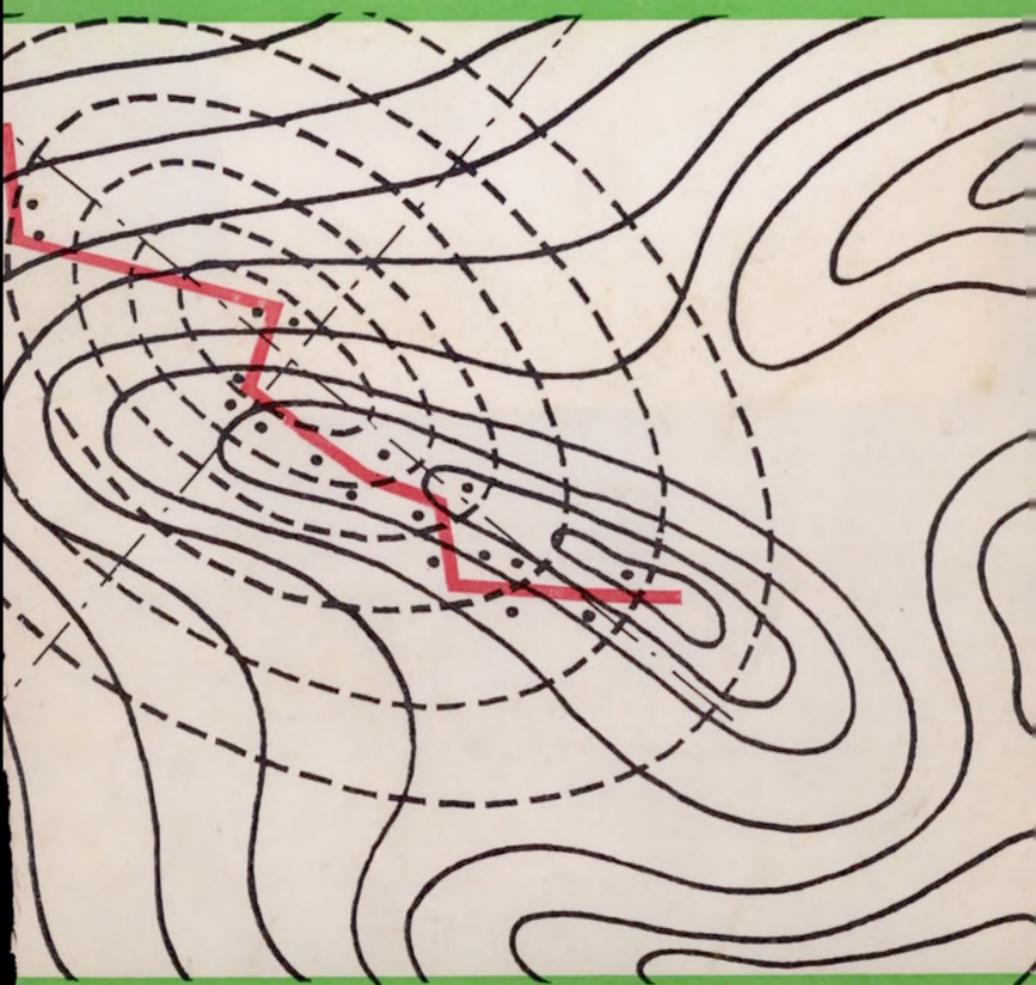


Уч. зап.

Г33

В. И. Геминтерн, Б. М. Каган

МЕТОДЫ ОПТИМАЛЬНОГО ПРОЕКТИРОВАНИЯ

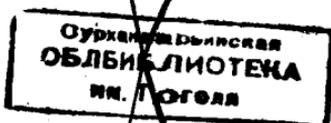
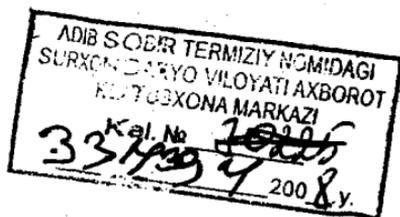


32.96.

Г33

В.И.Геминтерн, Б. М.Каган

МЕТОДЫ ОПТИМАЛЬНОГО ПРОЕКТИРОВАНИЯ



МОСКВА «ЭНЕРГИЯ» 1980

ББК 32.965

Г 33

УДК 658.012011.56

Геминтерн В. И., Каган Б. М.

Г 33 Методы оптимального проектирования. — М.: Энергия, 1980. — 160 с., ил.

45 к.

Содержит анализ и описание методов автоматизированного проектирования технических объектов, позволяющих определять оптимальные значения конструктивных параметров на основе технических и технико-экономических критериев и ограничений. Излагается подход к созданию расчетных моделей проектируемых объектов. Приводятся методы решения на ЭВМ задач, возникающих при оптимальном проектировании.

Для инженеров и научных работников, занимающихся вопросами применения вычислительной техники в проектно-конструкторских работах, а также для разработчиков АСУ. Она может быть полезна студентам и аспирантам, специализирующимся в области САПР и АСУ.

Г 30501-281
051(01)-80 161-80. 1502000000

ББК 32.965

6Ф6

© Издательство «Энергия», 1980 г.

ПРЕДИСЛОВИЕ

В настоящее время одним из главных путей повышения качества и эффективности проектно-конструкторских работ в машиностроении, строительстве и других областях является автоматизация проектирования на основе применения электронных вычислительных машин (ЭВМ). Книга посвящена важному направлению автоматизации проектно-конструкторских работ — оптимальному проектированию. Она содержит анализ и описание различных методов использования ЭВМ для определения оптимальных конструктивных параметров проектируемых объектов исходя из технических и технико-экономических критериев: оптимальности (минимальная стоимость, масса и т. п.) и ограничений.

Появление ЭВМ сделало практически возможным решение задач оптимального проектирования. Впервые автоматизированное оптимальное проектирование на ЭВМ было реализовано в широком масштабе в 1957 г. во ВНИИЭМ при проектировании серий электрических двигателей [48]. С тех пор оно получило широкое развитие в электротехнической промышленности (при проектировании электрических машин, трансформаторов, некоторых типов регуляторов и электроприводов) и во многих других отраслях народного хозяйства.

В результате усилий многих ученых-математиков и инженеров методы оптимального проектирования развились в важный раздел прикладной математики, а само автоматизированное оптимальное проектирование стало одной из наиболее эффективных областей применения ЭВМ в инженерном деле. Определение на ЭВМ оптимальных параметров проектируемого изделия дает прямой экономический и технический эффект.

В настоящее время ощущается необходимость в книге о методах автоматизации оптимального проектирования, адресованной инженерам, занятым разработкой новой техники. В книге рассматриваются различные ме-

тоды и алгоритмы решения на ЭВМ задач, возникающих при оптимальном проектировании (задач нелинейного программирования), излагаются подход к построению моделей проектируемых объектов, выбору технических и технико-экономических критериев оптимальности и ограничений, принципы построения программного обеспечения, ориентированного на автоматизацию оптимального проектирования. Приводятся примеры постановки и решения задач оптимального проектирования некоторых технических объектов.

Книга написана для широкого круга инженеров и студентов старших курсов вузов с целью помочь им эффективно использовать вычислительную технику в практической работе.

Авторы выражают благодарность д-ру техн. наук, проф. Д. А. Аветисяну, д-ру техн. наук, проф. Г. Н. Соловьеву и канд. физ.-мат. наук М. С. Штильману за полезные замечания, сделанные ими при чтении рукописи.

Авторы особенно благодарны за постоянное внимание и помощь в их работе по методам автоматизации оптимального проектирования академику АН Армянской ССР А. Г. Иосифьяну, по инициативе которого более четверти века назад были начаты исследования по применению ЭВМ в электротехнической промышленности.

Авторы

ПРОБЛЕМА ОПТИМАЛЬНОГО ПРОЕКТИРОВАНИЯ

1-1. ОБЩАЯ ПОСТАНОВКА ЗАДАЧИ ОПТИМАЛЬНОГО ПРОЕКТИРОВАНИЯ

При проектировании технических устройств важное значение имеет определение оптимальных вариантов конструкций машин и аппаратов, параметров схем, режимов работы технологического оборудования и т. д. Появление и быстрое развитие электронной вычислительной техники сделало практически возможным автоматизированное оптимальное проектирование.

Примерами задач оптимального проектирования являются: определение конструктивных параметров электрических двигателей, оптимальных по критерию минимальной стоимости; расчет элементов конструкций самолета максимальной грузоподъемности при заданной мощности двигателя и др. К рассматриваемому вопросу могут быть сведены многие задачи по синтезу оптимальных функций для систем управления (передаточных функций, нелинейных функций управления). В этом случае искомая функция аппроксимируется некоторой аналитической зависимостью (посредством рядов, дробно-рациональных функций и т. п.), коэффициенты которой определяются из условия оптимальности системы.

Задача поиска оптимальных значений параметров управления возникает во многих случаях применения управляющих вычислительных устройств для автоматической оптимизации технологических процессов. В подобных системах поиск оптимума осложняется действием помех, вносящих погрешность в измерения параметров производственного процесса.

Исходными данными при проектировании изделия являются указанные в техническом задании на разра-

ботку значения его номинальных параметров, требования стандартов и технологии изготовления к определенным характеристикам изделия.

Например, при проектировании электрического двигателя исходные данные включают номинальные значения мощности, питающего напряжения, частоты вращения, допустимое значение перегрева обмоток, требование к перегрузочной способности и коэффициенту мощности двигателя и некоторые другие данные.

В зависимости от назначения проектируемого изделия может быть указан критерий качества (или эффективности) для разрабатываемой машины, схемы или процесса. Естественным является стремление в процессе проектирования получить наилучшее значение этого критерия. Поэтому такой критерий будем называть *критерием оптимальности* проектируемого объекта.

В качестве критерия оптимальности в зависимости от характера и назначения проектируемого объекта может быть принята его стоимость, масса, потребляемая мощность, КПД или другая, более сложная характеристика.

Уменьшение (или увеличение) значения критерия оптимальности стеснено необходимостью выполнения требований стандартов, технических условий и технологии изготовления, ограничивающих значения определенных параметров и характеристик проектируемого изделия. Например, при проектировании электрического двигателя стремлению к экономии обмоточного провода ставит предел заданное техническими условиями ограничение по перегреву обмотки двигателя.

При оптимальном проектировании необходимо определить и обосновать критерий оптимальности проектируемого изделия и четко выделить показатели и характеристики R_1, R_2, \dots, R_m , принимаемые в качестве ограничений.

