

1. ОСНОВНЫЕ ПОЛОЖЕНИЯ ОПТИМИЗАЦИИ

1.1. Оптимизация в задачах управления и проектирования

Оптимизация – это процесс нахождения наилучшего решения задачи, определяемого по некоторому заранее установленному критерию. Другими словами, это целенаправленная деятельность, заключающаяся в получении наилучших результатов при соответствующих условиях.

Оптимизация в широком смысле слова находит применение в науке, технике и в любой другой области человеческой деятельности, например:

- при проектировании сложных инженерных сооружений и систем;
- в управлении производственными, техническими и экономическими системами;
- в процессах разработки автоматизированных информационных систем и т.д.

Так, **оптимальная система управления** может быть реализована в виде набора правил, стратегии или способа управления, согласно которым следует поступать в той или иной ситуации (военной, производственной и т.д.), или в виде комплекса технических средств управления (космическим или морским объектом, инженерным проектом и т.д.).

Поиски оптимальных решений привели к созданию специальных математических методов, и уже в XVIII в. были заложены математические основы оптимизации (вариационное исчисление, численные методы и др.). Однако до второй половины XX в. методы оптимизации во многих областях науки и техники применялись редко, поскольку практическое использование математических методов оптимизации требовало огромной вычислительной работы, которую без ЭВМ реализовать было крайне трудно, а в ряде случаев – невозможно. Например, большие трудности возникают при решении задач оптимизации процессов в химической технологии из-за большого числа параметров и их сложной взаимосвязи между собой. При наличии ЭВМ задача заметно упрощается.

Постановка задачи оптимизации предполагает существование следующих условий:

1. Наличие объекта оптимизации и цели оптимизации. Формулировка каждой задачи оптимизации должна требовать экстремального значения лишь одной величины, т.е. одновременно системе не должно приписываться два и более критерия оптимизации (практически всегда экстремум одного критерия не соответствует экстремуму другого).

Вот типичный пример неправильной постановки задачи оптимизации: “Получить максимальную производительность при минимальной себестоимости”. Ошибка заключается в том, что ставится задача поиска оптимума двух величин, противоречащих друг другу по своей сути. Правильная постановка задачи могла быть следующей: “Получить максимальную производительность при заданной себестоимости или получить минимальную себестоимость при заданной производительности”.

В первом случае критерий оптимизации – производительность, а во втором – себестоимость.

2. Наличие ресурсов оптимизации, под которыми понимают возможность выбора значений некоторых параметров оптимизируемого объекта. Объект должен обладать определенными степенями свободы – *управляющими воздействиями*.

3. Возможность количественной оценки оптимизируемой величины, поскольку только в этом случае можно сравнивать эффекты от выбора тех или иных управляющих воздействий.

4. Учет ограничений.

Часто оптимизируемая величина связана с экономичностью работы рассматриваемого объекта. Оптимизируемый вариант работы объекта должен оцениваться *критерием оптимальности*, под которым понимается количественная оценка оптимизируемого качества объекта.

На основании выбранного критерия оптимальности составляется целевая функция, представляющая собой зависимость критерия оптимальности от параметров, влияющих на ее значение. Вид критерия оптимальности или целевой функции определяется конкретной задачей оптимизации.

Таким образом, задача оптимизации сводится к нахождению экстремума целевой функции.

Наиболее общей постановкой оптимальной задачи является выражение критерия оптимальности в виде экономической оценки (производительность, себестоимость продукции, прибыль, рентабельность). Иногда, в частных задачах оптимизации, когда объект является частью технологического процесса, не всегда удается или не всегда целесообразно выделять прямой экономический показатель, который бы полностью характеризовал эффективность работы рассматриваемого объекта. В таких случаях критерием оптимальности может служить технологическая характеристика, косвенно оценивающая экономичность работы агрегата (время контакта, выход продукта, степень превращения, температура).

В задачах оптимизации важным моментом является использование системного подхода при постановке задачи. Сущность системного подхода заключается в комплексном, едином рассмотрении всех частей системы и их эффективном сочетании. Так, при космических полетах можно, увеличивая вес корабля, добиться максимальной автономии управления, независимо от Земли. Можно, наоборот, обеспечив хорошую связь с кораблем, больше аппаратуры разместить на Земле, максимально уменьшив вес корабля. Оптимальную границу распределения веса наземной и бортовой аппаратуры должны определить методы оптимизации исходя из критерия оптимальности и состояния техники, доступной проектировщикам.

Процесс проектирования любой системы включает в себя, как правило, следующие этапы:

- создание математической модели процесса, которым требуется управлять (объекта управления), и системы управления с обратными связями (контура управления);
- нахождение формального (математического) закона (описания) управления для полученной модели;
- разработка (приобретение) системного программно-технического комплекса управляющей информационно-вычислительной системы и средств связи;
- разработка информационного и программного обеспечения.

На каждом из этих этапов возникает проблема формулировки критериев оптимальности и оптимизации. Так, на этапе разработки информационного и программного обеспечения необходимо решать задачи *оптимизации программирования*, т.е. задачу порождения компилятором выходной программы, которая наилучшим способом использует ресурсы ЭВМ. При этом задача разбивается на ряд подзадач, например:

- *глобальная оптимизация программирования* – переупорядочивание установленной последовательности выполнения команд с целью исключения избыточных вычислений;
- *регистровая оптимизация программирования* – привязка регистров ЭВМ к переменным и промежуточным результатам, чтобы минимизировать число случаев “холостого” резервирования регистров;
- *локальная оптимизация программирования* – адаптация программы к конкретным особенностям архитектуры ЭВМ.

Хотя системный подход подразумевает общую оптимизацию всех этапов, каждый из них в отдельности может и не быть оптимальным.

Для решения задач оптимизации прежде всего необходимо уметь формулировать критерии оптимальности и владеть методами (процедурами) оптимизации.

1.2. Классические основы оптимизации

Классические основы оптимизации определяются методами исследования функций в математическом анализе.

Известно, что функция $f(x)$ имеет локальный минимум в точке x_0 , если существует некоторая положительная величина d , такая, что если $|x - x_0| < d$, то $f(x) \geq f(x_0)$, т.е. если существует окрестность точки x_0 , такая, что для всех значений x в этой окрестности $f(x)$ больше $f(x_0)$. Функция $f(x)$ имеет *глобальный минимум* в точке x^* , если для всех x справедливо неравенство $f(x) \geq f(x^*)$.

На рис. 1а дано графическое представление функции $f(x)$, которая имеет локальный минимум в точке x_0 и глобальный минимум в точке x^* .

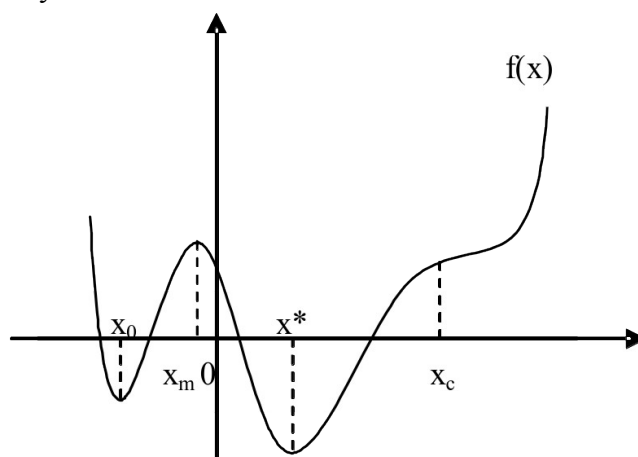


Рис. 1 а. Локальный и глобальный минимумы

Классический подход к задаче нахождения значений x_0 и x^* состоит в поиске уравнений, которым они должны удовлетворять. Представленная на рис. 1а функция и ее производные непрерывны, и видно, что в точках x_0 и x^* производная $f'(x)$ (градиент функции) равна нулю. Следовательно, x_0 и x^* будут решениями уравнения

$$f'(x) = 0.$$

Точка x_m , в которой достигается локальный максимум, и точка x_c , в которой имеется точка горизонтального перегиба функции, также удовлетворяют этому уравнению. Следовательно, уравнение $f'(x) = 0$ является только необходимым условием минимума, но не является достаточным условием минимума.

