрерывно дифференцируемые целевая функция $I(x) = I(x_1, \dots, x_n)$ и функции ограничений $g_j(x) = 0, j = 1, \dots, m; g_j(x) \leq 0, j = m + 1, \dots, p$, определяющие множество допустимых решений X . Требуется найти локальный минимум целевой функции на множестве X , т.е. такую точку $x^* \in X$, что

$$I(x^*) = \min_{x \in X} I(x),$$

где $X = \left\{ x \left\{ \begin{array}{l} g_j(x) = 0, \quad j = 1, \dots, m, \quad m < n; \\ g_j(x) \leq 0, \quad j = m + 1, \dots, p. \end{array} \right. \right\}$

Применение необходимых и достаточных условий применения условного экстремума эффективно для решения ограниченного числа задач, в которых вытекающие из этих условий соотношения имеют аналитическое решение. Для решения большинства практических задач используются численные методы, которые делятся на две группы.

1. Методы, использующие преобразование задачи условной оптимизации в последовательность задач безусловной оптимизации путём введения в рассмотрение вспомогательных функций: *методы последовательной безусловной оптимизации*.

2. Методы непосредственного решения задачи условной оптимизации, основанные на движении из одной допустимой точки, где выполнены все ограничения, к другой допустимой точке с лучшим значением целевой функции: *методы возможных направлений*.

Основная идея методов первой группы состоит в том, чтобы аппроксимировать исходную задачу условной оптимизации некоторой вспомогательной задачей, решение которой менее сложно, чем решение исходной. Естественно, что ограничившись одной вспомогательной задачей, можно получить, вообще говоря, лишь приближённое решение. Если же использовать последовательность задач, в определённом смысле «сходящихся» к исходной, то искомое точное решение в большинстве случаев окажется пределом соответствующей последовательности приближённых решений. Идея преобразования задачи с ограничениями в надлежащим образом построенную последовательность задач без ограничений представляется заманчивой главным образом в связи с наличием эффективных и надёжных методов безусловной минимизации.

На практике для получения решения исходной задачи с требуемой точностью достаточно бывает решить конечное (относительно небольшое) число вспомогательных задач. При этом нет необходимости решать их точно, а информацию, полученную в результате решения вспомогательной задачи, обычно удаётся эффективно использовать для решения следующей.

В рамках единой методологии можно выделить несколько подходов к решению задачи. Первый называется *методом штрафов (внешних штрафов)*. В этом методе к целевой функции добавляется функция, интерпретируемая как штраф за нарушение каждого из ограничений. Метод генерирует последовательность точек, которая сходится к решению исходной задачи.

Второй подход называется *методом барьеров (внутренних штрафов)*. Здесь к целевой функции исходной задачи добавляется слабое, которое не позволяет генерируемым точкам выходить за пределы допустимой области.

Третий подход связан с добавлением штрафной функции не к целевой функции, а к её функции Лагранжа. В результате возникает *модифицированная функция Лагранжа*, а методы, использующие эту функцию, называются *методами множителей*.

Четвёртый подход базируется на введении так называемых точных штрафных функций, позволяющих ограничиться решением лишь одной задачи безусловной минимизации.

Методы непосредственного решения задач условной оптимизации, образующие вторую группу, связаны с нахождением предела x^* последовательности $\{x^k\}$ допустимых точек при $k \rightarrow \infty$, таких, что $I(x^{k+1}) < I(x^k)$, $k = 0, 1, \dots$. К описанной группе методов относятся *метод проекции градиента* и *метод возможных направлений Зойтендейка*.

Упомянутая выше группа методов решения условных оптимизационных задач и большое количество разобранных примеров приведены в работах [10, 12, 16].

4.6. ОПТИМИЗАЦИЯ БИОТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА В УСЛОВИЯХ ЧАСТИЧНОЙ НЕОПРЕДЕЛЁННОСТИ

При проектировании/расчёте биотехнологических процессов (БТП) часто используют так называемую неопределённую («неточную») информацию о физико-химических параметрах, характеристиках исходных/«сырьёвых» веществ, конструктивных/геометрических показателях фрагментов технологического оборудования, внешних условиях эксплуатации процесса, финансовых/стоимостных данных и т.п., что опосредованно оказывает заметное влияние на качество функционирования БТП, характеризуемое известными технико-экономическими показателями, например прибылью. Количественный учёт ряда указанных «неточных»/случайных факторов составляет суть проектирования БТП в условиях частичной (неполной) неопределённости.

Пусть проектируемый БТП описывается векторным нелинейным уравнением

$$f(x, d, y, u, a) = 0, \quad (4.11)$$

где $d = \{d_l, l = \overline{1, p}\} \in D$, $a = \{a_\rho, \rho = \overline{1, r}\} \in A$ – векторы конструктивных и физико-химических параметров из замкнутых множеств D и A с неиспользуемыми замыканиями \bar{d}^- , \bar{d}^+ и \bar{a}^- и \bar{a}^+ соответственно.