Для проведения расчетов проектируемого объекта выделяют некоторую совокупность независимых конструктивных параметров (независимых переменных)

$$x_1, x_2, \dots, x_n,$$

значения которых однозначно определяют все остальные параметры и характеристики изделия, в том числе значения критерия оптимальности, а также показателей, принятых в качестве ограничений.

Переменные x_1, x_2, \dots, x_n представляют собой, например, геометрические размеры изделия или значения

других конструктивных параметров, параметры элементов схем, настройки регуляторов и т. п.

Критерий оптимальности является функцией независимых конструктивных параметров:

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

Показатели и характеристики, на значения которых наложены ограничения, являются также функциями независимых конструктивных параметров:

$$R_1 = R_1(x_1, x_2, \dots, x_n);$$

$$R_2 = R_2(x_1, x_2, \dots, x_n);$$

$$R_m = R_m(x_1, x_2, \dots, x_n).$$

Функции R_1, R_2, \dots, R_m будем в дальнейшем называть *функциями ограничений*.

В реальных задачах проектирования функция критерия оптимальности и функции ограничений (или, по крайней мере, часть функций ограничений) зависят нелинейно от независимых конструктивных параметров. Нелинейность критерия оптимальности и функций ограничений является важнейшей особенностью рассматриваемой проблемы.

В формализованном виде задача оптимального проектирования в общей постановке заключается в определении значений независимых переменных (независимых конструктивных параметров)

$$x_1, x_2, \dots, x_n,$$

при которых критерий оптимальности проектируемого объекта

$$F = F(x_1, x_2, \dots, x_n), \quad (1-1)$$

являющийся нелинейной функцией переменных, имеет минимально (или максимально) возможное значение при условии, что переменные x_1, x_2, \dots, x_n принимают лишь положительные значения, т. е.

$$x_j \geq 0, \quad j = 1, 2, \dots, n, \quad (1-2)$$

и выполняются ограничения, заданные в форме неравенств, для некоторых, в общем случае нелинейных, функций этих переменных (функций ограничений)

$$R_i(x_1, x_2, \dots, x_n) \leq 0, \quad i = 1, \dots, m; \quad m \leq n. \quad (1-3)$$

Ограничения, заданные в форме уравнений, можно использовать для исключения переменных и, таким образом, получить задачу с одним типом ограничений в виде неравенств.

В дальнейшем, если это особо не оговорено, будем предполагать, что функции R_i и F — непрерывные и дифференцируемые.

Ограничения вида

$$R_i(x_1, x_2, \dots, x_n) \geq 0$$

путем умножения левой части неравенства на минус единицу приводятся к форме (1-3).

Если на переменные наложены ограничения вида $a_j \leq x_j \leq b_j$ или $|x_j| \leq m_j$, то соответствующим преобразованием координат эти ограничения могут быть приведены к виду (1-2) и (1-3).

Совокупность формул, позволяющих для заданного набора значений конструктивных параметров $x_1, x_2, \dots, \dots, x_n$ рассчитать изделие и определить все его характеристики, в том числе значения функций ограничений и критерия оптимальности, называется *математической моделью проектируемого изделия*.

Математическая модель изделия разрабатывается на основе теории процессов функционирования изделия, методик его расчета. Уточнение модели может производиться путем сравнения результатов расчетов с характеристиками, полученными экспериментальным путем.

Математическая модель отличается от других способов описания объекта проектирования строгостью, компактностью. Использование математической модели позволяет во многих случаях заменить продолжительное и дорогостоящее экспериментирование с макетами проектируемых изделий серией расчетов с применением ЭВМ.

В математическом отношении оптимальное проектирование сводится к задачам *нелинейного программирования*. Этот класс задач значительно сложнее широко используемого в экономических оптимизационных расчетах линейного программирования, в котором предполагается, что критерий оптимальности и ограничения являются линейными функциями независимых переменных.

Прибегнем к геометрической интерпретации задачи нелинейного программирования.

В n -мерном пространстве переменных (конструктивных параметров) каждой точке x^r пространства соответствует совокупность вещественных чисел $x^r_1, x^r_2, \dots, x^r_n$, равных проекции вектора, проведенного из начала координат в точку x^r на координатные оси пространства x_1, x_2, \dots, x_n .

Функция F в каждой точке пространства имеет определенное значение, и, следовательно, n -мерное пространство является скалярным полем критерия оптимальности F . В рассматриваемом пространстве может быть построено семейство эквидотенциальных гиперповерхностей или гиперповерхностей постоянного уровня, соответствующих определенным значениям функции F .

Пространство переменных x_1, x_2, \dots, x_n является также скалярным полем функций ограничений R_i , и в нем могут быть построены гиперповерхности уровня функций R_i . Функциям ограничения соответствуют граничные гиперповерхности (в частном случае, гиперплоскости)

$$R_i(x_1, x_2, \dots, x_n) = 0, \quad i = 1, 2, \dots, m,$$

являющиеся $(n-1)$ -мерным многообразием.

Ограничениям (1-2) соответствуют гиперплоскости

$$x_i = 0, \quad i = 1, 2, \dots, n,$$

выделяющие в n -мерном пространстве первый пространственный квадрант.

Если условия (ограничения) заданы в виде уравнений, то переменные x_1, x_2, \dots, x_n могут выбираться лишь среди координат точек, лежащих на пересечении гиперповерхностей всех ограничений одновременно. В этом случае задача сводится к отысканию относительных минимумов (максимумов) и может (по крайней мере, принципиально) решаться методом множителей Лагранжа. На практике имеем дело с более общим и сложным случаем задания ограничений в форме неравенств, причем нелинейных.

Гиперповерхности, соответствующие нелинейным неравенствам ограничений, образуют в n -мерном пространстве сферический многогранник, выделяющий в первом квадранте область допустимых значений переменных (конструктивных параметров).

Сказанное выше иллюстрируется на рис. 1-1 для случая двух переменных. В плоскости $x_1 O x_2$ нанесены линии

уровня функции $F(x_1, x_2)$, причем $F_1 > F_2 > F_3 > \dots$, и функций ограничений $R_1(x_1, x_2)$, $R_2(x_1, x_2)$, $R_3(x_1, x_2)$.
Условия

$$\begin{aligned} x_1 \geq 0, x_2 \geq 0; R_1(x_1, x_2) \leq 0; \\ R_2(x_1, x_2) \leq 0; R_3(x_1, x_2) \leq 0 \end{aligned}$$

выделяют область допустимых значений переменных — область $OABCD$.

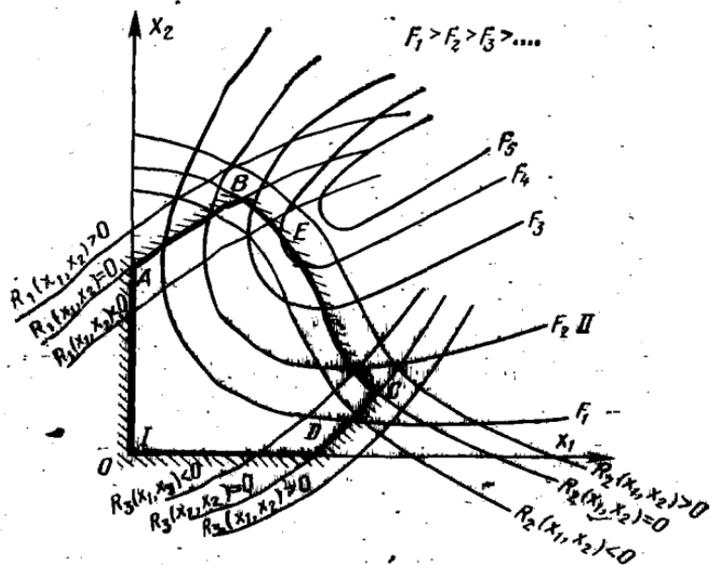


Рис. 1-1. Задача нелинейного программирования (случай двух переменных).

В рассматриваемом примере наименьшее возможное значение функция $F(x_1, x_2)$ принимает на границе области I допустимых значений переменных в точке E .

Встречаются задачи, в которых некоторые переменные могут принимать лишь определенные дискретные значения (например, диаметр обмоточного провода, выбираемый из определенного сортамента) либо только целочисленные значения (например, число проводов в пазу электрической машины).

В общем случае задачи нелинейного программирования ничего нельзя сказать заранее о расположении точки, в которой функция F имеет минимально (максимально) возможное значение. Эта точка может нахо-

даться как на поверхности сферического многогранника, образованного поверхностями ограничений, так и внутри его. Функция F может достигать экстремального значения не в определенной точке, а на некотором множестве точек (гиперлинии или гиперповерхности).

Другая особенность задачи нелинейного программирования (часто встречающаяся в практических задачах) состоит в том, что в области допустимых значений x_1, x_2, \dots, x_n оптимизируемая функция может иметь не один, а несколько локальных экстремумов. Решение задачи нелинейного программирования состоит в определении *глобального экстремума*, т. е. наименьшего (или наибольшего) значения $F(x_1, x_2, \dots, x_n)$ во всей области допустимых значений переменных.

Поиск максимума функции $\Phi(x_1, x_2, \dots, x_n)$ эквивалентен минимизации функции $F(x_1, x_2, \dots, x_n) = -\Phi(x_1, x_2, \dots, x_n)$. Поэтому для определенности в дальнейшем будем говорить о поиске минимального значения критерия оптимальности, предполагая, что в противном случае произведено указанное преобразование.

Формализация понятий допустимой области и оптимальных значений конструктивных параметров. Каждая точка x пространства E^n (в данном случае n -мерного пространства конструктивных параметров), в которой выполняются условия (1-2) и (1-3), называется *допустимой точкой* (или допустимым решением). Множество всех допустимых точек (решений) называется *допустимой областью* G . Допустимая область может быть непрерывной или дискретной. Она может быть односвязной, как это показано на рис. 1-1, или многосвязной, т. е. распадаться на отдельные изолированные подобласти.

Точка x^* называется локальным минимумом (локальным оптимумом), если неравенство

$$F(x^*) \leq F(x) \quad (1-4)$$

справедливо для всех точек из области G , которые находятся в некоторой окрестности точки x^* ¹.

¹ Окрестностью U точки $x \in E^n$ является множество таких точек $y \in E^n$, для которых норма разности $|x-y|$ не превышает некоторое положительное число, т. е.

$$|x-y| \leq r, \quad r > 0.$$

Внутри области допустимых значений переменных, включая его границу, может находиться несколько локальных оптимумов функции F .

Если точка x^* такова, что условие (1-4) оказывается справедливым для всех точек из области¹ G , то точка x^* называется *глобальным экстремумом*. (оптимумом) в области G .