Заметим, однако, что в точках x_0 и x^* производная $f'(x)$ меняет знак с отрицательного на положительный. В точке x_m знак меняется с положительного на отрицательный, в то время как в точке x_c он не меняется. Следовательно, производная в минимуме является возрастающей функцией, а поскольку степень возрастания $f'(x)$ измеряется второй производной, можно ожидать, что $f''(x_0) > 0$, $f''(x^*) > 0$, тогда как $f''(x_m) < 0$.

Если, однако, вторая производная равна нулю, ситуация остается неопределенной.

Полученные выше результаты могут найти надежное обоснование, если рассмотреть разложение функции $f(x)$ в ряд Тейлора в окрестности точки x_0 (или x^* , или x_m), что требует непрерывности функции $f(x)$ и ее производных:

$$f(x_0 + h) - f(x_0) = hf'(x_0) + (h^2/2!)f''(x_0) + \dots$$

Если в точке x_0 достигается минимум, то левая часть уравнения будет неотрицательной для любого достаточно малого h . Следовательно, первая производная $f'(x_0)$ должна быть равна нулю. Это является достаточным условием (см. уравнение $f'(x) = 0$). Если бы она была положительной, то достаточно малое отрицательное значение h делало бы правую часть

$$f(x_0 + h) - f(x_0) = hf'(x_0) + (h^2/2!)f''(x_0) + \dots$$

отрицательной, а если бы она была отрицательной, то достаточно малое положительное значение h делало бы правую часть отрицательной.

Так как всегда $h^2 > 0$, то, если $f''(x_0) > 0$, в точке x_0 достигается минимум. Если $f'(x_m) = 0$ и $f''(x_m) < 0$, то из аналогичных соображений в точке x_m достигается максимум. Для определения различия между локальным и глобальным минимумами необходимо сравнить значения функций $f(x_0)$ и $f(x^*)$.

Например, исследуем характер точек перегиба функции

$$f(x) = x^3 - 2x^2 + x + 1,$$

так как $f'(x) = 3x^2 - 4x + 1 = 0$ и $(3x - 1)(x - 1) = 0$, то $x = 1/3$ или $x = 1$.

При $x = 1/3$ производная $f'(x)$ меняет знак с положительного на отрицательный, а при $x = 1$ — отрицательного на положительный. Следовательно, в точке $x = 1/3$ достигается максимум, а в точке $x = 1$ — минимум.

Этот пример может быть решен более простым способом, если вычислить вторую производную $f''(x) = 6x - 4$:

$f''(1/3) = -2$, т.е. отрицательна, и при $x = 1/3$ достигается максимум;

$f''(1) = 2$, т.е. положительна, и при $x = 1$ достигается минимум. Неоднозначность, возникающую при $f''(x) = 0$, можно разрешить, увеличив количество членов в формуле разложения в ряд Тейлора:

$$f(x_0 + h) - f(x_0) = hf'(x_0) + (h^2/2!)f''(x_0) + (h^3/3!)f'''(x_0) + (h^4/4!)f''''(x_0) + \dots$$

При этом можно сформулировать следующее правило:

“Если функция $f(x)$ и ее производные непрерывны, то точка x_0 является точкой экстремума (максимума или минимума) тогда и только тогда, когда n четное, где n — порядок первой необращающейся в нуль в точке x_0 производной. Если $f''(x_0) < 0$, то в точке x_0 достигается максимум, если $f''(x_0) > 0$, то в точке x_0 достигается минимум”.

Например, найдем точку перегиба функции $f(x) = (x - 1)^6$.

$$f'(x) = 6(x - 1)^5 = 0 \text{ при } x = 1.$$

Первой необращающейся в нуль в точке $x = 1$ производной будет $f^6(1) = 6!$. Следовательно, функция $f(x)$ имеет минимум в точке $x = 1$.

Функцию n действительных переменных можно представить как

$$F(x_1, x_2, x_3, \dots, x_n).$$

Точка в n -мерном евклидовом пространстве с координатами $x_1, x_2, x_3, \dots, x_n$ обозначается вектором-столбцом x . Градиент функции, т.е. вектор с компонентами $\partial f/\partial x_1, \partial f/\partial x_2, \dots, \partial f/\partial x_n$, обозначается $\nabla f(x)$ или, иногда, $g(x)$. Наиболее распространенные задачи оптимизации заключаются в нахождении **минимума (или максимума) функции** или функционала. В первом

случае находят значение n переменных x_1, x_2, \dots, x_n , при которых функция $F(x_1, x_2, \dots, x_n)$ принимает экстремальное значение $F = \min(\max)$.

В простейшем случае дифференцируемости функции и неравенства нулю вторых производных задача сводится к решению n алгебраических (в общем случае нелинейных) уравнений

$$dF/dx_i = 0, i = 1, 2, \dots, n. \quad (1-1)$$

При оптимизации управления приходится оперировать с большим числом переменных. В этом специфика задач оптимизации, и это затрудняет решение уравнений оптимизации даже с помощью ЭВМ.

Если функция $F(x)$, где $x = (x_1, x_2, \dots, x_n)$, помимо переменных x_1, x_2, \dots, x_n зависит еще (параметрически) от другой переменной λ , то решение для каждого значения λ соотношения (1-1) дает оптимальный закон управления

$$x(\lambda) = \{x_1(\lambda), x_2(\lambda), \dots, x_n(\lambda)\}.$$

Здесь мы уже переходим, по существу, к понятию функционала, частным случаем которого является функция. Методы оптимизации, использующие этот критерий, составляют содержание раздела математики, названного вариационным исчислением. Понятие функционала в математике является дальнейшим обобщением понятия функции. Не очень строго **функционал** можно определить как функцию от функции, т.е. функцию, в которой в качестве независимой переменной выступает другая функция.

Если одному множеству M значений величины x ($x \in M$) соответствует другое множество N значений величины y , то говорят, что y является функцией x , т.е. $y = f(x)$.

Например, пусть $M \{1, 4, 9, 8\}$ и $N \{20, 70, 90, 110\}$ и пусть каждому значению x из множества M соответствует одно значение y из множества N . Это – метод задания функции в виде таблиц. Он часто встречается в кибернетике. Аналитически функция может быть задана в виде $y = x^2, y = \sin x$.

Если M – множество функций и каждой функции $f(x)$, принадлежащей $M \{f(x) \in M\}$, ставится в соответствие определенное значение величины y из множества N , то говорят, что на множестве M задан функционал.

Например

$$M \{\sin x, \cos x, \operatorname{tg} x, \operatorname{ctg} x\}$$

и

$$N \{3, 4, 5, 18\}.$$

Другим примером функционала может служить определенный интеграл

$$I = I \{y(x)\} = \int_0^l y(x) dx.$$

Каждой функции $y(x)$ будет соответствовать числовое значение I . Так, при $y = x$ $I = 1/2$, при $y = x^2$ $I = 1/3$.