Рассмотрим более подробно аргументы функции $f(\cdot)$ и их взаимосвязи. Координату x можно представить в форме $x_n^p + \Delta x$, где x_n^p – заданная нагрузка по «сырью» на БТП; Δx – неопределённая знакопеременная составляющая, такая, что $\|\Delta x\| < \|x_n^p\|$ или даже $\|\Delta x\| \ll \|x_n^p\|$; вектор-параметр a также можно представить в форме $\bar{a} + \Delta a$, где \bar{a} – «точная» или заданная составляющая, а Δa зависит от случайной ком-

поненты Δx , изменений показателей внешней среды (температуры, влажности, давления и т.д.) и иногда от вариаций Δd конструктивных параметров $d = \bar{d} + \Delta d$. Будем упрощённо считать Δx и x_n^p , а также Δa и \bar{a} , Δd и \bar{d} линейно независимыми, а x и d независимыми друг от друга (хотя Δd может влиять на Δa), но Δd существует только на этапе проектирования, а Δx и Δa – на всём жизненном цикле БТП. Управление u также существенно изменяется и влияет на y на этапе проектирования БТП, а на стадии функционирования $u = \text{const}$, что обеспечивается соответствующими автоматическими системами регулирования.

Для удобства изложения материала объединим неопределённые параметры $x_n^p + \Delta x$ и $\bar{a} + \Delta a$ в вектор ξ из ограниченного открытого числового множества $\Xi = \{\xi_p : \xi_p^- < \xi_p < \xi_p^+\}$, $j = \overline{1, n_\xi}$.

Выбор множества параметров, которые должны рассматриваться как неопределённые, очень важен: выбор большого числа параметров в качестве неопределённых может сделать задачу математического моделирования БТП в условиях неопределённости крайне трудоёмкой, а невключение некоторых параметров в число неопределённых может значительно повлиять на результаты её решения. Для выбора таких параметров, которые в наибольшей степени влияют на процесс, применяют методы анализа чувствительности математической модели объекта.

Обозначим через ξ , y и $\bar{\xi}$, \bar{y} совокупности неопределённых параметров и выходных координат объекта, которые соответственно могут и не могут быть непосредственно измерены на этапе функционирования БТП. Предположим, что мы имеем достаточно точные приборы для измерения значений ξ , y , которые проводятся при некотором $u = u^*$. Если размерность вектор-функции $f(\cdot)$ с компонентами в виде функций $f_i(x, d, y, u, \xi, \bar{\xi})$, $i = \overline{1, n}$, больше или равна размерности вектора $\bar{\xi}$, то значения параметров $\bar{\xi}$ на этапе функционирования могут быть найдены решением обратной задачи

$$\min_{\bar{\xi}} \left\{ \left[f(x, d, y, u, \xi, \bar{\xi}) \right]^T f(x, d, y, u, \xi, \bar{\xi}) \right\}.$$

Интервалы неопределённости измеряемых параметров могут быть найдены, если известны максимальные ошибки измерения применяемых датчиков и приборов [19]. Отметим, что решение $\bar{\xi}$ обратной за-

дачи при известных ошибках $[\Delta\xi^j, \Delta y^j]$ измерения неопределённых параметров и выходных координат объекта ξ, y – это некоторая функция от $\Delta\xi^j$ и Δy^j , т.е. $\bar{\xi} = \bar{\xi}(\Delta\xi^1, \Delta\xi^2, \dots, \Delta\xi^N, \Delta y^1, \Delta y^2, \dots, \Delta y^N)$. В общем случае область неопределённости параметров ξ задаётся следующими условиями:

$$\Xi = \{\xi(\Delta\xi^1, \Delta\xi^2, \dots, \Delta\xi^N, \Delta y^1, \Delta y^2, \dots, \Delta y^N) : |\Delta\xi^j| \leq \delta_1, |\Delta y^j| \leq \delta_2\},$$

где δ_1, δ_2 – максимальные значения погрешности измерения ξ и y .

Использование области неопределённости в таком виде очень неудобно при решении задач оптимизации. В связи с этим в качестве области неопределённости может оказаться целесообразным использовать наименьший многомерный прямоугольник, включающий область Ξ . Для этого необходимо найти n_ξ пар $[\xi_i^L, \xi_i^U]$ решением следующих задач:

$$\xi_i^L = \min_{\Delta\xi^j, \Delta y^j} \xi_i(\Delta\xi^1, \Delta\xi^2, \dots, \Delta\xi^N, \Delta y^1, \Delta y^2, \dots, \Delta y^N);$$

$$\xi_i^U = \max_{\Delta\xi^j, \Delta y^j} \xi_i(\Delta\xi^1, \Delta\xi^2, \dots, \Delta\xi^N, \Delta y^1, \Delta y^2, \dots, \Delta y^N);$$

$$|\Delta\xi^j| \leq \delta_1, \quad |\Delta y^j| \leq \delta_2, \quad j = \overline{1, N}.$$

Для обеспечения безопасности БТП и заданного качества продукции используют «жёсткие» ограничения типа $g_j(x_n^p, d, y, u, \xi) \leq 0$, $j = \overline{1, m}$, которые должны выполняться для любых значений аргументов d, y, u и даже ξ , а также «мягкие» ограничения, выполняемые с заданной доверительной вероятностью P_d , т.е.

$$\text{Вер} \{g_j(x_n^p, d, y, u, \xi) \leq 0, \quad j = \overline{1, m}\} \geq P_d.$$

Рассмотрим некоторые постановки оптимизационных задач математического проектирования БТП в условиях частичной неопределённости и основные подходы к их решению.

Одноэтапная задача математического проектирования

Характерной особенностью одноэтапной задачи проектирования БТП является невозможность уточнения информации относительно неопределённых параметров $\{\Delta x, \Delta a\}$ и, соответственно, $\xi = \{\xi_\rho, \rho = \overline{1, n_\xi}\}$, а конструктивные параметры $d \in D$ и управления $u \in U$ равноправны в том смысле, что они не изменяются на этапе функционирования объекта.