Формирование критерия оптимальности. Важнейшим моментом в оптимальном проектировании является определение и обоснование критерия оптимальности.

Выбор критерия оптимальности значительно облегчается, если характер и назначение проектируемого изделия таковы, что среди его характеристик (показателей) имеется характеристика (показатель), которая со всей очевидностью определяет качество изделия. Такое положение часто имеет место при проектировании агрегатов бортового оборудования, для которых в зависимости от условий применения очевидными критериями оптимальности могут служить, например, значение массы, потребляемая мощность, КПД.

Во многих практических проектных задачах вопрос оказывается значительно сложнее: качество изделия оценивается значениями целого ряда экономических и технических характеристик. В этом случае оптимальное проектирование должно обеспечивать улучшение показателей не по одной, а сразу по нескольким характеристикам. Такого рода задачи называют *многокритериальной оптимизацией* [34].

Примитивным подходом является выбор одной из функций, описывающих модель изделия, в качестве критерия оптимальности и рассмотрение всех других как ограничений. Более общий подход — разбиение указанного множества функций на критерии и ограничения.

Пусть $\varphi_1(x)$, $\varphi_2(x)$, ..., $\varphi_p(x)$ — выделенный в результате такого разбиения набор технико-экономических характеристик, которые должны быть улучшены в результате оптимального проектирования. Выбор комплексного критерия оптимальности состоит в формировании такой функции $F = F(\varphi_1(x), \varphi_2(x), \varphi_p(x))$, максимальное (минимальное) значение которой соответствует наиболее эффективному проектному решению.

В общем случае формирование комплексного критерия оптимальности является весьма сложной задачей,

¹ Принадлежность точки x^* области G обозначается $x^* \in G$.

требующей всестороннего анализа требований к проектируемому объекту.

Одним из наиболее распространенных способов формирования такого критерия является аддитивный учет критериев φ_i , $i=1, 2, \dots, p$, при котором $F(x)$ представляется с помощью линейной комбинации:

$$F(x) = \sum_{i=1}^p c_i \varphi_i(x),$$

где весовые коэффициенты c_i определяются на основе экспертных оценок. Такой метод удобен, когда критерии φ_i измерены в одинаковых единицах или для их оценки используются относительные величины.

Примером подобного критерия может служить критерий оптимальности приведенных затрат при оптимальном проектировании асинхронных двигателей [12]. Этот критерий включает затраты на изготовление двигателя, а также на его эксплуатацию (затраты на электроэнергию) в течение определенного периода.

Следующий подход к решению многокритериальной задачи оптимизации состоит в упорядочении критериев по степени их важности и введении допусков Δ_i по каждому из критериев. Оптимизацию производят сначала по наиболее важному критерию φ_1 . Пусть минимальное значение этого критерия равно φ^*_1 . Затем в модель вводится ограничение

$$\varphi_1(x) \leq \varphi^*_1 + \Delta_1 \quad (1-5)$$

и осуществляется поиск оптимума по второму критерию $\varphi_2(x)$.

Если φ^*_2 — минимум по второму критерию, то в модель добавляется ограничение

$$\varphi_2(x) \leq \varphi^*_2 + \Delta_2$$

и осуществляется поиск оптимума по критерию $\varphi_3(x)$ и т. д.

Использование вектора допусков $\Delta = (\Delta_1, \Delta_2, \dots, \Delta_p)$ позволяет также ввести такой комплексный критерий, чтобы производился поиск варианта изделия, характеристики которого наименее всего отклонялись бы от значений $\varphi^*_1, \varphi^*_2, \dots, \varphi^*_p$ (взятых с соответствующими

допусками). В частности, этому условию удовлетворяет критерий

$$F(x) = \min_{1 \leq i \leq p} \max_x (\varphi_i(x) - \varphi_i^* - \Delta_i). \quad (1-6)$$

В результате проблема формирования критерия оптимальности сводится к выбору вектора допусков. Выбор численных значений допусков может существенно повлиять на характеристики проектируемого изделия. Необоснованное ужесточение допуска по одному критерию может существенно ухудшить качество изделия и не позволит добиться улучшений по другим критериям.

Поскольку конструктор нередко формирует свое представление о критерии оптимальности и оптимальном варианте изделия в процессе проектирования, целесообразно корректировать определение комплексного критерия оптимальности по результатам серии оптимизационных расчетов. В качестве исходной информации для формирования критерия могут использоваться значения исходных критериев $\varphi_1(x)$, $\varphi_2(x)$, ..., $\varphi_p(x)$, соответствующие допустимым значениям проектируемых параметров. Такой подход носит название адаптивного способа формирования критерия.

Поиск оптимума при адаптивном формировании критерия оптимальности разбивается на две стадии: построение комплексного критерия на основе вычислительных экспериментов и поиск оптимума по выбранному критерию. Указанные стадии могут многократно повторяться. При этом для построения комплексного критерия осуществляется вычисление гиперповерхности отклика с помощью методов планирования эксперимента [4].

Адаптивные методы формирования критерия получают в настоящее время значительное распространение благодаря применению в проектировании диалоговых режимов взаимодействия с ЭВМ, позволяющих оперативно пересчитывать оптимум, более полно использовать опыт проектировщика, в наглядной форме представлять результаты расчетов.

1-2. КЛАССИФИКАЦИЯ МЕТОДОВ РЕШЕНИЯ ЗАДАЧ ОПТИМАЛЬНОГО ПРОЕКТИРОВАНИЯ

При создании новых видов изделий возникают задачи оптимального проектирования в самой общей постановке, когда критерий оптимальности и ограничения

являются нелинейными функциями конструктивных параметров. В общем случае эти задачи являются многоэкстремальными и поэтому требующими определения глобального экстремума.

Если данная конкретная задача оптимального проектирования является одноэкстремальной, на практике, как правило, этот факт заранее, до ее решения, не может быть установлен из-за сложности функций критерия оптимальности и ограничений. Поэтому даже к таким задачам приходится подходить как к многоэкстремальным.

Решение задачи оптимального проектирования достаточно сложно. В последние годы развитию методов решения этих задач (задач нелинейного программирования) уделялось большое внимание. Был предложен ряд методов, позволяющих получить (точное или приближенное) решение для специальных классов задач. Однако, по-видимому, вообще невозможно создать метод, позволяющий получить точное решение общей задачи нелинейного программирования за конечное число итераций (шагов). Теоретически сложно доказать сходимость для процедур, применяемых при практических расчетах.

В теоретических работах исследуются главным образом частные случаи общей задачи, например минимизация вогнутой функции при выпуклой области допустимых значений переменных, задачи с линейными ограничениями, с квадратичным критерием оптимальности и т. п.

Методы численного решения задач нелинейного программирования с помощью ЭВМ продолжают создаваться. В то же время многие практически важные задачи оптимального проектирования уже удалось решить на ЭВМ.

На рис. 1-2 представлена классификация методов решения задач оптимального проектирования для детерминированных моделей проектируемых объектов, в которых все параметры изменяются непрерывно.

Каждый уровень классификационной таблицы, представленной на рис. 1-2, соответствует либо некоторому этапу расчета, либо конкретному методу, применяемому на этом этапе. Пунктирными вертикальными линиями разделены этапы поиска, входящие в качестве составных элементов в процедуру поиска более высокого уровня.

ня. Так, поиск глобального оптимума включает в себя один из способов генерации начальных точек, а также поиск локальных оптимумов. Поиск локального оптимума состоит из следующих этапов:

- выбор формы представления математической модели;
- определение направления движения к оптимуму;
- определение длины шага поиска.

Поиск глобального оптимума										
Поиск локального оптимума								Выбор начальных точек		
Метод определения длины шага поиска		Метод представления математической модели				Метод определения направления движения к оптимуму				
Параболической аппроксимации	"Золотого сечения"	Формирование обобщенного критерия оптимальности	Барьерные функции	Штрафные функции	Метод проекций в переменной метрике	Линейное локальное моделирование	Проекционно-градиентный метод	"Сравнитель"	Оценка градиента	Случайного поиска
	Детерминированный	Последовательные испытания с адаптацией	Использование априорной информации	Последовательность независимых испытаний						

Рис. 1-2. Классификация методов поиска оптимума при оптимальном проектировании.

Сплошной вертикальной линией на рис. 1-2 разделены альтернативные варианты. Например, формирование обобщенного критерия оптимальности осуществляется одним из методов: штрафных или барьерных функций.

Поскольку решение общей задачи оптимального проектирования происходит обычно в условиях отсутствия информации о числе локальных оптимумов, поиск всегда направлен на выявление глобального оптимума.

В основе всех алгоритмов поиска глобального опти-

34394
2-51
40225

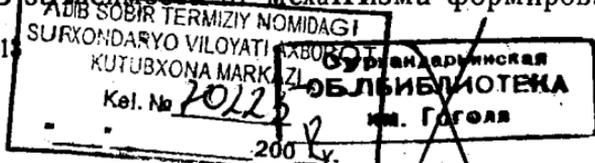
мума лежит получение и использование информации о расположении локальных экстремумов. Алгоритмы поиска глобального оптимума различаются по способу генерации начальных точек. Простейшими являются так называемые последовательности независимых испытаний. Каждое испытание состоит в вычислении в некоторой допустимой точке значения критерия оптимальности и сравнении его со значениями, полученными в предыдущих точках поиска. Последовательности испытаний могут быть детерминированными (например, обход узлов ~~и~~ мерной сетки в пространстве конструктивных параметров) или стохастическими. Для обоих случаев можно указать правила прекращения поиска после достаточного числа испытаний.

Следующая группа методов поиска глобального экстремума основана на использовании априорной информации о характере критерия оптимальности. Обычно эти методы пригодны лишь для специальных классов задач, поскольку указанная информация может быть получена лишь на основе исследования специфики проектируемого объекта. В дальнейшем рассмотрим два априорных условия: Липшица и сепарабельности критерия оптимальности. При справедливости первого из них оказывается возможным перейти к перебору значений конструктивных параметров на неравномерной сетке (с учетом предыдущих значений критерия оптимальности), а при справедливости второго — применить методы динамического программирования.

Наиболее общим и в то же время достаточно эффективным подходом к поиску глобального оптимума является использование последовательных испытаний с адаптацией. Адаптация в данном случае состоит в постепенном автоматическом совершенствовании процедуры поиска (сокращении числа испытаний) по мере накопления информации, в частности процедуры выбора начальных точек, а также направлений поиска локальных экстремумов.

Следующий уровень классификации охватывает методы поиска локального оптимума. Методы поиска локального оптимума включают: выбор способа представления математической модели для поиска, процедуру оценки направления на экстремум и определение длины шага поиска.