Можно поставить задачу отыскивания такой функции $y(x)$, которая обращала бы этот функционал в минимум (максимум), т.е.

$$I\{y\} \rightarrow \min(\max).$$

В общем случае подынтегральное выражение в функционале может зависеть явно от аргумента x , y и производной y' :

$$I\{y\} = \int_a^b F(x, y, y') dx. \quad (1-2)$$

Нетрудно показать, что запись (1-2) включает в себя функцию. Так, если положить

$$F(x, y, y') = f(x) g(t-x),$$

то

$$I = \int_a^b f(x) g(t-x) dx = f(x),$$

при $a < t < b$.

Аналогично, если положить

$$F(x, y, y') = c_i f_i(x) g(t-x),$$

то

$$I = \sum_{i=1}^n c_i f_i(t_i).$$

Иногда считают, что функционал является функцией бесконечного числа переменных. Соответственно, **вариационное исчисление** можно рассматривать как обобщение методов отыскивания экстремума функции на случай большого или бесконечного числа переменных. Действительно, функцию $y(x)$ в формуле (1-2) можно заменить приближенно ломаной линией (рис. 1) с вершинами $y_0 = y(x_0) = y(a)$, $y_1 = y(x_0 + \Delta x)$, $y_n = y(x_0 + \Delta n x) = y(b)$, а интеграл – суммой:

$$I = \sum_{i=0}^{n-1} F(x_i, y_i, (y_i - y_{i-1}) / \Delta x) \Delta x.$$

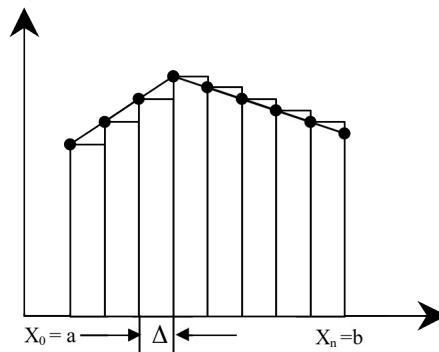


Рис. 1. Замена интеграла на сумму

После этого вариационная задача приближенно решается как обычная задача на отыскивание экстремума функции n переменных

$$I = (y_1, y_2, \dots, y_n).$$

Такой вывод использовал для своего основного уравнения вариационного исчисления Эйлер.

Однако большинство задач оптимизации содержит ограничения на исходную функцию. Пусть требуется минимизировать функционал

$$I = \int_a^b F(x, y, y') dx \rightarrow \min$$

при наличии ограничений

$$\varphi_k(x, y, y') \leq 0;$$

$$k = 1, 2, 3, \dots, m,$$

где φ_k - некоторые функции.

Смысл этих соотношений состоит в том, что отыскивается не любая функция $y(x)$, обращающая функционал в минимум, а такая, которая удовлетворяет системе ограничений.

Нетрудно убедиться, что тем самым значение **условного экстремума** (экстремум функции при условии, когда на критерий оптимизации наложены ограничения) не может быть меньше значения *абсолютного экстремума* (без ограничений). Аналогичным образом формируется критерий оптимальности (с ограничениями) для функции

$$\begin{aligned} F(x_1, x_2, \dots, x_n) &\rightarrow \min; \\ \varphi_k(x_1, x_2, \dots, x_n) &\leq 0, \\ k &= 1, 2, 3, \dots, m. \end{aligned} \quad (1-3)$$

В классическом вариационном исчислении в функционале интеграл понимается в обычном смысле как предел сумм Дарбу. Далее в целях соблюдения прикладного уровня изложения будет использовано классическое понятие функционала (если не дано специальных оговорок).

Следует заметить, что во всех приведенных формулах переменные x и y могут быть векторами:

$$\begin{aligned} x &= (x_1, x_2, \dots, x_n); \\ y &= (y_1, y_2, \dots, y_n); \\ F &= (F_1, F_2, \dots, F_p). \end{aligned}$$

Такая форма записи широко используется, и, в частности, она учитывает случаи оптимизации функции многих переменных и оптимизации нескольких функций.

Задачи оптимизации при наличии ограничений, по существу, привели к пересмотру классических методов и созданию новых методов, известных под названием методов программирования. Если в формулах (1-3) все функции линейные, налицо задача линейного программирования. В общем случае эти соотношения определяют задачу нелинейного программирования.

1.3. Критерии оптимизации

Критерий оптимизации - количественный или порядковый показатель, который выражает предельную меру экономического, научно-технического или другого эффекта принимаемого решения для сравнительной оценки возможных решений (альтернатив) и выбора наилучшего.

Вопрос о выборе критерия оптимизации - один из самых важных в процессах оптимизации, в то же время он один из наиболее сложных, и процесс выбора критерия содержит существенную творческую составляющую. Рассмотрим более подробно требования, которые должны предъявляться к критерию оптимизации.

Критерий оптимизации должен:

- выражаться количественно;
- быть единственным;
- отражать наиболее существенные стороны процесса;
- иметь ясный физический смысл и легко рассчитываться.

При постановке конкретных задач оптимизации критерий оптимизации должен быть записан в виде аналитического выражения. В том случае, когда случайные возмущения невелики, а их воздействие на объект можно не учитывать, критерий оптимизации может быть представлен как функция входных, выходных и управляющих параметров:

$$K_{\text{опт.}} = K(X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_n, U_1, U_2, \dots, U_n),$$

где K – критерий оптимизации; Y – выходные параметры; X – контролируемые входные параметры; U – регулируемые, управляющие параметры.

Так как $Y=f(U)$, то при фиксированных X можно записать $K = K(U_{\text{i ср.}})$

При этом всякое изменение значений управляющих параметров двояко сказывается на величине K :

- прямо, т.к. управляющие параметры непосредственно входят в выражение критерия оптимизации;
- косвенно – через изменение выходных параметров процесса, которые зависят от управляющих.

Если же случайные возмущения достаточно велики и их необходимо учитывать, то следует применять экспериментально-статистические методы, которые позволят получить модель объекта в виде функции

$$Y = \varphi (X_i, U_i),$$

которая справедлива только для изученной локальной области. Тогда критерий оптимальности примет следующий вид:

$$K = K(X_{cp}, U_{cp}).$$

В принципе, для оптимизации вместо математической модели можно использовать и сам объект, однако оптимизация опытным путем имеет ряд следующих недостатков:

- необходим реальный объект;
- необходимо изменять технологический режим в значительных пределах, что не всегда возможно;
- длительность испытаний и сложность обработки данных.

Наличие математической модели (при условии, что она достаточно надежно описывает процесс) позволяет значительно проще решить задачу оптимизации аналитическим либо численным методами.

Критерии оптимизации могут классифицироваться разными способами, в зависимости от конкретных задач.

Так, иногда различают простые и сложные критерии оптимизации. Критерий оптимизации называется простым, если требуется определить экстремум целевой функции без задания условий на какие-либо другие величины. Такие критерии обычно используются при решении частных задач оптимизации (например определение максимальной концентрации целевого продукта, оптимального времени пребывания реакционной смеси в аппарате и т.п.).

Критерий оптимизации называется сложным, если необходимо установить экстремум целевой функции при некоторых условиях, которые накладывают ограничения на ряд других величин (например, определение максимальной производительности при заданной себестоимости, определение оптимальной температуры при ограничениях по термостойкости). Как правило, процедура решения задачи оптимизации обязательно включает, помимо выбора управляющих параметров, еще и установление ограничений на эти параметры. Ограничения могут накладываться как по технологическим, так и по экономическим соображениям.

В некоторых случаях выделяются два других вида критериев оптимизации. Это, во-первых, **критерии оптимизации прагматические** – выработанные практикой качественные или количественные характеристики оптимальности работы различных систем и, во-вторых, **критерии оптимизации математические** – разработанные математиками математические критерии оптимальности, положенные в основу аналитических, графоаналитических, численных и машинных методов оптимизации.