Для упрощения задачи проектирования БТП примем $n = 3$, $d = \{d_1, d_2, d_3\} \in D$, $\{D = d : d_l^- \leq d_l \leq d_l^+, l = \overline{1, 3}\}$, $u = \{u_1, u_2, u_3\} \in U$, $\{U = u : u_j^- \leq u_j \leq u_j^+, j = \overline{1, 3}\}$, $\xi = \Xi = \{\xi_k : \xi_k^- < \xi_k < \xi_k^+, k = \overline{1, 2}\}$, ограничения $g_\lambda(x_n^p, d, y, u, \xi) = g_\lambda(\cdot) \leq 0$, $\lambda = 1, 2$ – гладкие, выпуклые; технико-экономические показатели типа прибыли $\phi(x_n^p, d, y, u, \xi)$ – выпуклые и гладкие.

Поскольку ограничения являются жёсткими, то при наличии неопределённости необходимо гарантировать работоспособность технологического объекта, т.е. гарантировать наличие таких значений векторов $d \in D$ и $u \in U$, которые обеспечивают выполнение ограничений $g_\lambda(x_n^p, d, y, u, \xi) = g_\lambda(\cdot) \leq 0$, $\lambda = 1, 2$ для всех возможных значений $\xi = \{\xi_k, k = \overline{1, 2}\} \in \Xi$. Запишем логическое условие работоспособности БТП в виде

$$\forall \xi \in \Xi, \forall \lambda = 1, 2 \quad g_\lambda(x_n^p, d, y, u, \xi) \leq 0,$$

которое в соответствии с соотношением эквивалентности

$$\max_{z \in Z} \psi(z) \leq 0 \Leftrightarrow \psi(z) \leq 0, \quad z \in Z$$

можно преобразовать в аналитическое условие вида

$$\max_{\xi \in \Xi} g_\lambda(x_n^p, d, y, u, \xi) \leq 0, \quad \lambda = 1, 2.$$

Поскольку математическое ожидание $M_\xi(\phi(x_n^p, d, y, u, \xi))$ даёт среднее значение прибыли $\phi(x_n^p, d, y, u, \xi)$ на этапе функционирования БТП, то естественно использовать выражение $I(d, u) = M_\xi(\phi(x_n^p, d, y, u, \xi))$ как целевую функцию задачи оптимизации в условиях неопределённости.

Сформулируем одноэтапную задачу математического проектирования с жёсткими ограничениями в условиях неопределённости параметров $\xi \in \Xi \subset \Xi^2$: при заданной нагрузке x_n^p по сырью требуется определить максимальную прибыль $I^*(d, u)$, векторы конструктивных параметров $d^* \in D$ и управления $u^* \in U$, такие, что

$$I^*(d, u) = \max_{d, u} \{M_\xi(\phi(x_n^p, d, y, u, \xi))\} \quad (4.12)$$

при связи $f(x_n^p, d, y, u, \xi) = 0$ и ограничениях:

$$\max_{\xi \in \Xi} g_\lambda(x_n^p, d, y, u, \xi) \leq 0, \quad \lambda = 1, 2. \quad (4.13)$$

Если применять «мягкие» ограничения, то вместо ограничений (4.13) следует использовать вероятностное условие

$$\text{Вер}\{g_\lambda(x_n^p, d, y, u, \xi) \leq 0\} \geq P_d, \lambda = 1, 2. \quad (4.14)$$

Сформулированные задачи оптимизации (4.12), (4.11), (4.13) или (4.12), (4.11), (4.14) аналогичны задаче оптимизации установившегося режима БТП и для их решения можно использовать методы, рекомендованные в гл. 3.

Главной трудностью в этих задачах является вычисление многомерного интеграла (математического ожидания $M_\xi(\varphi(x_n^p, d, y, u, \xi))$ и вероятностей $\text{Вер}\{g_\lambda(\cdot) \leq 0\}$ выполнения ограничений $\{g_\lambda(x_n^p, d, y, u, \xi) \leq 0\}$, $\lambda = 1, 2$). Можно рекомендовать ряд методов их экономичного вычисления, программные реализации которых имеются в пакетах прикладных программ: методы, основанные на аппроксимации подынтегрального выражения с помощью конечного множества точек в пространстве интегрирования, методы Монте-Карло и близкие к ним процедуры латинского гиперкуба и последовательности проб Хаммерслея, основанные на генерации псевдослучайных чисел (чаще всего с равномерным законом распределения).

В первой группе методов вычисление $M_\xi(\varphi(x_n^p, d, y, u, \xi))$ осуществляется по приближённой формуле

$$I(d, u) = M_\xi(\varphi(x_n^p, d, y, u, \xi)) \approx \sum_{j \in J_1} \omega_j \varphi(x_n^p, d, y, u, \xi^j),$$

где ω_j – весовые коэффициенты, удовлетворяющие условиям $\omega_j \geq 0$, $\sum_{j \in J_1} \omega_j = 1$ (далее мы будем опускать эти условия); ξ^j ($j \in J_1$) – аппроксимационные точки, равномерно покрывающие область неопределённости $\Xi = \{\xi_k : \xi_k^- < \xi_k < \xi_k^+, k = 1, 2\}$. В соответствии с техникой Монте-Карло для оценки интеграла используется следующее соотношение:

$$M_\xi(\varphi(x_n^p, d, y, u, \xi)) \approx N^{-1} \sum_{j \in J_1} \varphi(x_n^p, d, y, u, \xi^j),$$

где ξ^j ($j \in J_1$) – последовательность N векторов ξ^j , имеющих равномерную плотность распределения вероятностей.