Взаимности от механизма формирования пробных



точек способы оценки направления движения к оптимуму разделяются на *детерминированные* и *стохастические*. Стохастические методы содержат алгоритмы выбора направлений на основе некоторой случайной функции с заданным распределением.

Схема отражает два существенно различных подхода к выбору способа представления математической модели для поиска оптимума.

Первый из них, метод штрафных функций, объединяет критерий оптимальности и ограничения, в один обобщенный критерий оптимальности. При приближении к границе допустимой области (или при выходе из нее) обобщенный критерий оптимальности начинает резко возрастать за счет штрафных коэффициентов, и поиск автоматически возвращается в допустимую область. Наиболее часто используемыми на практике методами формирования обобщенного критерия оптимальности являются метод штрафных функций и метод барьеров.

Второй подход, объединяющий методы возможных направлений, основан на разделении этапов определения направления убывания критерия оптимальности и коррекции этого направления в соответствии с ограничивающими условиями (на основе проекций градиента на допустимую область). В качестве направления движения к оптимуму наиболее часто используется антиградиент критерия оптимальности, а способы построения проекций различаются по признакам, которые положены в основу проекции. Так, построение направления с максимальной длиной проекции осуществляется методами квадратичного программирования, проектирование на ограничивающие гиперповерхности — проекционным градиентным методом и т. д. Методы возможных направлений наиболее эффективны при линейных ограничениях. В случае нелинейных ограничений используются их различные модификации.

Трудности, возникающие при переходе к нелинейным ограничениям, обусловлены тем, что проектирование на криволинейную поверхность приводит к необходимости ее кусочно-линейной аппроксимации вдоль направления движения. В результате значительно снижается эффективность методов и возрастает погрешность вычислений. В то же время формирование обобщенного критерия оптимальности дает наиболее эффективные результаты при нелинейных ограничениях.

Синтезом рассмотренных подходов является построение обобщенного критерия оптимальности только для нелинейных ограничений. При наличии в исходной задаче линейных ограничений используются методы возможных направлений. В этом случае допустимая область представляет собой многогранник, а обобщенный критерий оптимальности учитывает все оставшиеся ограничения.

Алгоритмы поиска локального оптимума x^* являются, как правило, итеративными, т. е. порождают последовательность векторов $\{x^k\} = x^1, x^2, \dots, x^k$, сходящуюся к вектору x^* .

Говорят, что вектор x^* является пределом сходящейся последовательности $\{x^k\}$, если для любого $\varepsilon > 0$ найдется такой номер N , что при $k > N$ выполняется неравенство $|x^k - x^*| < \varepsilon$. Отсюда следует, что допустимая область G должна вместе с любой сходящейся последовательностью содержать и ее предел. Такая область называется *замкнутой*. Примером замкнутой области может служить множество всех точек, которые удовлетворяют ограничениям (1-2) и (1-3).

Множество всех точек пространства E^n , которые не содержатся в замкнутой области G (этот факт обозначается $E^n \setminus G$), называется *открытым*. Каждая точка открытого множества M обладает окрестностью, полностью содержащейся в M . Точки, обладающие указанным свойством, называются *внутренними*. Следовательно, в открытом множестве M все точки являются внутренними. В отличие от открытых множеств в замкнутой области G (если она не совпадает со всем пространством E^n) имеются точки, в любой окрестности которых имеются точки из $E^n \setminus G$. Такие точки области G называются *граничными*. Множество всех граничных точек образует границу области G . В частности, если область G определяется условиями (1-2) и (1-3), его границу составляют те точки, в которых хотя бы одно из ограничений выполняется как строгое равенство.

Эффективность методов поиска локального оптимума определяется их скоростью сходимости к x^* . Говорят, что последовательность $\{x^k\}$ сходится к x^* быстрее последовательности $\{y^k\}$, если начиная с некоторого $k > N$ выполняется неравенство

$$|x^k - x^*| < |y^k - x^*|, \quad k = N+1, N+2, \dots \quad (1-7)$$

Критериями оценки качества выбора направления по ряду обращений к модели обычно являются:

- 1) улучшение значения критерия оптимальности по сравнению с имеющимся в данной точке;
- 2) наиболее быстрое убывание (возрастание) критерия оптимальности в окрестности данной точки;
- 3) наиболее вероятное расположение экстремума с учетом кривизны гиперповерхности, представляющей критерий оптимальности.

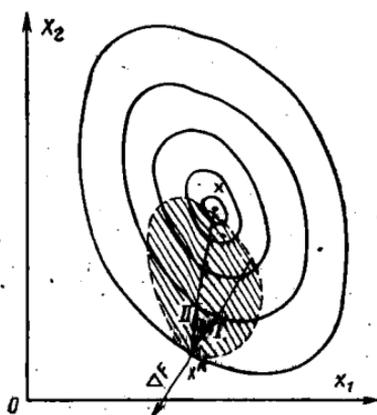


Рис. 1-3. Выбор направления поиска.

I — направление антиградиента; *II* — направление, выбранное с учетом кривизны критерия оптимальности (метод Ньютона).

Использование указанных трех критериев показано на рис. 1-3. Выбор направления по критерию 1 является неоднозначным. Заштрихована область, внутри которой все направления из точки x^k позволяют улучшить значение критерия оптимальности. Однако в зависимости от выбора направления критерий оптимальности при фиксированном шаге убывает на различное значение. Наибольшее значение убывания достигается при движении в сторону, противоположную градиенту (направление антиградиента), т. е. при выборе направления по критерию 2.

Градиентом называется вектор

$$\nabla F^T(x) = \left(\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \dots, \frac{\partial F}{\partial x_n} \right), \quad (1-8)$$

координатами которого служат частные производные критерия оптимальности по переменным x_1, x_2, \dots, x_n .

Направление вектора-градиента соответствует направлению наибольшей скорости возрастания критерия оптимальности.

Во многих случаях движение вдоль антиградиента до минимального значения (метод наискорейшего спуска) позволяет достаточно быстро определить оптимальное значение. Но наиболее точная оценка направления на экстремум может быть получена при учете кривизны критерия оптимальности, определяемой с помощью гессiana (метод Ньютона, методы переменной метрики)

$$\Gamma = \begin{vmatrix} \frac{\partial^2 F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 F}{\partial x_1 \partial x_n} \\ \frac{\partial^2 F}{\partial x_1 \partial x_2} & \frac{\partial^2 F}{\partial x_2^2} & \dots & \frac{\partial^2 F}{\partial x_2 \partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 F}{\partial x_1 \partial x_n} & \frac{\partial^2 F}{\partial x_2 \partial x_n} & \dots & \frac{\partial^2 F}{\partial x_n^2} \end{vmatrix}, \quad (1-9)$$

т. е. на основе критерия 3.

С его помощью удается учитывать кривизну гиперповерхности уровня критерия оптимальности по различным направлениям и определять наиболее рациональную стратегию поиска.

Использование каждого из трех критериев выбора направления движения к оптимуму требует различного числа обращений к модели. Ввиду того что математические модели объектов являются довольно сложными и каждое обращение к ним (вычисление в одной точке критерия оптимальности и проверка ограничений) занимает значительное время работы ЭВМ, необходимо очень внимательно подходить к выбору критерия оценки направления. Необоснованное усиление критерия выбора направления поиска может привести к резкому возрастанию числа обращений к модели, а ослабление — к беспорядочному сканированию в окрестности оптимума. В обоих случаях увеличивается время счета на ЭВМ.

При выборе направления на экстремум обычно принимается следующая стратегия. На начальных итерациях, когда с большой вероятностью уже несколько первых вычислительных значений критерия оптимальности дают информацию о направлении движения к экстремуму, используется критерий 1. Реализация этого критерия достигается использованием прямых детерминированных методов и методов случайного поиска.

Затем по мере приближения к экстремуму и усложнения геометрии критерия оптимальности используются методы, основанные на оценке градиента (стохастическая аппроксимация, градиентные методы, а также методы планирования эксперимента), соответствующие критерию 2. Наконец, в окрестности оптимального значения используются наиболее трудоёмкие, но и наиболее информативные методы («овражные», переменной метрики), реализующие критерий 3 выбора направления на экстремум.

В рассмотренной стратегии поиска допускаются исключения при резко меняющихся значениях критерия оптимальности в процессе поиска оптимального решения. Локальная информация о характере изменения критерия оптимальности быстро устаревает при движении к оптимуму. В частности, использование в подобных случаях таких характеристик, как градиент и гессиан критерия оптимальности, не ускоряет движение к экстремуму. Для такого класса функций оказываются эффективными специальные овражные методы, использующие адаптацию в процессе поиска.

С методами выбора направления на экстремум тесно связаны методы определения длины шага поиска, хотя на некоторых, особенно начальных, стадиях поиска достаточно эффективным оказывается движение с фиксированным шагом либо с шагом, пропорциональным модулю градиента. В общем случае метод выбора длины шага существенно влияет на скорость сходимости алгоритма. Обычно шаг поиска вдоль выбранного направления производится либо до границы допустимой области, либо до достижения минимального значения критерия оптимальности по заданному направлению. Поэтому алгоритмы формирования длины шага поиска представляют итерационные процедуры, включающие проверку условий допустимости всех промежуточных значений. Наиболее распространенные методы поиска минимума вдоль заданного направления (методы однопараметрической оптимизации): Фибоначчи, «золотого сечения» и квадратичной аппроксимации.

При нахождении оптимального решения следует в первую очередь разделить параметры на дискретные x_1, x_2, \dots, x_p и непрерывные x_{p+1}, \dots, x_n .

Наличие дискретно изменяющихся параметров существенно осложняет решение задачи оптимального

проектирования. Если число дискретных параметров и принимаемых ими значений невелико, то можно для каждого набора значений дискретных параметров $x^i_1, x^i_2, \dots, x^i_p$ решать задачу поиска

$$\min_{x_{p+1}, \dots, x_n} F(x^i_1, x^i_2, \dots, x^i_p, x_{p+1}, \dots, x_n),$$

где вектор (x_{p+1}, \dots, x_n) содержится в непрерывной допустимой области G_i , соответствующей фиксированному выбранному вектору $x^i_1, x^i_2, \dots, x^i_p$.

В более сложных случаях применяются специальные методы дискретного программирования. Специфика этих методов не позволяет изложить их в рамках настоящей книги. Будут рассмотрены только некоторые эвристические преобразования, которые позволяют перейти от дискретной задачи минимизации к последовательности задач с непрерывной областью изменения параметров.