В настоящее время наблюдается сближение этих двух критериев: с одной стороны, появились новые математические методы оптимизации, такие, как принцип максимума и динамическое программирование, которые лучше приспособлены для решения практических задач оптимизации, с другой стороны, практика проектирования все чаще пользуется критериями оптимальности, удобными в математическом смысле. Так произошло с критерием среднеквадратической ошибки,

принятым в качестве оценки точности работы автоматических систем регулирования, и критерием вероятности, утвердившимся как количественная оценка эффективности для работы системы.

На практике используется целый ряд критериев оптимизации. Выбор того или иного критерия зависит от экономиста или конструктора системы управления, и в этом содержится элемент нестрогости. В то же время, при использовании в оптимизирующих расчетах библиотек прикладных программ для языков программирования все чаще применяют типовые критерии, которые становятся общепринятыми и заносятся непосредственно в технические задания.

Далее рассматривается ряд примеров критериев оптимизации, которые используются в инженерной практике.

Критерий среднего квадрата ошибки – требование минимума дисперсии или квадрата ошибки между заданным $h(t)$ и выходным $x(t)$ сигналами системы.

Этот критерий применяется при оценке качества работы автоматизированных систем регулирования и удобен при решении математических задач оптимизации. Требование минимума дисперсии или квадрата ошибки (рис. 2) между заданным (или желаемым) $h(t)$ и выходным $x(t)$ сигналами системы записывается как

$$e^2_{\text{cp}} = [h(t) - x(t)]^2_{\text{cp}} \rightarrow \min$$

и означает большую нежелательность (по сравнению с линейным законом) больших (чем малых) по значению ошибок (рис. 3). Причем, в соответствии с рис.2 $h(t)$ получается из полезного входного сигнала $m(t)$ с помощью заданного оператора $x(t)$, в то время как реальный сигнал $x(t)$ получается из входного сигнала системы $s(t)$ с помощью оператора $k(t)$, который следует найти. При использовании критерия

$$|e_{\text{cp}}| = |h(t) - x(t)|_{\text{cp}} = \min,$$

где “cp” означает усреднение по ансамблю, считается, что вред, наносимый ошибкой, пропорционален ее величине.

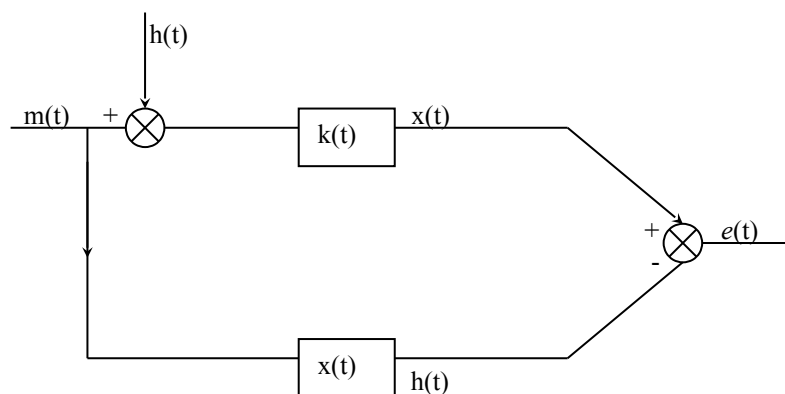


Рис. 2. Схема определения ошибки автоматической системы регулирования

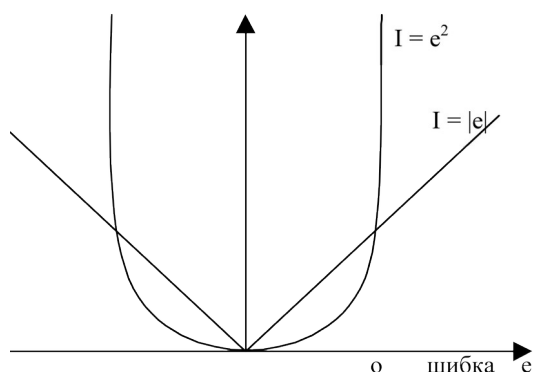


Рис. 3. Оценка качества управления по величине дисперсии

Средний квадрат ошибки e_{cp}^2 - это функционал от ее импульсной переходной функции и для стационарных сигналов и линейной системы имеет вид

$$e_{cp}^2 = R_h(0) - 2 \int_0^{\infty} K(\tau) R_{hs}(\tau) d\tau + \int_0^{\infty} K(\tau) d\tau \int_0^{\infty} K(\tau_1) R_s(\tau - \tau_1) d\tau_1,$$

где $R_h(\tau)$ и $R_s(\tau)$ - функции корреляции желаемого $h(t)$ и входного $s(t)$ сигналов, а $R_{hs}(\tau)$ - их взаимная функция корреляции.

Если при минимизации требуется обеспечить заданное значение динамической ошибки, то добавляются условия

$$\int_0^{\infty} K(\tau) (-\tau)^z d\tau = \mu_z, z = 1, 2, \dots, m,$$

где μ_z - заданные числа, и задача становится вариационной задачей на условный экстремум.

Интегральный критерий – это критерий, имеющий вид интеграла по отрезку (в общем случае по области), на котором (ой) задана искомая функция. Такой критерий, в частности, используется для определения параметров системы управления, оптимальной в переходном режиме, и в качестве него иногда выбирается минимум функционала

$$I_1 = \int_0^{\infty} [e^2 + T^2 (de/dt)^2] dt,$$

где e - ошибка рассогласования в системе (рис. 4).

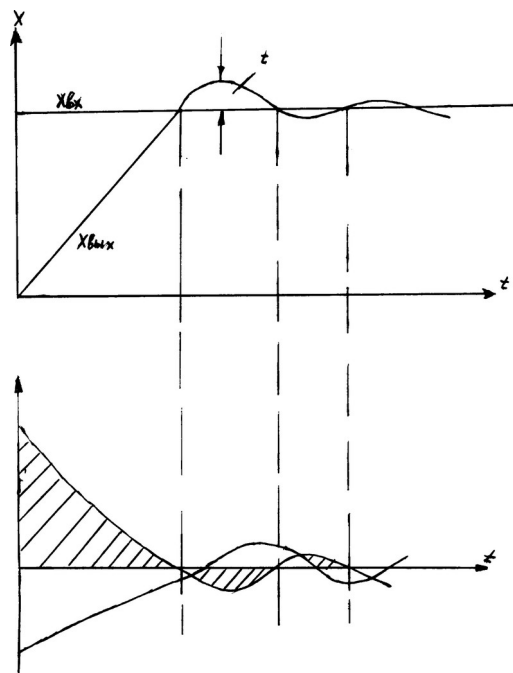


Рис. 4. Интегральный критерий качества

Оптимальная система управления - система, реализованная в виде набора правил, стратегии, согласно которым следует поступать в соответствующих ситуациях для получения оптимального решения.

Требованием минимума функционала

$$I_1 \rightarrow \min$$

можно обеспечить в системе переходный процесс с малыми отклонениями, величина которых определяет величину интеграла

$$\int_0^{\infty} e^2 dt.$$

Переходный процесс достаточно плавный, без резких колебаний, что определяется величиной интеграла

$$\int_0^{\infty} T^2 (de/dt)^2 dt,$$

который по существу включает управление по производной. Дело в том, что при переходе нулевого значения ошибки e производная e' принимает большие значения и из-за этого возникают большие переколебания. Наличие второго члена в функционале уменьшает значение производной, способствуя меньшим колебаниям.

Интегральный критерий требует минимума суммарной площади, ограниченной кривыми ошибки, ее производной и осью абсцисс, причем последняя площадь берется с весом T . Из рис. 4 видно, что эти две кривые “работают” со сдвигом примерно в 90 градусов, и коэффициент T осуществляет оптимальный баланс между управлением по ошибке и по производной.