Опишем один из возможных итерационных подходов к решению одноэтапной задачи (4.12), (4.11), (4.13) математического проектирования БТП в условиях неопределённости параметров $\xi \in \Xi \subset E^2$. Ап-

приори введём множества S_1 аппроксимационных $S_1 = \{\xi^j : \xi^j \in \Xi, j \in J_1\}$ и S_2 «критических» точек, в которых могут нарушаться ограничения задачи $\max_{\xi \in \Xi} g_\lambda(x_n^p, d, y, u, \xi) \leq 0, \lambda = 1, 2$.

Поскольку функции $g_\lambda(\cdot), \lambda = 1, 2$, выпуклы, то в начальное множество критических точек $S_2^{(0)}$ целесообразно включить угловые точки $\xi_k^-, \xi_k^+, k = 1, 2$, области неопределённости (прямоугольника) $\Xi \subset E^2$. Сформулируем вспомогательную задачу (А): определить максимальное значение функции $\hat{I}(d, u)$, векторы конструктивных параметров $\hat{d} \in D$ и управлений $\hat{u} \in U$, такие, что

$$\hat{I}(d, u) = \max_{d, u} \sum_{j \in J_1} \omega_j \Phi(x_n^p, d, y, u, \xi^j) \quad (\text{А})$$

при связи $f(x_n^p, d, y, u, \xi) = 0$ и ограничениях:

$$g_1(x_n^p, d, y, u, \xi^j) \leq 0, g_2(x_n^p, d, y, u, \xi^j) \leq 0, \xi^j \in S_1, j \in J_1;$$

$$g_1(x_n^p, d, y, u, \xi^l) \leq 0, g_2(x_n^p, d, y, u, \xi^l) \leq 0, \xi^l \in S_2, l \in J_2.$$

Для решения задачи (4.12), (4.11), (4.13) покажем следующий алгоритм.

На первом шаге задаём начальный номер итерации, например $v = 1$, совокупность аппроксимационных точек $\xi^j, j \in J_1, \xi^j \in S_1$, начальную совокупность критических точек $S_2^{(v-1)} = \{\xi^l : \xi^l \in \Xi, l \in J_2^{(v-1)}\}$ и начальные приближения конструктивных параметров $d^{(0)}$ и управляющих переменных $u^{(0)}$.

На втором шаге находим решение вспомогательной задачи (А) одним из методов, например, дихотомии, золотого сечения и др., и определяем значения $I(d^{(v)}, u^{(v)})$ и векторов $d^{(v)}, u^{(v)}$.

Далее на третьем шаге решаем две экстремальные задачи: $\max_{\xi \in \Xi} g_\lambda(x_n^p, d^{(v)}, y, u^{(v)}, \xi) \leq 0, \lambda = 1, 2$, и определяем две точки $\xi_1^{(v)}$ и $\xi_2^{(v)}$, доставляющие максимум функциям $g_\lambda(x_n^p, d^{(v)}, y, u^{(v)}, \xi) \leq 0, \lambda = 1, 2$.

На четвёртом шаге образуем новое множество критических точек

$$R^{(v)} = \{\xi_1^{(v)}, \xi_2^{(v)} : g_\lambda(x_n^p, d^{(v)}, y, u^{(v)}, \xi) > 0, \lambda = 1, 2.$$

Если это множество пустое, то решение задачи получено, т.е. $d^* = d^{(v)}, u^* = u^{(v)}$ и алгоритм заканчивает свою работу, иначе переходим к следующему шагу.

На пятом шаге формируем новое множество критических точек $S_2^{(v)} = S_2^{(v-1)} \cup R^{(v)}$ и, полагая $v: v + 1$, переходим к шагу 2.

Определённым недостатком описанного алгоритма решения задачи (4.12), (4.11), (4.13) является увеличение числа критических точек на каждой итерации и, соответственно, числа учитываемых ограничений. В некоторых случаях при большом числе критических точек число ограничений может стать слишком большим.

Рассмотрим один из возможных эвристических подходов к решению одноэтапной задачи математического проектирования БТП с «мягкими» ограничениями (4.12), (4.11), (4.14), предварительно переформулировав её в эквивалентную форму:

$$I^*(d_{\alpha^*}, u_{\alpha^*}) = \min_{\alpha \in \Lambda} \left(\max_{d, u} \sum_{i \in J_1} \omega_i \varphi(x_{\text{н}}^{\text{п}}, d, y, u, \xi^i) g_{\lambda}(x_{\text{н}}^{\text{п}}, d, y, u, \xi) \leq \alpha_{\lambda}, \quad (\text{Б}) \right. \\ \left. \lambda = 1, 2 \right), \\ \Lambda = \left\{ \alpha \mid \forall \lambda \text{ Вер} \left[g_{\lambda}(x_{\text{н}}^{\text{п}}, d_{\alpha}, y, u_{\alpha}, \xi) \leq 0 \right] \geq P_{\lambda}, \lambda = 1, 2 \right\}.$$

Напомним, что черта «|» означает «при условии», а символ \forall – «для любого».

В задаче (Б) вероятностные ограничения (4.14) заменяются на жёсткие ограничения вида $g_{\lambda}(x_{\text{н}}^{\text{п}}, d, y, u, \xi) \leq \alpha_{\lambda}$, $\lambda = 1, 2$, где α_1 , α_2 – скалярные переменные ($\alpha_1 < 0$, $\alpha_2 < 0$), а далее подбираются минимальные значения $|\alpha_1^*|$, $|\alpha_2^*|$ и соответствующие им конструктивные параметры d_{α^*} и управления u_{α^*} , такие, при которых выполняются вероятностные ограничения (4.14).