Рассмотренные выше элементы поиска оптимума являются составными частями значительного числа методов решения задачи оптимального проектирования. Эти методы ориентированы на использование ЭВМ в процессе оптимизации и учитывают следующие особенности задач оптимального проектирования:

сложность вычислений критерия оптимальности и ограничений;

многоэкстремальность, обусловленную как сложностью критерия оптимальности, так и геометрией допустимой области;

детерминированный характер критерия оптимальности и ограничений;

трудность проверки априорных математических свойств модели, таких как дифференцируемость, выпуклость, условие Липшица и т. д.;

наличие случайной погрешности в расчетах, обусловленной, с одной стороны, сложностью модели, а с другой — конечным числом разрядов, используемых при вычислениях на ЭВМ.

1-3. ПОНЯТИЕ О СИСТЕМЕ ОПТИМАЛЬНОГО ПРОЕКТИРОВАНИЯ. ОСОБЕННОСТИ ОРГАНИЗАЦИИ ПРОГРАММНОГО КОМПЛЕКСА

Системой оптимального проектирования называется совокупность методов, алгоритмов и программ, реализующих с помощью ЭВМ автоматизированный поиск на

математической модели проектируемого объекта оптимальных конструктивных параметров, при которых критерий оптимальности конструкции имеет минимально (или максимально) возможное значение при условии выполнения ограничений, накладываемых требованиями стандартов, технических условий и технологии.

Сложность математической модели проектируемых инженерно-технических объектов приводит к необходимости использования значительного числа различных методов оптимизации, эмпирических приемов ускорения сходимости, а также формирования комбинированных алгоритмов, ориентированных на изменение методов поиска в зависимости от геометрии допустимой области и скорости убывания (возрастания) критерия оптимальности. Реализация указанных процедур на ЭВМ требует создания сложного программного комплекса. Основными требованиями к такому комплексу являются:

гибкая система ввода-вывода информации, позволяющая вносить оперативные изменения в модель проектируемого объекта, а также в исходные данные как по результатам вычислений, так и в процессе счета (система должна работать как в режиме пакетной обработки, так и в диалоговом режиме);

обеспечение эффективного использования алгоритмов оптимизации, включенных в систему, путем реализации процессов адаптации при поиске, предусматривающих своевременную смену алгоритмов в процессе решения задачи оптимального проектирования;

возможность расширения системы за счет включения в нее алгоритмов, прошедших испытания на тестовых и практических задачах;

простота управляющего языка системы, использование мнемоники и смысловых конструкций при формировании заданий с целью сделать систему допустимой для инженеров-проектировщиков, не знакомых с программированием на ЭВМ;

возможность использования программного комплекса автоматизации оптимального проектирования в системах коллективного пользования.

Система управления комплексом программ автоматизированного оптимального проектирования, таким образом, может рассматриваться как некоторая проблемно-ориентированная операционная система. Она предназначена для поиска оптимума конструктивных

параметров при произвольных ограничениях на пределы их изменения.

Структура программного комплекса, реализующего оптимальное проектирование объектов на ЭВМ, представлена на рис. 1-4. Прекращение поиска производится на основе сравнения результатов расчетов двух или нескольких последовательных итераций. Наиболее рас-

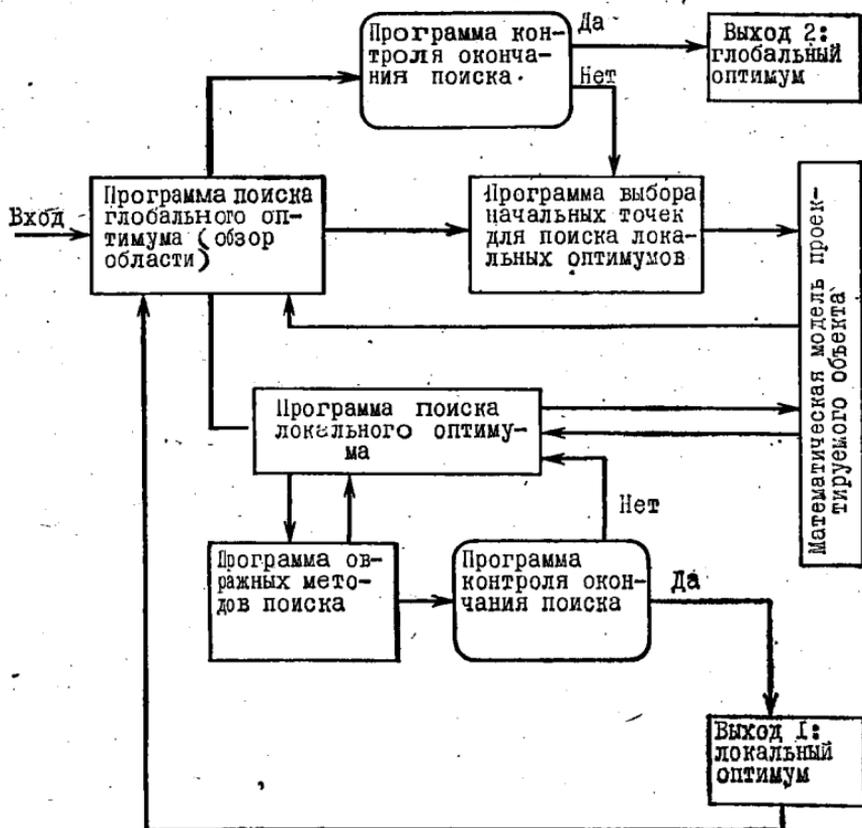


Рис. 1-4. Структура программного комплекса оптимального проектирования.

пространенными оценками являются норма разности значений двух последовательных итераций x^k и x^{k+1} и модуль разности двух последовательных значений критерия оптимальности. Так, поиск оптимума прекращается, если норма разности

$$\epsilon_k = |F(x^k) - F(x^{k+1})| \quad (1-10)$$

становится меньше заданного значения.

Такая оценка основана на предположении монотонного убывания ϵ_k в зависимости от числа итераций. Однако одна только величина ϵ_k лишь частично отражает действительное местонахождение x^k относительно локального оптимума x^* .

При медленно меняющемся критерии оптимальности, т. е. при пологом характере функции $F(x)$ вблизи оптимума, разность ϵ_k может мало меняться даже при большом шаге поиска. Поэтому наряду с условием (1-10) признаком прекращения поиска служит также норма разности двух последовательных итераций:

$$\delta_k = (x^k - x^{k+1})^T (x^k - x^{k+1}). \quad (1-11)$$

Ввиду того что при резких изменениях критерия оптимальности (в частности, в «овражной» ситуации) δ_k не может характеризовать близость очередной итерации к локальному оптимуму, как правило, используются одновременно обе оценки.

При поиске оптимума с помощью метода штрафных функций, когда критерий и ограничения заменяются обобщенным критерием оптимальности, используется оценка модуля градиента этого критерия. Поиск прекращается, если величина

$$\gamma_k = \nabla F^T \nabla F. \quad (1-12)$$

на k -й итерации становится меньше заданного числа.

Как известно, величина γ_k положительна и обращается в нуль в точке локального оптимума, поэтому по скорости ее убывания можно судить о приближении к оптимуму.

Система по минимальным управляющим указаниям со стороны пользователя (инженера-проектировщика) должна выполнять следующие операции: формировать модель объекта, выбирать для каждой конкретной задачи проектирования наиболее эффективные методы решения и осуществлять их применение в требуемой последовательности.

Для формирования рабочей модели объекта и поисковой программы система должна использовать следующие признаки, определяющие, по существу, характер и особенности решаемой задачи оптимального проектирования:

модификацию проектируемого объекта (указание используемых модулей программной модели объекта);

список ограничений;
критерий оптимальности (в случае нескольких критериев должны быть указаны отношения приоритета);
указание о дискретной или непрерывной области изменения отдельных конструктивных параметров;
требования к точности решения.

Интерактивные (диалоговые) режимы поиска оптимального решения. С развитием вычислительной техники все возрастающее значение приобретают так называемые интерактивные (диалоговые) методы проектирования. Они основаны на оперативном обмене информацией между ЭВМ и проектировщиком в процессе оптимального проектирования.

Техническим средством для такого обмена информацией служит устройство представления данных на электронно-лучевой трубке — дисплей. С помощью дисплея осуществляется формирование задания на проектирование изделия, а также вывод результатов расчетов.

В интерактивном режиме осуществляется управление алгоритмами, данными и специальными подсистемами проектирования.

Управление алгоритмами включает в себя изменение поискового алгоритма, изменение процедур вычисления некоторых характеристик (градиента, гессиана и др.), а также критерия прекращения поиска оптимума.

Управление данными предполагает изменение параметров модели после нахождения оптимального решения или в процессе поиска. Эти изменения направлены на исследование окрестности оптимального решения, устойчивости оптимального решения при небольших изменениях начальных данных, определения наличия локальных экстремумов и т. д. Кроме того, в процессе управления данными может быть изменен комплексный критерий оптимальности, в частности приоритеты отдельных его составляющих, добавлен или исключен из рассмотрения ряд ограничений, т. е. произведено изменение модели.

Управление специальными подсистемами проектирования предоставляет конструктору широкие возможности для исследования оптимального варианта изделия. Специальные подсистемы в режиме диалога могут осуществлять:

анализ окрестности оптимума с помощью представления на экране линий или поверхностей уровня кри-

терия оптимальности, а также отдельных проекций допустимой области;

выделение ограничений, существенных в точке оптимума, и варьирование их для определения возможностей улучшения определенных технико-экономических характеристик изделия за счет ослабления некоторых из ограничений;

варьирование проектировщиком отдельных проектируемых параметров с целью анализа выбранного технического решения;

наглядное представление результатов расчетного проектирования (поиска оптимума) в виде таблиц или чертежей.

1-4. АНАЛИЗ ЭФФЕКТИВНОСТИ АЛГОРИТМОВ ПОИСКА ОПТИМУМА

Функционирование системы автоматизированного оптимального проектирования предполагает наличие априорной информации об эффективности алгоритмов поиска и их предпочтительной области (стадии) применения. Такая информация накапливается на основе статистики, полученной на математических моделях проектируемых изделий либо с помощью тестовых исследований.

Эффективность алгоритма отражает возможность достижения с его помощью конечной цели при поиске оптимального решения. Так, целью поиска локального экстремума является определение оптимума с заданной точностью из произвольных начальных точек, причем за достаточно короткий промежуток времени. Целью поиска глобального оптимума является полный обзор допустимой области, позволяющий сделать вывод о том, что глобальный оптимум содержится среди уже найденных локальных оптимумов.

Алгоритмы поиска локальных оптимумов различаются по объему вычислений на каждой итерации, числу обращений к модели, точности определения оптимума, надежности алгоритма и т. д.