Критерий максимального быстродействия состоит в минимизации времени, за которое управляемый объект должен перейти в заданное состояние.

Этот критерий возник при зарождении вариационного исчисления (первая задача которого – задача о “брахистохроне”, кривой наискорейшего спуска, и была задачей быстродействия) и основное развитие получил при решении военных задач в 60-е гг. XX в., в которых быстродействие является принципиальным фактором.

В наиболее распространенном случае задача оптимизации по быстродействию сводится к получению переходного процесса, заканчивающегося в кратчайшее время. Будь то система управления ракетой или антенной радиолокационной станции, требуется, чтобы переходный процесс заканчивался в минимальное время. Дело в том, что до окончания переходного процесса система не может выполнить своего основного назначения следящей системы, например антенна радиолокационной станции не может следить за входным сигналом (рис. 5). Для этого случая в функционале

$$I = \int_a^b F(x, y, y') dx \rightarrow \min$$

необходимо положить

$$F(x, x', t) = 1$$

и, изменив $x \rightarrow t$, $y \rightarrow x$, $y' \rightarrow x'$, получить

$$I_2 = \int_{t_0}^{t_k} F(x, x', t) dt = t_k - t_0 \rightarrow \min,$$

причем t_0 соответствует начальным координатам процесса, а t_k - конечным.

Наиболее распространено получение оптимального переходного процесса с включением максимального ускорения в начале с последующим переходом на максимальное торможение. В этом случае необходимо только определить момент переключения t_n . Задача максимального быстродействия может возникнуть не обязательно в связи с переходным процессом следящих систем. Если, например, требуется за наименьшее время доставить летательный аппарат с Земли на Луну, то эта задача также сведется к задаче о максимальном быстродействии.

Критерий минимума стоимости в единицу времени – стоимость функционирования совокупности систем массового обслуживания в единицу времени.

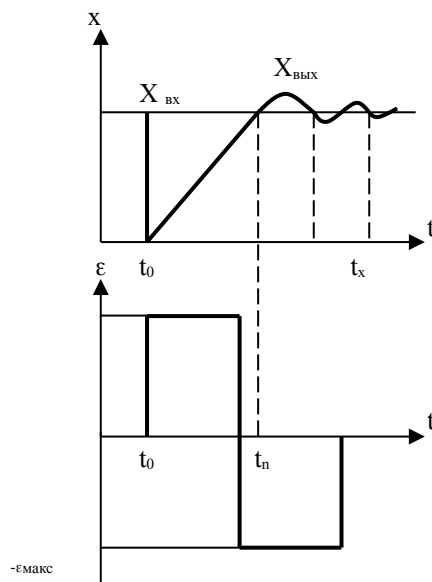


Рис. 5. Система, оптимальная по быстродействию

В качестве другого примера функционала, встречающегося в практике проектирования оптимальных систем, можно привести стоимость функционирования совокупности систем массового обслуживания (или сети массового обслуживания) в единицу времени

$$C = c_1 p_{cp} + c_2 w_{cp},$$

где p_{cp} — среднее число свободных приборов; w_{cp} — среднее число заявок, ожидающих своей очереди; c_1 и c_2 — стоимости простоев одного прибора и заявки, соответственно, в единицу времени.

Из-за случайности потока заявок и времени их обслуживания всегда имеется какое-то среднее число простаивающих приборов или ожидающих заявок. Требуется так распределить число приборов по разным системам, чтобы $C = \min$. Такая кибернетическая модель и критерии широко используются при создании оптимальных систем управления в сфере обслуживания, например управления системой гостиниц. В качестве заявок здесь выступают клиенты, в качестве приборов — администраторы. Поток клиентов и время их обслуживания — случайные величины. Плохо, и когда простаивают администраторы, и когда ожидают клиенты. Минимум функционала

$$C = c_1 p_{cp} + c_2 w_{cp}$$

означает оптимальный выбор числа администраторов, основанный на определенной статистике потока клиентов. Для решения задачи должны быть известны значения стоимости простоя клиентов и администраторов. Величины p_{cp} и w_{cp} являются функциями числа приборов на отдельных участках.

В отличие от предыдущих случаев, эта задача имеет дело с дискретным изменением параметров, так как число приборов может изменяться только дискретным образом. Поэтому здесь неприменимо классическое вариационное исчисление, и необходимо применять специальные методы типа дискретного и целочисленного программирования.

Критерий минимума критического времени выполнения работы — это минимизация критического пути по графу при ограниченных ресурсах.

В связи с интенсивным внедрением сетевых методов планирования и управления возникают задачи о критическом пути. Например, выполнение сложного проекта сводится к

последовательности выполнения какого-то определенного количества работ. Для отображения этого процесса прибегают к геометрической интерпретации с помощью графа (рис. 6).

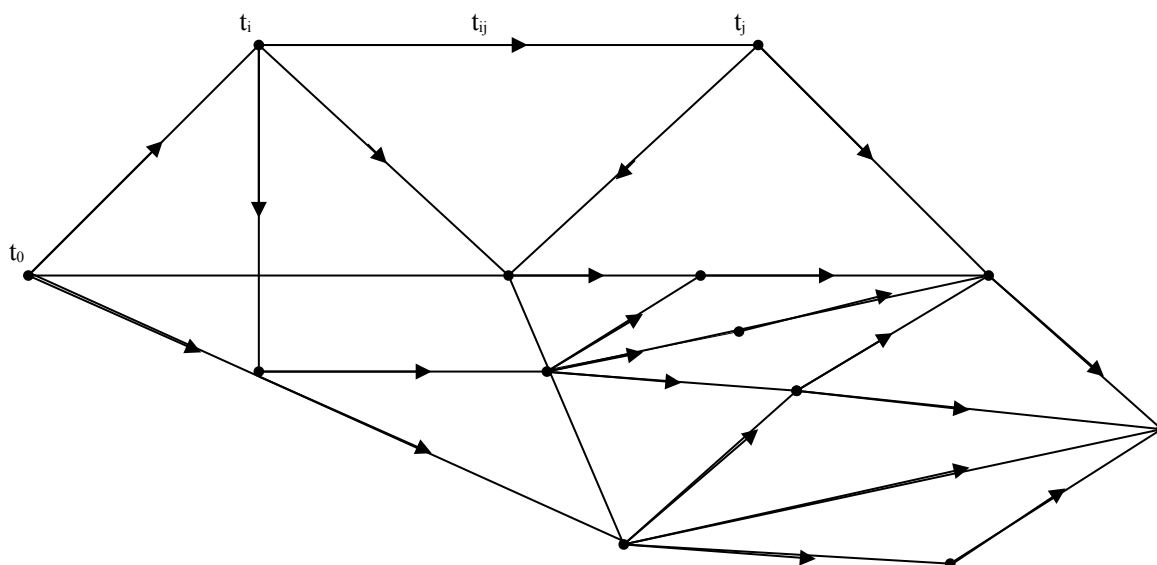


Рис. 6. Сетевой граф комплекса работ

Узел графа обозначает конец одной работы и начало другой. Дуга, или ребро, графа - работа, а длина ребра пропорциональна времени выполнения работы. Пусть длительность выполнения отдельной работы

$$t_{ij} = t_j - t_i,$$

где t_i и t_j — моменты времени, соответствующие началу и концу работы.