Покажем итерационный алгоритм для решения задачи (4.12), (4.11), (4.14).

На первом шаге задаём начальный номер итерации, например $v = 1$, совокупность аппроксимационных точек ξ^i , $i \in J_1$, $\xi^i \in S_1$, значение вероятности P_{λ} , точности ε решения задачи оптимизации, начальные приближения $d^{(0)}$, $u^{(0)}$ и $\alpha^{(0)} = (\alpha_1^{(0)}, \alpha_2^{(0)})$.

На втором шаге решаем задачу (Б) одним из методов, например, дихотомии, золотого сечения и др., и определяем значения векторов $d^{(v)} = d_{\alpha^{(v-1)}}$, $u^{(v)} = u_{\alpha^{(v-1)}}$ и $I(d^{(v)}, u^{(v)})$.

На шаге 3 в точке $d^{(v)}$, $u^{(v)}$ методом Монте-Карло вычисляются вероятности выполнения ограничений $g_{\lambda}(x_{\text{н}}^{\text{п}}, d^{(v)}, y, u^{(v)}, \xi) \leq 0$, $\lambda = 1, 2$,

с использованием генератора псевдослучайных чисел ξ_k , $k = 1, 2$, с равномерным законом распределения и математической модели $f(x_H^p, d^{(v)}, y, u^{(v)}, \xi) = 0$ и проверяется выполнение условий $\text{Вер}\{g_\lambda(x_H^p, d, y, u, \xi) \leq 0\} \geq P_d$, $\lambda = 1, 2$.

Если вероятностные ограничения не выполняются, т.е. $\alpha_\lambda^{(v-1)} \notin \Lambda$, $\lambda = 1, 2$, то производится «ужесточение» ограничений во вспомогательной задаче (Б) посредством увеличения заданных скалярных переменных $|\alpha_1^{(v-1)}|$ или $|\alpha_2^{(v-1)}|$ для нарушенных ограничений, т.е. определяются новые значения скалярных переменных $\alpha^{(v)} = (\alpha_1^{(v)}, \alpha_2^{(v)})$, число итераций v увеличивается на 1, т.е. $v := v + 1$, и следует переход к шагу 2. В противном случае с заданной точностью ε подбираются минимальные числа $|\alpha_1^{(v)*}|$, $|\alpha_2^{(v)*}|$, принадлежащие допустимой области Λ , а соответствующие им конструктивные параметры $d^* = d_{\alpha^{(v)*}}$, управления $u^* = u_{\alpha^{(v)*}}$ и значение целевой функции $I^*(d_{\alpha^{(v)*}}, u_{\alpha^{(v)*}})$ принимаются за решение задачи (4.12), (4.11), (4.14).

Рассмотрим теперь другую формулировку одноэтапной задачи математического проектирования ТП, в которой в качестве критерия будет использоваться нижняя граница α прибыли $\varphi(x_H^p, d, y, u, \xi)$, которая не может быть нарушена с заданной вероятностью P_d . Перепишем задачу с «мягкими» ограничениями в виде

$$I^*(d, u) = \max_{d, u, \alpha} \alpha \quad (4.15)$$

при связи $f(d, y, u, \xi) = 0$ и ограничениях:

$$\text{Вер}\{g_0 = \alpha - \varphi(x_H^p, d, y, u, \xi) \leq 0\} \geq P_d, \quad (4.16)$$

$$\text{Вер}\{g_\lambda(x_H^p, d, y, u, \xi) < 0\} \geq P_d, \lambda = 1, 2, \quad (4.17)$$

где α – скалярная переменная; P_d – заданное значение вероятности выполнения ограничений.

В задаче (4.15) – (4.17) мы ищем наибольшее значение α^* переменной α , для которой условие (4.16) удовлетворялось бы с заданной вероятностью P_d . Таким образом, решение задачи (4.15) – (4.17) позволяет определить d^* и u^* , при которых гарантируется, что в течение всего этапа функционирования целевая функция $\varphi(x_H^p, d^*, y, u^*, \xi)$ будет не меньше, чем α^* с вероятностью P_d .

Отметим, что существенным недостатком постановок одноэтапных оптимизационных задач в условиях частичной неопределённости является то, что при определении оптимальных управлений u^* на этапе проектирования не учитывается потенциальная возможность уточнения вектора ξ и, соответственно, u^* на этапе функционирования ТП. Такую возможность предоставляют постановки двухэтапных оптимизационных задач математического проектирования ТП [19].

Двухэтапная задача математического проектирования

При постановке двухэтапной задачи математического проектирования БТП в условиях неопределённости будем предполагать, что на этапе функционирования объекта конструктивные параметры $d \in D$ по-прежнему постоянны, а неопределённые параметры $\{x_n^p + \Delta x, \bar{a} + \Delta a\}$ и, соответственно, $\xi = \{\xi_k, k = \overline{1, n_\xi}\}$ либо могут быть точно измерены, либо оценены с использованием доступной экспериментальной информации.

Сформулируем внутреннюю задачу оптимизации установившихся режимов БТП при фиксированных значениях конструктивных $d \in D$ и неопределённых $\xi \in \Xi$ параметров: требуется определить вектор управлений $\hat{u} \in U$, такой, что $\hat{\phi}(x_n^p, d, y, \hat{u}, \xi) = \max_{u \in U} \phi(x_n^p, d, y, u, \xi)$ при связи $f(x_n^p, d, y, u, \xi) = 0$ и ограничениях $g_\lambda(x_n^p, d, y, u, \xi), \lambda = 1, 2$. Внутренняя задача разрешима, устойчива и имеет одно или сколь угодно много решений \hat{u} (но при выпуклой функции $\phi(x_n^p, d, y, u, \xi)$ $\hat{\phi}$ – единственный $\max \phi(\hat{u})$), т.е. эта задача – аргументная; если $\phi(u)$ – невыпукла, то имеем много стационарных точек u , а за решение выбираем управление \hat{u} , на котором $\hat{\phi}(\hat{u})$ максимальна.