Использование ЭВМ при оптимизации вносит определенные особенности в представление об эффективности методов. Так, например, ввиду ограниченности разрядной сетки ЭВМ все вычисления производятся с некоторой погрешностью. Наличие этой погрешности (особенно при сложной модели) приводит к тому, что

вычисления на каждой итерации выполняются с ошибкой. Во многих случаях приближенный характер вычислений существенно снижает скорость сходимости методов, что делает удачные в теоретическом плане методы непригодными для практических расчетов.

Следует отметить также существенное влияние размерности пространства конструктивных параметров на скорость сходимости поиска. Особенно значительно это влияние для детерминированных поисковых алгоритмов. В результате уже для 25—30 конструктивных параметров решение практических задач оптимального проектирования становится затруднительным даже при значительных затратах времени счета на ЭВМ.

В соответствии с изложенными особенностями оптимизационных расчетов на ЭВМ приняты два типа характеристик алгоритмов поиска локального оптимума: локальные, в которых оценка скорости движения к оптимуму осуществляется на каждой итерации, и общие (или глобальные), сопоставляющие каждому алгоритму некоторую численную оценку.

Локальные характеристики алгоритмов поиска вычисляются на каждой итерации. Они используются для определения момента окончания поиска или при переходе к другому (возможно более эффективному) алгоритму поиска. Кроме того, локальные оценки являются во многих случаях составной частью общих характеристик поиска.

Для вычисления локальных оценок используются уже известные характеристики ε , δ и γ [см. формулы (1-10) — (1-12)].

Так, если хотя бы одна из величин ε_k или δ_k не приближается к нулю с ростом k , то это может свидетельствовать о недостаточной эффективности выбранного метода оптимизации на данном этапе поиска (например, для нескольких последних итераций).

Для контроля эффективности поиска на каждом шаге вычисляются отношения $\varepsilon_{k+1}/\varepsilon_k$ и δ_{k+1}/δ_k , которые должны быть меньше единицы. Если эти отношения приближаются к единице, метод поиска следует модифицировать или заменить более эффективным.

Для исследования алгоритмов поиска в условиях возрастающего объема вычислений по мере приближения к оптимуму используется логарифмическая оценка

$$\lambda_k = \ln[F(x^k) - F(x^{k+1})]. \quad (1-13)$$

Ее отношение к числу одномерных поисков может служить эффективным критерием сравнения алгоритмов. По этой же характеристике с помощью тестовых примеров удастся установить необходимое возрастание требований к точности одномерных (однопараметрических) алгоритмов поиска по мере приближения к оптимуму.

Некоторые локальные методы оценки эффективности формулируются в терминах алгоритмов определенного класса и служат для сравнения алгоритмов внутри этого класса. Примером такой характеристики может служить показатель потерь на оценку (см. § 5-2), представляющий собой отклонение выбранного направления поиска от градиента как функцию числа пробных вычислений. Она используется для сравнения градиента с его оценками, полученными стохастическими методами.

Общие оценки эффективности алгоритмов поиска отражают такие важнейшие их свойства, как:

точность поиска, т. е. значение окрестности локального оптимума, в которую приводит алгоритм оптимизации после заданного числа итераций;

скорость сходимости, т. е. число итераций, необходимое для достижения заданной точности;

время счета, т. е. время поиска на ЭВМ оптимального значения с заданной точностью, отнесенное к некоторой численной величине (например, к быстродействию ЭВМ, коэффициенту сложности задачи и т. п.), позволяющей осуществить сравнение алгоритмов для достаточно широкого круга задач и типов ЭВМ;

стабильность (или устойчивость), которая характеризуется незначительным увеличением числа итераций при малых возмущениях, возникающих при выборе начальной точки, а также вследствие погрешности вычислений;

надежность, т. е. свойство алгоритма, формировать в процессе работы такую последовательность итераций, которая приводит к оптимуму при многократном повторении поиска из различных начальных точек.

Для сравнения алгоритмов по этим критериям следует производить расчеты в одинаковых или близких условиях. При этом следует учитывать целый ряд обстоятельств, которые могут повлиять на процесс счета: сложность модели проектируемого объекта, выбор начального приближения и т. п. В частности, для того чтобы оценки эффективности не зависели от выбора на-

чальной точки поиска, ее выбирают далеко от оптимума либо поиск производят из нескольких точек, равномерно распределенных в области поиска, и выбирают среднее значение характеристики.

Рассмотрим подробно оценки эффективности алгоритмов поиска по каждой группе.

Оценка алгоритмов по *точности поиска* производится путем вычисления уже известных локальных характеристик ϵ , δ или γ [см. (1-10)—(1-12)] после выполнения заданного числа итераций N . Характеристики $\epsilon(N)$ и $\delta(N)$ применяются для анализа практически всех алгоритмов локального поиска. Они отражают степень приближения к оптимуму как по координатам вектора конструктивных параметров, так и по критерию оптимальности. Характеристика γ используется лишь в специальных случаях: учет ограничений в неявной форме (штрафные функции), поиск с помощью градиентных методов, методов переменной метрики и т. д.

Ввиду того, что многие алгоритмы включают в себя на каждой итерации как выбор направления поиска, так и одномерный поиск по выбранному направлению, в качестве оценки скорости сходимости часто используется такая характеристика, как число одномерных поисков, необходимых для достижения оптимума с заданной точностью. Эта характеристика позволяет исследовать эффективность выбора направления при многопараметрической оптимизации. Она определяет, насколько часто производится смена направления поиска. Если в процессе применения сложного алгоритма имеет место частая смена направлений, то это приводит к существенному снижению эффективности поиска не только за счет большого числа обращений к модели, а также и за счет многократного обращения к алгоритму. Естественно, такие алгоритмы не следует включать в систему.

Оценка алгоритма по *скорости сходимости* в некотором смысле обратна оценке по точности. При ее вычислении фиксируется требуемая точность ϵ или δ и производится сравнение алгоритмов по числу пробных шагов N , необходимых для достижения этой точности. Эта характеристика не зависит от быстродействия ЭВМ (а лишь от длины разрядной сетки), сложности модели проектируемого объекта и поэтому является показателем эффективности алгоритма. Каждое пробное вычисление используется либо для выбора направления по-

иска, либо для одномерного поиска вдоль этого направления. Поэтому скорость сходимости зависит от точности на каждой из двух стадий поиска. Задание высокой точности одномерного поиска на каждой итерации может существенно увеличить общее число обращений к модели (пробных вычислений) даже в случае весьма совершенного способа выбора направления. Необоснованное снижение требований к точности одномерного поиска может нарушить теоретические предположения, положенные в основу выбора направления (например, условие ортогональности в методе сопряженных направлений).

В качестве компромиссных вариантов могут быть использованы критерии с заданным числом проб при одномерном поиске. Число проб может быть фиксированным на все время поиска или изменяться (возрастать) по определенному правилу по мере увеличения числа итераций.

Сравнение алгоритмов *по времени счета* позволяет оценить стоимость вычислений при выборе направления поиска и сравнить его с общим временем счета. Для того чтобы эта оценка не зависела от быстродействия используемой ЭВМ, применяются относительные оценки времени, необходимого для решения задач с точностью до заданной погрешности ε . Пусть n — общее число задач, которые решались с помощью программы a , и n_s — число задач, для которых было получено оптимальное решение ($n_s \leq n$).

Вводится условный коэффициент

$$b_{a,p} = \begin{cases} 1, & \text{если программа } a \text{ решает задачу } p; \\ 0 & \text{— в противном случае,} \end{cases} \quad (1-14)$$

и $\hat{t}_{a,p}$ — относительное время, которое затрачено программой a на решение задачи p . Величина $\hat{t}_{a,p}$ представляет собой отношение $\hat{t}_{a,p}$ — реального времени, затраченного на решение задачи, к среднему времени τ_p , необходимому для решения этой же задачи с помощью ранее известных алгоритмов. Тогда формула для вычисления эффективности имеет вид:

$$\hat{i}_a = \frac{1}{n_s} \sum_{p=1}^n b_{a,p} \hat{t}_{a,p} = \frac{1}{n_s} \sum_{p=1}^n \frac{b_{a,p} t_{a,p}}{\tau_p}. \quad (1-15)$$

Трудность решенных задач может оцениваться с помощью суммарных затрат машинного времени на решение всех n задач из заданного набора:

$$T_a = \sum_{p=1}^n t_{a,p} \quad (1-16)$$

причем величина $t_{a,p}$ принимается равной удвоенному времени решения задачи p самой медленной программой, решившей ее, если программа a не решила эту задачу. Такое добавление позволяет сравнивать между собой даже те программы, которые дают оптимальное решение лишь для части задач из выделенного набора.

Оценка алгоритмов с помощью (1-15) и (1-16) и последующее сравнение результатов с числом обращений к модели показали, что число обращений к модели не дает объективную оценку эффективности алгоритмов, в которых время обработки информации о модели проектируемого объекта сравнимо или превосходит время расчета модели объекта в данной точке. Именно такие алгоритмы получили широкое развитие в настоящее время, поскольку ввиду возрастающей сложности математических моделей (и времени обращения к ним) необходимо наиболее полно использовать накопленную в ходе поиска информацию. Это, естественно, увеличивает время обработки.

Довольно сложной проверяемой характеристикой является *стабильность* алгоритмов или их *устойчивость* по отношению к небольшим изменениям траектории поиска. Ценность стабильных алгоритмов при решении на ЭВМ задач оптимального проектирования очевидна: небольшие погрешности, возникающие вследствие ограниченной разрядной сетки ЭВМ, неточностей одномерного поиска и т. д., всегда присутствуют при расчетах. Поэтому алгоритм должен быть построен таким образом, чтобы такие погрешности приводили лишь к незначительному увеличению числа итераций. Стабильными являются прямые методы и алгоритмы случайного поиска.

С повышением скорости сходимости алгоритмов их стабильность обычно снижается (т. е. возрастают требования к точности вычислений на каждой итерации). При этом, даже в одном классе методов, многие нестабильные алгоритмы существенно ускоряют сходимость к оптимальному решению в тех случаях, когда их

удаётся применить. Так, например, в рассматриваемой в гл. 4 группе алгоритмов переменной метрики метод Барнса—Розена менее стабилен, чем *DFP*-метод, однако позволяет приблизиться к оптимуму с помощью меньшего числа итераций.