Критическим в графе будет такой путь, на котором суммарное время выполнения работ максимально. Этот путь задерживает всю разработку: без его уменьшения невозможно сократить время выполнения всей разработки. Критический путь определяется решением экстремальной задачи, в которой обращается в максимум следующее выражение:

$$t_{\text{кр}} = \sum_{i,j} t_{ij} \rightarrow \max.$$

Суммирование в этой двойной сумме необходимо производить таким образом, чтобы получить путь из начальной вершины t_0 в конечную t_k без разрыва и дважды не пройти по одной и той же дуге. Критерий минимума критического времени

$$I = t_{\text{кр}} \rightarrow \min$$

означает минимизацию критического пути при ограниченных ресурсах. При этом наличные ресурсы распределяются на те работы, которые лежат на критическом пути. В данном случае, по существу, решается минимаксная задача, характерная для теории игр:

$$\min \max \sum_{i,j} t_{ij} = ?,$$

причем максимум выбирается среди всех путей в графе по времени выполнения работ, а минимум берется по ресурсам для работ, лежащих на критическом пути.

Минимаксный критерий – используется для определения оптимальной стратегии при наличии ситуации, когда интересы двух сторон противоположны.

Минимаксный критерий широко используется для определения оптимальной стратегии при наличии конфликтной ситуации, когда интересы двух сторон противоположны, т.е. выигрыш одной стороны означает проигрыш другой. В этом случае часто приходится выбирать среди худших для себя стратегий противника наименее худшую, т.е. брать максимум по множеству

стратегий противника и минимум по собственным стратегиям. В этом и заключается минимаксный критерий, широко используемый в теории игр. В теории матричных игр задается матрица платежей

$$\|a_{ij}\|,$$

где $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$.

Каждый элемент этой матрицы означает платеж противнику в случае, когда он применяет стратегию j , а наша сторона - стратегию i . Требуется найти среди множества худших для нас стратегий противника наименее плохую, т.е. решить минимаксную задачу

$$\min_i \max_j a_{ij} = ?$$

Нетрудно убедиться, что если поменять знаки a_{ij} на обратные, а это физически означает замену проигрыша нашей стороны на выигрыш, то будет решаться максиминная задача

$$\max_i \min_j (-a_{ij}) = ?$$

В некоторых задачах, имеющих так называемую седловую точку,

$$\min_i \max_j a_{ij} = \max_i \min_j a_{ij}.$$

1.4. Классификация методов оптимизации

Классификация методов оптимизации представляет собой достаточно сложную задачу, так как в основном они исторически развивались независимо один от другого с использованием различных концепций, математических аппаратов и т.д. Однако специалисту по информатике и вычислительной технике необходимо ознакомиться с ними.

Разумеется, что приводимая классификация (как и любая классификация) носит условный характер, но в целом она позволяет сразу охватить все особенности методов.

Возможно несколько подходов к классификации. Следует различать методы определения экстремума функции и функционала (рис. 7). Поскольку функция является частным случаем функционала, методы отыскания экстремума функции проще. Методы динамического программирования и принципа максимума применяются для отыскания экстремума функционала и функции. Прямые методы вариационного исчисления (методы Рунге, Эйлера и др.), как и дискретный вариант уравнения Эйлера, сводят задачу отыскания экстремума функционала к экстремуму функции.

Методы отыскания экстремума функции получили большое развитие в связи с вычислительными трудностями решения системы алгебраических уравнений вида

$$dF(x_1, x_2, \dots, x_n)/dx_j = 0, j = 1, 2, \dots, n,$$

особенно при наличии ограничений на координаты x_j .

Классическая математика ограничивалась разработкой методов решения и доказательствами принципиальной разрешимости таких систем уравнений, что привело к созданию **аналитических методов оптимизации** (методы принципиальной разрешимости уравнений оптимизации). При решении конкретных задач важно владеть процедурами, позволяющими доводить решение до числовых данных. Это заставило искать и разрабатывать **численные методы оптимизации** (методы решения конкретных инженерных задач с доведением решения до числовых данных). Практика проектирования конкретных систем управления требует применения обоих методов.

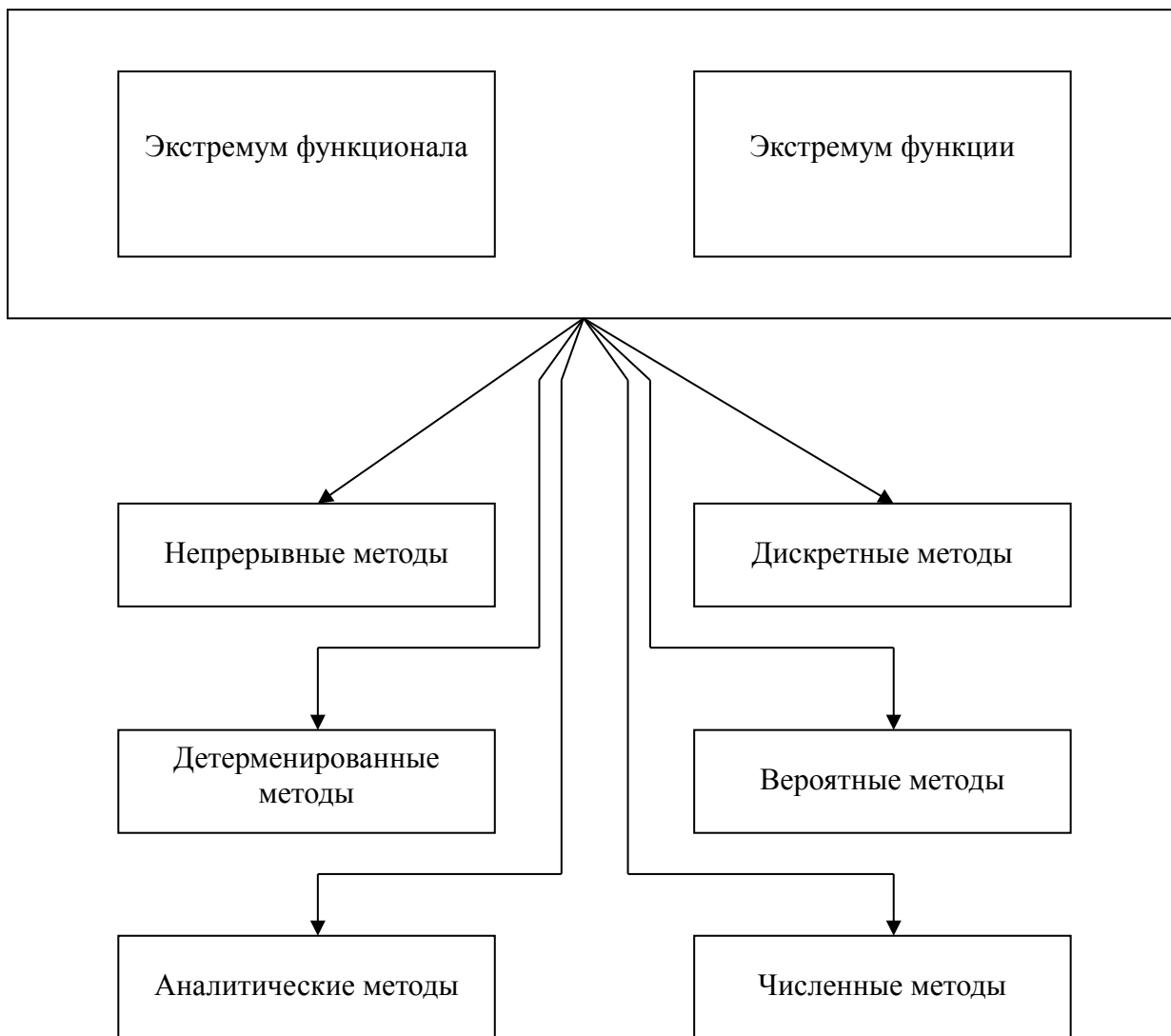


Рис. 7. Схема методов оптимизации

За аналитическими методами остается преимущество, заключающееся в возможности определения качественной картины поведения оптимальной системы при изменении ее параметров и структуры. Численные же методы обеспечивают получение конкретных числовых значений параметров. Рациональное объединение этих методов производится при разработке человеко-машинных методов оптимизации, использующих диалог человека и ЭВМ, базы данных и возможности современных телекоммуникационных систем. При этом обычно повторяют вычисления, изменяя условия, используя, где необходимо, аналитические методы, представленные в виде стандартных библиотечных программ, и, самое главное, оперативно включая в процедуру отыскания оптимального управления интеллектуальные способности человека.