Поскольку на этапе функционирования объекта значение целевой функции оптимизации установившихся режимов будет равно

$$\hat{\phi}(x_n^p, d, y, \hat{u}, \xi) = \max_{u \in U} \phi(x_n^p, d, y, u, \xi),$$

то на этапе проектирования ТП можно оценить эффективность его будущей работы, подсчитав оценку математического ожидания $M_\xi(\hat{\phi}(x_n^p, d, y, \hat{u}, \xi))$ от максимальной прибыли $\hat{\phi}(x_n^p, d, y, \hat{u}, \xi)$ при фиксированном значении вектора ξ . Эту величину будем далее использовать как целевую функцию двухэтапного проектирования БТП в условиях неопределённости параметров $\{\Delta x, \Delta a\}$.

Пусть, как и в одноэтапной задаче, $n = 3$, $d = \{d_1, d_2, d_3\} \in D = \{d : d_l^- \leq d_l \leq d_l^+, l = \overline{1, 3}\}$, $u = \{u_1, u_2, u_3\} \in U = \{u : u_j^- \leq u_j \leq u_j^+, j = \overline{1, 3}\}$, $\xi \in \Xi = \{\xi_k : \xi_k^- < \xi_k < \xi_k^+, k = \overline{1, 2}\}$, имеются выпуклые ограничения $g_\lambda(x_H^p, d, y, u, \xi) = g_\lambda(\cdot) \leq 0$, $\lambda = 1, 2$, функция $I(d, u)$ – гладкая, выпуклая. Сформулируем двухэтапную задачу математического проектирования БТП: при заданной нагрузке x_H^p по сырью требуется определить максимальную прибыль $I^*(d, u)$, векторы конструктивных параметров $d^* \in D$ и управляющих переменных $u^*(\xi) \in U$ БТП, такие, что

$$I^*(d, u) = \max_{d, u(\xi)} \Phi(x_H^p, d, y, u, \xi) \quad (4.18)$$

$\forall \xi$ при связи $f(x_H^p, d, y, u, \xi) = 0$ и ограничениях:

$$g_1(x_H^p, d, y, u(\xi), \xi) \leq 0, \quad g_2(x_H^p, d, y, u(\xi), \xi) \leq 0. \quad (4.19)$$

Задача (4.18), (4.19) имеет бесконечное число ограничений и поисковых переменных (одна многомерная функция $u(\xi)$ эквивалентна бесконечному числу обычных поисковых переменных). Её решение d^* , u^* гарантирует работоспособность БТП, так как внутренняя задача оптимизации статических режимов решается во всех точках $\xi \in \Xi$. При этом, однако, нельзя гарантировать, что внутренняя задача оптимизации имеет решение для каждого $d \in D$ и $\xi \in \Xi$. Чтобы преодолеть этот недостаток, введём условие работоспособности технологического объекта: объект считается работоспособным, если для каждого $\xi \in \Xi$ можно найти такие значения управлений $u \in U$, что ограничения (4.19) будут удовлетворены. Логическое представление этого условия имеет следующий вид: $\forall \xi \in \Xi, \exists u \in U \max_{\lambda=1,2} g_\lambda(x_H^p, d, y, u, \xi) \leq 0$, что согласно соотношениям эквивалентности $\exists z \quad F(z) \Leftrightarrow \min_z F(z) \leq 0$, $\max_{z \in Z} \Psi(z) \leq 0 \Leftrightarrow \Psi(z) \leq 0$, $z \in Z$ может быть представлено в аналитическом виде $\chi(d) = \max_{\xi \in \Xi} \min_u \max_{\lambda=1,2} g_\lambda(x_H^p, d, y, u, \xi) \leq 0$, здесь символ \exists означает «существует».

Таким образом, для того чтобы гарантировать работоспособность БТП для каждого $d \in D$ и $\xi \in \Xi$, необходимо постановку задачи (4.18), (4.19) дополнить ограничением $\chi(d) = \max_{\xi \in \Xi} \min_u \max_{\lambda=1,2} g_\lambda(x_H^p, d, y, u, \xi) \leq 0$ [19]. Введение данного ограничения резко усложняет решение задачи математического проектирования прямыми методами из-за необходимости вычисления на каждой итерации значений $\chi(d)$.

Для удобства проведения дальнейших рассуждений введём функцию

$$h(d, \xi) = \min_{u \in U} \max_{\lambda=1,2} g_\lambda(x_H^p, d, y, u, \xi),$$

тогда $\chi(d) = \max_{\xi \in \Xi} h(d, \xi)$. Поскольку функции $g_\lambda(x_n^p, d, y, u, \xi)$, $\lambda = 1, 2$, выпуклы по переменным u и ξ , то максимум функции $h(d, \xi)$ по ξ находится в одной из угловых точек области неопределённости (прямоугольника) $\Xi = \{\xi_k^- : \xi_k^- < \xi_k < \xi_k^+, k=1, 2\}$, т.е. $\chi(d) = \max_{\xi \in \Xi} h(d, \xi)$, здесь $\{\Xi\}$ – множество угловых точек прямоугольника Ξ . Таким образом, максимум функции $h(d, \xi)$ по ξ и, следовательно, значения $\chi(d) = \max_{\xi \in \Xi} h(d, \xi)$ можно находить путём последовательного перебора угловых точек $\{\Xi\}$ области неопределённости Ξ .