Для оценки, насколько часто оказывается полезным применение того или иного алгоритма, используется такая его характеристика, как надёжность [42], т. е. вероятность получения оптимума с заданной точностью. Практическое определение этой характеристики основано на статистической обработке результатов поиска для целого ряда тестовых примеров. Основная трудность такой обработки связана с исключением индивидуальных особенностей задачи, влияющих на алгоритм: геометрии эквипотенциальных линий критерия оптимальности, выбора начальной точки поиска, длины разрядной сетки ЭВМ, организации программы и т. д.

1-5. ТЕСТОВЫЕ ЗАДАЧИ

Для проверки функционирования алгоритмов по рассмотренным критериям используются так называемые тестовые задачи. Это могут быть практические задачи с известным оптимумом или искусственным образом подобранные функции критерия оптимальности и ограничений. Они используются как для отладки программ, реализующих методы оптимизации, так и для сравнения этих программ в различных геометрических ситуациях поиска. Тестовые задачи позволяют глубже понять специфику алгоритмов оптимального проектирования, оценить их эффективность, влияние погрешностей вычислений на результат и т. д.

Известно большое число работ, посвященных созданию эффективных тестовых задач. Эта проблема оказывается тесно связанной с выбором критерия эффективности алгоритма поиска экстремума. Однако до сих пор отсутствует сколько-нибудь удовлетворительное ее решение.

Преимуществом использования тестовых задач для апробирования алгоритмов, является тот факт, что потери на каждое обращение к модели минимальны. Существенным является лишь время работы самого поискового алгоритма.

В тестовых задачах моделируются ситуации, соответствующие особенностям функции критерия оптимальности и ограничений, создающим характерные затруднения при поиске оптимального решения. Такими трудностями являются:

- двумерные или многомерные «овраги»;
- маломеняющийся критерий оптимальности («долина») при значительных изменениях проектируемых параметров;
- сложная геометрия ограничений (многосвязные допустимые области, наличие оптимума в «клювообразной» области и т. п.);
- многоэкстремальные задачи.

Наибольшие трудности при поиске локального оптимума доставляют так называемые «овражные» ситуации. Тестовой задачей,

имитирующей «овраг», является так называемая функция Розенброка [78]

$$R(x_1, x_2) = \alpha(x_2 - x_1^2)^2 + (1 - x_1)^2. \quad (1-17)$$

График этой функции для $\alpha = 100$ приведен на рис. 1-5. Линии уровня функции Розенброка сильно вытянуты в одном направлении, а в ортогональном к нему расположены довольно тесно.

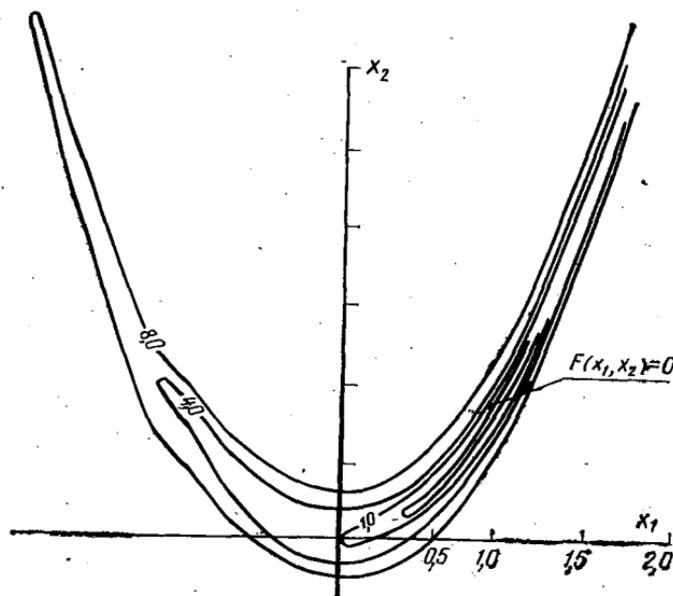


Рис. 1-5. График функции Розенброка $F(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$.

В общем случае, когда критерий оптимальности представлен с помощью функции, зависящей от n переменных, под «оврагом» понимается ситуация, когда в некоторой области все переменные можно разбить на две группы: при изменениях одних функция меняется резко, а при изменении других — незначительно. Форма «оврага» может быть различной. Так, например, функция

$$F(x_1, x_2, x_3) = 100[(x_3 - 100)^2 + (r - 1)^2] + x_2^2, \quad (1-18)$$

где $x_1 = r \cos 2\theta$, $x_2 = r \sin 2\theta$, представляет собой «овраг», линиями постоянного уровня которого являются спирали.

Сложность геометрии линий уровня функций (1-17) и (1-18) позволяет оценить с их помощью точность и быстродействие различных методов поиска и, следовательно, возможность их применения для решения практических задач оптимального проектирования. Для поиска оптимума при наличии «оврагов» (так называемая «овражная» ситуация) многие методы оказываются неэффективными и часто требуется разработка специальных, более сложных алгоритмов.

Другой часто встречающейся трудностью является пологий характер изменения критерия оптимальности, когда из-за ошибок

округления при счете на ЭВМ не удается установить, улучшаются ли значения критерия. Пологие функции критерия оптимальности встречаются, например, при проектировании электрических машин [27]. Для решения таких задач требуется точное вычисление значений критерия, его градиента, а также функций ограничений.

Тестовые функции, реализующие пологий критерий оптимальности, называют «долинами». Примером такой функции может служить выражение

$$F(x) = \int_0^1 T(x) dx - \frac{1}{n} \sum_{j=1}^n T(x_j)$$

для произвольной функции $T(x)$ [18]. Выражение $F(x)$, рассматриваемое как функция переменных x_1, x_2, \dots, x_n , представляет собой тестовую задачу размерности n .

Эффективность алгоритмов, предназначенных для решения общей задачи оптимального проектирования, может быть проверена на тестовых задачах, содержащих ограничения. Эти задачи должны включать в себя нелинейные зависимости как в критерии оптимальности, так и в функциях ограничений. Приведем два тестовых примера, используемых для сравнения различных модификаций алгоритмов, учитывающих ограничения в явной форме [69].

Пример 1-1. Найти минимум критерия оптимальности

$$F(x) = x_1 x_2 x_3 x_4 x_5$$

при ограничениях:

$$x^2_1 + x^2_2 + x^2_3 + x^2_4 + x^2_5 - 10 = 0;$$

$$x_2 - x_3 - 5x_4 x_5 = 0;$$

$$x^2_1 + x^2_2 + 1 = 0.$$

Эта функция имеет минимальное значение $F^* = 2,9197$ в точке $x^* = (-1,1712; 1,5957; 1,8272; -0,7636; -0,7636)$. Начальной точкой поиска обычно служит $x^0 = (-2; 1,5; 2; -1; -1)$.

Пример 1-2. Найти минимум критерия оптимальности

$$F(x) = (x_1 - 10)^2 + 5(x_2 - 12)^2 + x^4_3 + 3(x_4 - 11)^2 + \\ + 10x^6_5 + 7x^2_6 + x^4_7 - 4x_6 x_7 - 10x_8 - 8x_7$$

при Ограничениях:

$$-2x^2_1 - 3x^2_2 - x_3 - 4x^2_4 - 5x_5 + 127 \geq 0;$$

$$-7x_1 - 3x_2 - 10x^2_3 - x_4 + x_5 + 282 \geq 0;$$

$$-23x_1 - x^2_2 - 6x^2_6 + 8x_7 + 196 \geq 0;$$

$$-4x_1 - x^2_2 + 3x_1 x_2 - 2x^2_3 - 5x_8 + 11x_7 \geq 0.$$

Эта задача имеет оптимальное значение $F^* = 680,97$ в точке $x^* = (2,30; 1,95; -0,47; 4,37; 0,51; 1,03; 1,68)$. Используемая начальная точка поиска $x^0 = (1; 2; 0,4; 0; 1; 1)$.

Тестирование алгоритмов решения многоэкстремальных задач можно осуществлять на моделях, представляющих собой суперпозицию (чаще всего линейную комбинацию) одноэкстремальных функций, квадратичных в окрестности оптимума [6]. Однако эффектив-

ность методов поиска глобального оптимума существенно определяется спецификой решаемой задачи оптимального проектирования. Поэтому формирование тестовых задач для этих алгоритмов обычно состоит в упрощении математических моделей реальных задач с сохранением их инженерной интерпретации.

1-6. ПРИМЕР. СИСТЕМА АВТОМАТИЗАЦИИ ОПТИМАЛЬНОГО ПРОЕКТИРОВАНИЯ АСИНХРОННЫХ ДВИГАТЕЛЕЙ

Проектирование общепромышленных асинхронных двигателей требует проведения значительного объема электромагнитных, тепло-вентиляционных, виброакустических и экономических расчетов. Выбор оптимального варианта асинхронного двигателя осуществляется путем оптимального расчетного проектирования на экономико-технической математической модели двигателя [26, 13]. Оптимальное проектирование является центральной частью системы автоматизации проектирования асинхронных двигателей.

Целью создания системы автоматизированного проектирования асинхронных двигателей (САПР АД) является повышение качества асинхронных двигателей, сокращение сроков разработки и освоения их в производстве. Система автоматизированного проектирования АД включает комплекс технических и программных средств, информационную базу, а также организационные положения, определяющие функции и взаимодействие расчетных, конструкторских и технологических подразделений, осуществляющих проектирование.

Общая постановка задачи оптимального проектирования асинхронных двигателей предполагает одновременный поиск оптимальных параметров (в частности, размеров) как активной, так и конструктивной частей машины. Однако при этом задача настолько усложняется, что в настоящее время является едва ли практически разрешимой. С другой стороны, существующий опыт проектирования показывает, что раздельный поиск оптимальных размеров активной электромагнитной части двигателя с последующим определением размеров других конструктивных элементов может обеспечить получение результатов проектирования, по своему качеству вполне удовлетворяющих требованиям практики.

Содержание и последовательность процедур расчета вариантов двигателя определяются его математической моделью. Для активной части электрической машины математической моделью является методика (совокупность формул) электромагнитного и тепло-вентиляционного расчетов. Следует заметить, что при рассмотрении более широкого круга показателей машины в состав математической модели могут быть включены также и методики других расчетов, например виброакустических характеристик, показателей надежности и др.

В качестве критерия оптимальности в САПР АД приняты суммарные затраты на производство и эксплуатацию двигателей, учитывающие стоимость потребляемой двигателем электроэнергии за определенный период эксплуатации. Цель оптимального проектирования — минимизация этого критерия.

Число независимых конструктивных параметров, варьируемых в процессе поиска оптимального варианта конструкции асинхронного двигателя, обычно не превышает восьми — десяти. К ним относятся внешний и внутренний диаметры сердечника статора, длина активной части двигателя, размеры пазов статора и ротора, число витков обмотки статора и др. Значения остальных размеров и других пара-

метров двигателя однозначно определяются по основным конструктивным параметрам.