При таком способе оптимизации исходный критерий оптимальности может быть математически нестрого формализован в виде функции или функционала. Так, он может состоять из нескольких положений, сформулированных достаточно четко на словесном, содержательном уровне, что при наличии диалога человек-машина вполне допустимо.

В задачах с ограничениями большое развитие получили методы линейного и нелинейного программирования, прямые методы отыскания экстремума функции, а также дискретные принципы максимума и динамическое программирование (рис. 8).

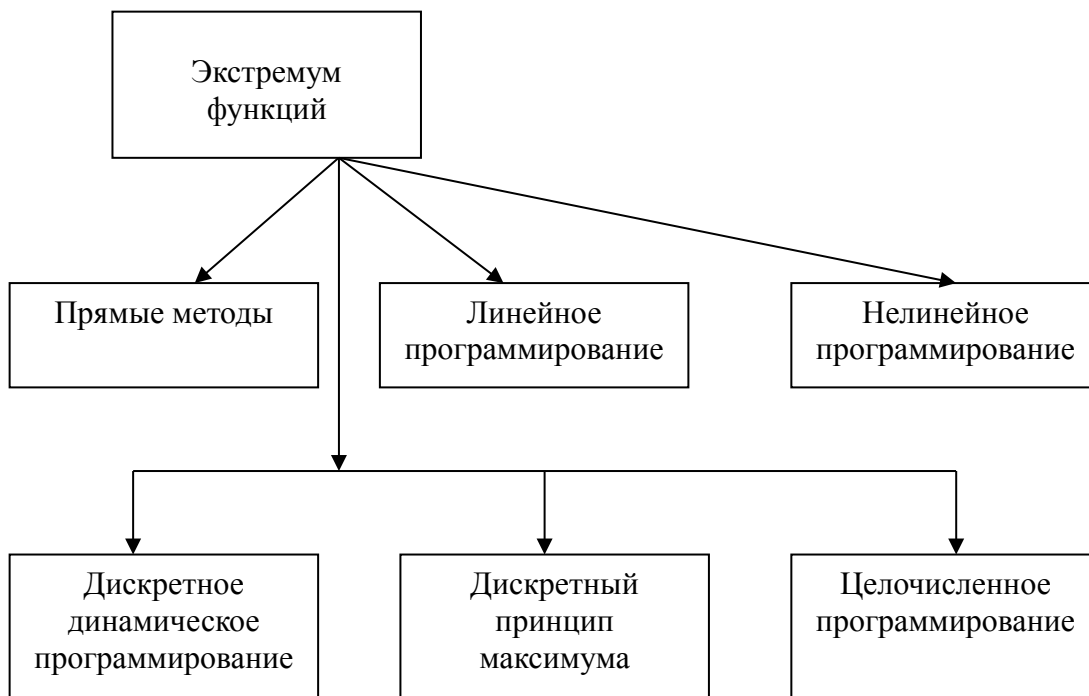


Рис. 8. Методы отыскания экстремумов функций

Методы отыскания экстремума функционала ведут свое начало от классических методов Эйлера-Лагранжа-Гамильтона и заканчиваются динамическим программированием и принципом максимума (рис. 9). В них также содержатся прямые методы отыскания экстремума функционала и различные частные методы.

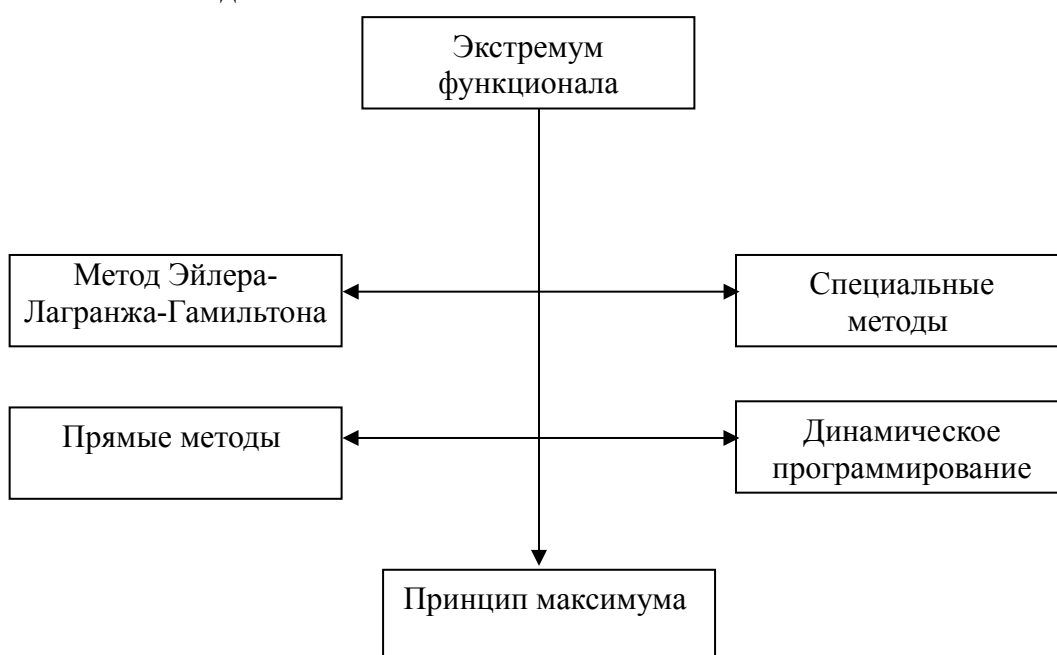


Рис. 9. Методы отыскания экстремумов функционалов

Почти во всех методах:

- Эйлера — Лагранжа — Гамильтона;
- принципе максимума;
- динамическом программировании;
- прямых методах отыскания экстремума функции

можно выделить дискретные и непрерывные модификации, детерминированные и случайные варианты.

Дискретная задача оптимизации возникает всякий раз, когда требуется найти оптимальное значение переменных, которые заданы, так же как и функция от них, в виде набора дискретных значений.

В связи с разработкой автоматизированных информационных систем управления (АИСУ) большое значение приобрели методы целочисленного программирования. Одна из основных задач АИСУ – оперативно-календарное планирование, относится к задачам целочисленного программирования. Единого общего метода решения задач целочисленного программирования нет. Разработано много способов решения пригодных для частного класса задач, среди которых большое место занимают нестрогие эвристические методы. Особого внимания среди них заслуживают **лингвистические методы оптимизации** (методы, имитирующие применяемые человеком методы оптимизации и планирования с добавлением эффективных аналитических и числовых процедур). Они ближе всего подходят к человеческому способу мышления и удобны для реализации на ЭВМ с помощью проблемно-ориентированных языков. Человеко-машинные методы решения целочисленных задач оптимизации являются по существу своему лингвистическими в широком смысле.

В заключение надо отметить, что одни и те же методы, рассматриваемые с разных точек зрения, могут использоваться как аналитические и как численные. Так, градиентный метод используется в прямых методах отыскания экстремума функции и методах нелинейного (квадратичного и выпуклого) программирования. Во втором случае он используется как аналитический, в первом – как численный метод.