Перепишем задачу (4.18), (4.19) в дискретном виде. Для этого заменим математическое ожидание в выражении (4.18) некоторой конечной суммой и бесконечное число ограничений (4.19) конечным числом ограничений в аппроксимационных точках $\xi^i (i \in J_1)$: $\tilde{I}^*(d, u) = \max_{d, u^i} \sum_{i \in J_1} \varphi(x_n^p, d, y, u^i, \xi^i)$ при связи $f(x_n^p, d, y, u, \xi) = 0$ и ограничениях: $g_\lambda(x_n^p, d, y, u^i, \xi^i) \leq 0, \lambda = 1, 2, i \in J_1, \chi(d) \leq 0$.

Множество аппроксимационных точек $S_1 = (\xi^i : i \in J_1)$ и весовые коэффициенты ω_i будем задавать из инженерных соображений. Например, аппроксимационные точки должны выбираться таким образом, чтобы они попадали в область наиболее вероятных значений, которые параметры ξ могут принимать при функционировании БТП.

В некоторых случаях двухэтапную задачу оптимального проектирования БТП решают без ограничения $\chi(d) \leq 0$. При этом данная задача упрощается, а её постановка бывает оправдана, если аппроксимационные точки достаточно плотно покрывают область Ξ . Для решения задачи двухэтапного проектирования без ограничения $\chi(d) \leq 0$ можно использовать вышеописанный итерационный алгоритм для одноэтапной задачи (4.12), (4.11), (4.13) математического проектирования.

Двухэтапное проектирование позволяет определить более экономичные конструкции оборудования и технологические режимы на этапе проектирования БТП по сравнению с одноэтапным проектированием. Но при этом остаются проблемы, связанные с экономичным вычислением многомерного интеграла (математического ожидания $M_\xi(\varphi(x_n^p, d, y, u, \xi))$ и вероятностей $\text{Вер}\{g_\lambda(\cdot) \leq 0\}$ выполнения ограничений в задаче оптимизации, реализацией оптимального режима и обеспечением устойчивости и надёжности функционирования объекта.

Указанные проблемы требуют скорейшего разрешения, и мы призываем молодых учёных, аспирантов и студентов внести свой вклад в разработку теоретических основ и инженерных методов математического проектирования автоматизированных биотехнологических процессов.

ЗАКЛЮЧЕНИЕ

В учебном пособии рассмотрено введение в современную теорию и практику системного анализа, математического моделирования и оптимизации технологических процессов биотехнологических производств. Рассмотрены: структура и функциональные подсистемы биохимического производства, иерархическая структура и связи в биотехнологической системе на примерах нескольких производств, основы разработки и проектирования автоматизированных биотехнологических процессов и производств, особенности их эксплуатации.

Особое внимание уделено вопросам математического моделирования и оптимизации биотехнологических процессов. Рассмотрены одноэтапные и двухэтапные задачи математического проектирования. Для решения поставленных задач можно использовать современные пакеты прикладных программ, включающие набор математических моделей типовых объектов и специальные функции, реализующие численные методы решения. Следует отметить, что представленные в них математические модели практически всегда нуждаются в серьёзном анализе адекватности на базе экспериментальных данных, полученных на функционирующем объекте или на специализированных лабораторных установках. Указанные операции весьма трудоёмки, а отказ от их выполнения не гарантирует получения достаточно точных и надёжных результатов математического моделирования и оптимизации процесса.

При постановке одноэтапных и двухэтапных задач проектирования следует отметить, что остаётся немало проблем, связанных с вычислением критерия оптимальности, экономичным (в смысле затраченного времени) вычислением математического ожидания, разработкой новых достаточно простых и быстродействующих алгоритмов решения одноэтапных и двухэтапных задач оптимального проектирования процессов биотехнологических производств, обеспечения устойчивости и надёжности проектируемых технологических объектов. Для успешного решения рассмотренных оптимизационных задач необходимо создание методов/способов анализа степени выпуклости и особенно невыпуклости критериев. Ещё одна проблема связана с реализацией оптимального режима на этапе функционирования технологического процесса.

Подводя итог, можно отметить, что современные средства компьютерного моделирования в совокупности с системным подходом открывают всё большие возможности для решения задач расчёта биотехнологических процессов, проектирования нового перспективного оборудования и поиска новых способов повышения его эффективности.