1.5. Системный анализ и оптимизация

Системный анализ - это синтез идей и принципов теории исследования операций и методов теории управления с возможностями современной вычислительной техники. В настоящее время используются следующие понятия: “системный анализ”, “теория систем” и “системный подход”. Слово *система* и связанные с ним термины получили широкое распространение. Это произошло потому, что на передний план все более и более выступает необходимость изучения сложных комплексов (систем). Такая необходимость определяется усложнением создаваемых технических конструкций, устройств, технологий и всех совокупностей хозяйственных связей, с которыми приходится иметь дело экономистам, хозяйственным руководителям и инженерам.

Потребность изучения биологических объектов и проблем экологии, которые с каждым годом становятся все актуальнее, также приводит исследователя к сложным системам.

Понятие *система* относится к числу тех, для которых трудно дать строгое определение. Часто системой называют совокупность элементов, между которыми существуют те или другие связи (например система двух притягивающихся масс). Мы не будем пытаться давать строгое определение системы. Для наших целей достаточно того интуитивного понятия системы, которое имеется у каждого, изучающего инженерные дисциплины.

В ответ на потребности изучения сложных систем возникла дисциплина “системный анализ”. Одной из центральных проблем системного анализа является проблема принятия решений. Инструмента теории исследования операций для этих целей часто оказывается недостаточно. Как уже было сказано выше, неопределенность цели в задачах оптимизации состоит в многокритериальности. Трудно соизмерить и сопоставить между собой различные требования -

трудно формализовать понятие “цель”, объединить показатели. В ряде задач либо вообще нельзя сколько-нибудь точно поставить цели, либо те цели, которые хотелось бы поставить, нереальны (примеры подобных ситуаций дает порой экономика). В таких случаях на помощь приходит системный анализ.

Допустим, что речь идет о планах перспективного развития топливно-энергетического комплекса. Как определить цели? Конечно, всегда можно сформулировать требования: топлива побольше, затрат поменьше и т.д. Но для проекта плана необходимы более или менее точные показатели и реалистические цели, которые согласуются с потребностями страны и могут быть обеспечены существующими ресурсами. В таких проблемах самый главный момент - сформулировать цели, которые должен преследовать проект. Цель перестает быть внешним фактором, как, например, в теории исследования операций или теории управления, она становится самостоятельным объектом исследования.

Что надо исследователю для того, чтобы установить (правильно сформулировать) те реальные цели, осуществление которых должны обеспечивать создаваемая конструкция или хозяйственный комплекс? Очевидно, что для этого необходимо, прежде всего, представить себе функционирование будущей конструкции, сопоставить ее возможности с теми ресурсами, которыми будут располагать разработчики. Достичь этого можно с помощью физических (макеты будущих конструкций) или математических моделей.

Таким образом, чтобы использовать математику, сначала необходимо описать систему моделей и создать математический аппарат, который позволит провести анализ изучаемого процесса, увидеть последствия принятых решений, оценить возможности при различных альтернативах, и только на основе такого анализа можно сформулировать цели. Сложность изучаемых и проектируемых систем приводит к необходимости создания специальной техники исследования, использующей аппарат имитации, - воспроизведения на ЭВМ специально организованными системами математических моделей функционирования проектируемого или изучаемого комплекса.

Последнее не означает недооценки классических методов анализа. Более того, эффективное использование имитационных систем необходимо предполагает предварительную обработку модели. Исследование динамики процесса, позволяющее увидеть перспективы и наметить цели, - это лишь один из аспектов системного анализа, может быть, и самый важный, но не исчерпывающий всего многообразия вопросов, на которые он в состоянии дать ответ. Это всего лишь первый шаг исследования. Следующая проблема состоит в том, чтобы реализовать намеченные цели, т.е. сформулировать решения, в результате выполнения которых будет обеспечено достижение этих целей (выбраны параметры создаваемых конструкции или проекта).

Среди задач, возникающих в связи с созданием соответствующих проектов, большое место занимают проблемы сочетания структурных и функциональных аспектов. Один из трудных вопросов, связанных с этим, относится к проблемам проектирования иерархической организации. Любые более или менее сложные системы всегда организованы по иерархическому принципу в связи с тем, что централизованная обработка информации и принятие решений часто бывают невозможны из-за большого объема информации, которую следует собирать и перерабатывать, из-за возникающих при этом задержек и искажений и т.д. Если речь идет о проектировании технических систем, то задача исследования систем (задача проектировщика) состоит в разработке функциональной схемы, которая может быть реализована заведомо не единственным способом, и в определении частных целей.

Значительно сложнее обстоит дело, когда речь идет о народнохозяйственных комплексах, функционирование элементов которых зависит от того, как управляют ими люди. В отличие от машины, человек всегда имеет собственные цели и интересы, и проектировщику системы уже недостаточно только формулировать цели для нижних звеньев. Необходимо еще быть уверенным, что эти цели будут достигнуты, что нижние звенья выполняют требования верхних звеньев. Для этого, в свою очередь, должен быть спроектирован специальный “механизм”, одного приказа для достижения цели бывает недостаточно. Вот почему возникает потребность в специальной теории, которая должна развивать принципы создания иерархии в управлении и методы их анализа. Теория иерархических систем, которая изучает некоторые из аспектов этой проблемы, является одной из важнейших частей системного анализа.

Таким образом, системный анализ - это дисциплина, развивающая методы проектирования сложных технических, народнохозяйственных и экологических систем. Системный анализ как дальнейшее развитие теории исследования операций и теории управления включает в себя эти дисциплины со всем арсеналом средств, развитых в их рамках.

В настоящее время сложилась некоторая условная классификация математических моделей по характеру и способу использования произвольных функций и параметров, которые они содержат. В их число входят следующие виды моделей:

1. Модели без управления, которые описывают динамические процессы (с помощью, например, дифференциальных или разностных уравнений) и не содержат свободных параметров или функций. К их числу относится большинство прогностических моделей, когда заданное начальное состояние определяет траекторию процесса. Модели такого рода могут быть и стохастическими, например, они могут содержать случайные величины и функции вида

$$x' = f(x, t, \xi),$$

где x - некоторый случайный вектор с известным законом распределения. В этом случае исследуются не отдельные траектории, а их статистические свойства, например среднее значение. Модели подобного рода являются типичными для описания процессов, происходящих в неживой природе.

2. Модели, которые могут быть использованы для оптимизации некоторых действий. Например, динамический процесс, модель которого описывается уравнением вида

$$x' = f(x, t, u),$$

где выбор функции $u(t, x)$ находится в распоряжении какого-то субъекта. Вектор-функция $u(f, x)$ называется управлением. Управление выбирается из условия достижения некоторой цели. Распространенный класс задач с помощью данной модели можно описать следующим образом: за время T перевести систему из состояния

$$x(0) = x_0$$

в состояние

$$x(T) = x_T$$

так, чтобы “затраты” были минимальными, т.е.

$$\int_0^T F(x, u, t) dt \rightarrow \min.$$

Ограничения $x(0) = x_0$, $x(T) = x_T$ и целевую функцию в понятие модели не включают. Для одной и той же модели $x' = f(x, t, u)$ могут ставиться разные задачи.

3. Модели, которые могут использоваться для анализа конфликтных ситуаций. Предположим, что динамический процесс определяется действиями нескольких субъектов, в распоряжении которых имеются управления: u, v, w, \dots Тогда

$$x' = f(x', t, u, v, w, \dots),$$

причем управления будут выбираться из условий вида

$$\int_0^T F_i(x, u, v, w, \dots, t) dt \rightarrow \min,$$

каждое из которых отражает вполне определенные интересы того или другого субъекта.