СПИСОК ЛИТЕРАТУРЫ

1. **Бирюков, В. В.** Основы промышленной биотехнологии / В. В. Бирюков. – М. : КолосС, 2004. – 296 с.
2. **Баранов, Д. А.** Процессы и аппараты : учебник / Д. А. Баранов, А. М. Кутепов. – М. : Академия, 2004. – 304 с. : ил.
3. **Бесков, В. С.** Общая химическая технология и основы промышленной экологии : учебник / В. С. Бесков, В. С. Сафронов. – М. : Химия, 1999. – 472 с.
4. **Перегудов, Ф. И.** Введение в системный анализ / Ф. И. Перегудов, Ф. П. Тарасенко. – М. : Высшая школа, 1989. – 367 с.
5. **Моисеев, Н. Н.** Математические задачи системного анализа / Н. Н. Моисеев. – М. : Наука, 1981. – 488 с.
6. **Балакирев, В. С.** Оптимальное управление процессами химической технологии (экстремальные задачи в АСУ) : Сер. Химическая кибернетика / В. С. Балакирев, В. М. Володин, А. М. Цирлин. – М. : Химия, 1978 – 384 с.
7. **Основы** проектирования химических производств : монография / С. И. Дворецкий, Д. С. Дворецкий, Г. С. Кормильцин, А. А. Пахомов. – М. : ИД «Спектр», 2014. – 356 с.
8. **Ахназарова, С. Л.** Оптимизация эксперимента в химии и химической технологии : учебное пособие / С. Л. Ахназарова, В. В. Кафаров. – М. : Высшая школа, 1978. – 319 с.
9. **Математическое моделирование** технологических процессов : учебное пособие / под. ред. В. С. Балакирева. – Ярославль : Изд-кий дом Н. П. Пастухова, 2018. – 352 с.
10. **Пантелеев, А. В.** Методы оптимизации в примерах и задачах : учебное пособие / А. В. Пантелеев, Т. А. Летова. – М. : Высшая школа, 2005. – 544 с.
11. **Базара, М.** Нелинейное программирование. Теория и алгоритмы / М. Базара, К. Шетти. – М. : Мир, 1982. – 583 с.
12. **Реклейтис, Г.** Оптимизация в технике. В 2-х кн. / Г. Реклейтис, Ф. Рейвиндран, Л. Рэгсдел. – М. : Мир, 1986. – Т. 1, 2.
13. **Моисеев, Н. Н.** Методы оптимизации / Н. Н. Моисеев, Ю. П. Иванилов, Е. М. Столярова. – М. : Наука, 1978. – 351 с.
14. **Сухарев, А. Г.** Курс методов оптимизации / А. Г. Сухарев, А. В. Тимохов, В. В. Федоров. – М. : ФИЗМАТЛИТ, 2005. – 368 с.
15. **Мину, М.** Математическое программирование. Теория и алгоритмы / М. Мину. – М. : Наука, 1990. – 488 с.
16. **Поляк, Б. Т.** Введение в оптимизацию / Б. Т. Поляк. – М. : Наука, Главная редакция физико-математической литературы, 1983. – 384 с.
17. **Теплообменные аппараты** : учебное пособие / Б. Е. Байгалиев, А. В. Щелчков, А. Б. Яковлев, П. Ю. Гортышов. – Казань : Изд-во Казан. гос. техн. ун-та, 2012. – 180 с.
18. **Павлов, К. Ф.** Примеры и задачи по курсу процессов и аппаратов химической технологии : учебное пособие / К. Ф. Павлов, П. Г. Романков, А. А. Носков. – Л. : Химия, 1987. – 576 с.
19. **Островский, Г. М.** Технические системы в условиях неопределённости / Г. М. Островский, Ю. М. Волин. – М. : БИНОМ. Лаборатория знаний, 2008. – 319 с.

СОДЕРЖАНИЕ

Предисловие	3
Введение	5
Глава 1. Общая характеристика процессов биотехнологии	7
Глава 2. Системный анализ биотехнологических процессов и систем	16
2.1. Определение системного анализа	17
2.2. Характеристика этапов системного анализа	20
2.3. Биотехнологический процесс как система	24
2.4. Состояние биотехнологической системы	35
2.5. Анализ биотехнологической системы	58
2.6. Разработка и проектирование автоматизированных биотехнологических процессов и производств	66
Глава 3. Математическое моделирование биотехнологических процессов	72
3.1. Математические модели технологических процессов	72
3.2. Построение математических моделей	82
3.3. Параметрическая идентификация математической модели	87
3.4. Анализ адекватности модели биотехнологического процесса	90
3.5. Типовые задачи математического моделирования проектируемого биотехнологического процесса	91
Глава 4. Оптимизация биотехнологических процессов и детерминированные оценки альтернатив	111
4.1. Постановка и анализ основных компонентов оптимизационных задач	111
4.2. Необходимые и достаточные условия безусловного экстремума	114
4.3. Необходимые и достаточные условия условного экстремума	116
4.4. Численные методы решения безусловных оптимизационных задач	124
4.5. Численные методы решения условных оптимизационных задач	144
4.6. Оптимизация биотехнологического процесса в условиях частичной неопределённости	146
Заключение	157
Список литературы	158

Учебное издание

Дворецкий Дмитрий Станиславович
Дворецкий Станислав Иванович
Акулинин Евгений Игоревич
Голубятников Олег Олегович
Темнов Михаил Сергеевич

СИСТЕМНЫЙ АНАЛИЗ И ОПТИМИЗАЦИЯ БИОТЕХНОЛОГИЧЕСКИХ ПРОИЗВОДСТВ

Учебное пособие

Редактор Л. В. Комбарова
Компьютерное макетирование М. А. Евсейчевой

Подписано в печать 18.11.2019.

Выход в свет 25.12.2019.

Формат 60 × 84 / 16. 9,3 усл. печ. л.

Тираж 100 экз. (1-й з-д 70) Заказ № 85

ISBN 978-5-8265-2097-0



Издательский центр ФГБОУ ВО «ТГТУ»

392000, г. Тамбов, ул. Советская, д. 106, к. 14.

Тел. 8(4752) 63-81-08.

E-mail: izdatelstvo@admin.tstu.ru

Отпечатано в Типографии ФГБОУ ВО «ТГТУ»

392008, г. Тамбов, ул. Мичуринская, д. 112А.

Тел. 8(4752) 63-07-46.

E-mail: tipo_tstu68@mail.ru