В качестве функций ограничений $R_i(x)$ при проектировании принимаются:

а) зависимости, обеспечивающие соблюдение требований по пусковым характеристикам двигателя: минимально допустимые пусковой и максимальный моменты, максимально допустимый пусковой ток;

б) требования, связанные с надежностью и сроком службы двигателя: предельно допустимое превышение температуры обмотки, а также скорость нарастания температуры при заторможенном роторе;

в) конструкторско-технологические ограничения, направленные на обеспечение прочности, жесткости и технологичности изготовления деталей и узлов. В частности, ограничения максимального значения коэффициента заполнения паза, минимальных значений высоты спинки сердечника ротора, ширины зубцов и т. д.

В результате задача оптимального проектирования асинхронных двигателей формулируется как общая задача математического программирования с нелинейными критерием оптимальности и функциями ограничений.

Универсальная математическая (программная) модель асинхронного двигателя охватывает большое число его конструктивных модификаций, отличающихся формой пазов статора (открытые, полуоткрытые, полужакрытые, закрытые), типом ротора (одинарная или двойная клетка) и другими особенностями.

Универсальная программная модель двигателя построена по модульному принципу. Отдельные модули соответствуют фрагментам расчетной методики. Такая универсальная модель может быть представлена в виде графа, вершинами которого являются модули, а дугами — возможные переходы от одного модуля к другому. Задание конструктивной модификации электрической машины определяет связь между модулями, т. е. активизирует на графе некоторый путь.

Соответствующее программное обеспечение предусматривает возможность модификации проектируемой машины заданием некоторого вектора b с координатами, принимающими значение 0 или 1. Каждому модулю соответствует определенная координата вектора b , равная единице только в том случае, когда этот модуль должен быть включен в модель. Вектор b задается перед каждым расчетом, и по нему генерируется программная модель для заданной модификации проектируемой машины.

Для выбранных критерия оптимальности и ограничений осуществляется поиск оптимальных конструктивных параметров. С помощью стохастических методов производится обзор допустимой области, позволяющий выбрать начальные точки для применения методов поиска локальных оптимумов. Уточнение расположения локальных оптимумов производится с помощью метода линейного локального моделирования (см. § 6-2). Адаптация при поиске оптимума заключается в сокращении времени поиска за счет использования при последующем шаге информации, полученной на предыдущих шагах.

Решение задачи оптимизации (поиск глобального оптимума) конструктивных параметров асинхронных двигателей на ЭВМ средней производительности занимает 0,5—1 ч.

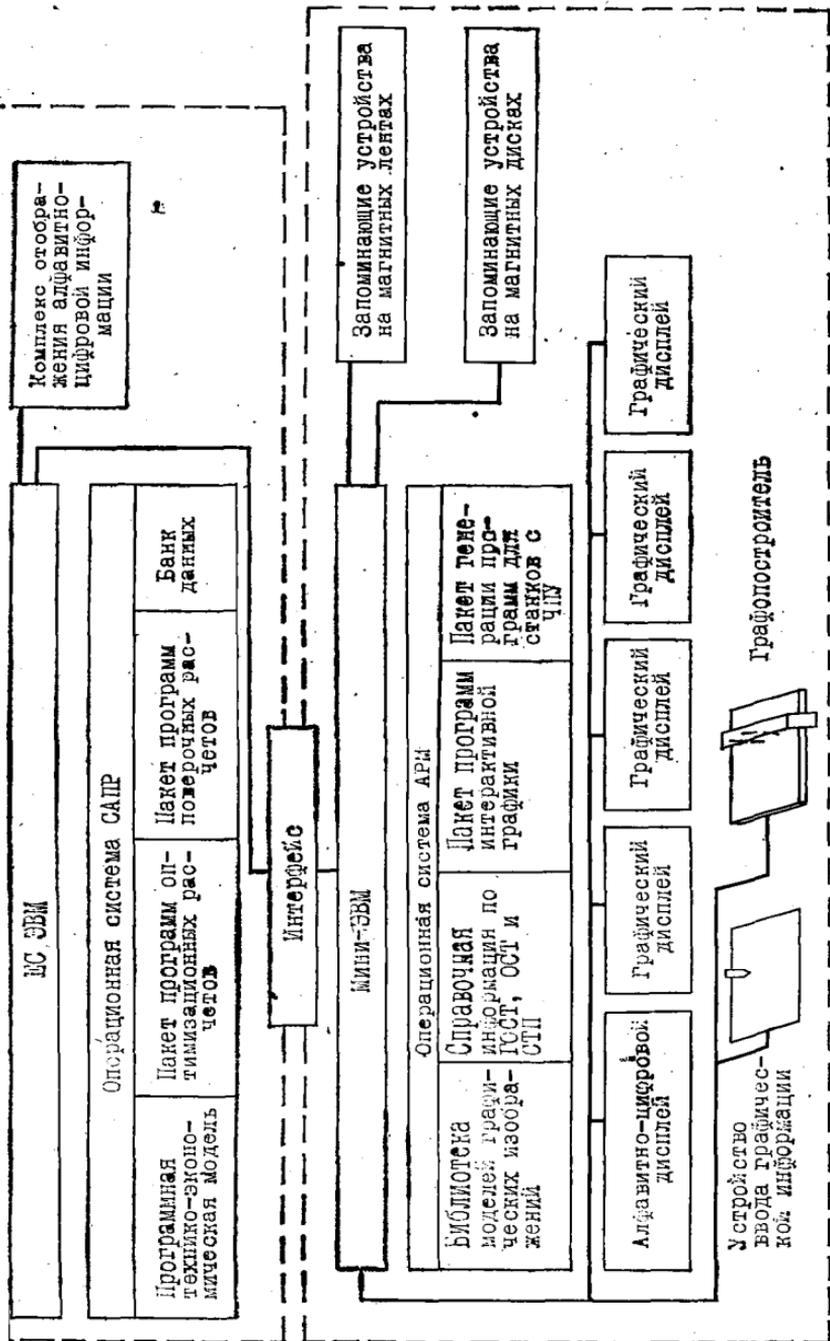


Рис. 1-6. Система автоматизированного проектирования асинхронных двигателей.

Система САПР АД имеет двухуровневую структуру вычислительных средств (рис. 1-6). Верхний уровень образован ЭВМ Единой системы, на которой выполняется поиск оптимальных параметров проектируемого объекта. Нижний уровень состоит из комплекса аппаратуры и программных средств автоматизированного рабочего места конструктора (АРМ), включающего мини-ЭВМ, чертежную машину и дисплейные пульты. Комплекс АРМ связан с ЭВМ верхнего уровня через соответствующее сопряжение, позволяющее производить обмен информацией между ЭВМ обоих уровней.

Программное обеспечение собственно оптимального проектирования включает: операционную систему САПР, программную модель асинхронного двигателя, пакет программ алгоритмов поиска оптимума, банк данных. Программное обеспечение нижнего уровня содержит: библиотеку моделей графических изображений, справочную информацию по ГОСТ, ОСТ и СТП, пакет программ интерактивной графики, пакет генерации программ для станков с ЧПУ.

Работа проектировщика в диалоговом режиме позволяет быстро производить расчеты, направленные на доводку рассчитанных на ЭВМ оптимальных вариантов машин. В частности, при проектировании серий двигателей доводочная работа заключается в пересчетах двигателя на стандартные размеры обмоточных проводов и в унификации длин сердечников.

В режиме диалога инженер-расчетчик может оперативно производить следующие операции: коммутацию блоков математической модели, необходимых для того или иного вида расчетов, а также изменения перечня ограничений и критериев оптимальности. Эти операции выполняются при помощи директив, набираемых на клавиатуре дисплея.

Интерактивный режим взаимодействия конструктора с ЭВМ позволяет осуществить широкие исследовательские работы для решения таких принципиальных вопросов проектирования серий, как выбор целесообразного класса нагревостойкости изоляции, рациональных чисел и конфигурации пазов, типа обмотки и т. д.

Глава вторая

ПОИСК ГЛОБАЛЬНОГО ОПТИМУМА

2-1. ПРОБЛЕМА ПОИСКА ГЛОБАЛЬНОГО ОПТИМУМА

Поиск оптимальных конструктивных параметров осуществляется в условиях отсутствия информации о числе локальных экстремумов функции критерия оптимальности, ее свойствах гладкости, сложности геометрии допустимой области и т. д. Поэтому задача оптимального проектирования является многоэкстремальной задачей, которая состоит в определении глобального оптимума по информации о математической модели проектируемого объекта, накапливаемой в процессе поиска.

На рис. 2-1 приводится пример геометрической интерпретации многоэкстремальной задачи оптимального

проектирования. На рисунке видны два локальных оптимума. Локальные оптимумы возникают, как правило, на границе допустимой области, однако возможно их наличие и во внутренних точках.

Основная трудность разработки алгоритмов поиска глобального оптимума объясняется тем, что с помощью ЭВМ удается собрать лишь локальную информацию

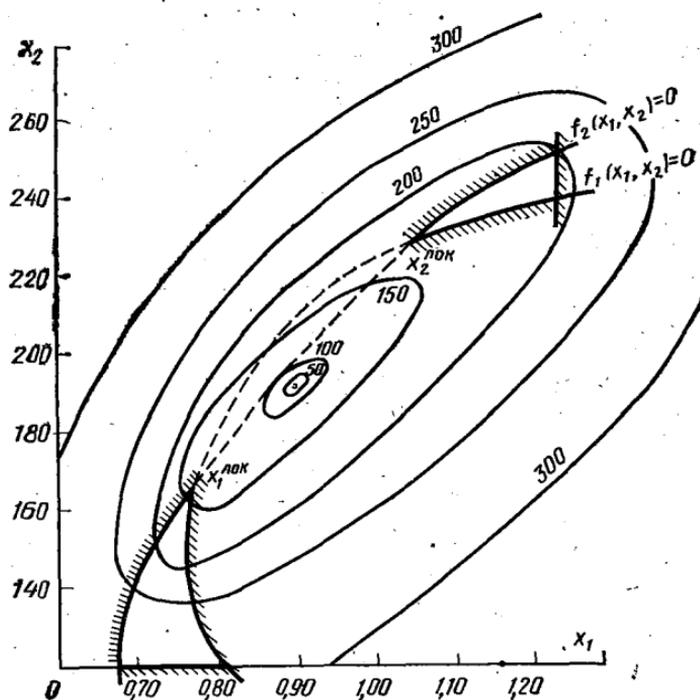


Рис. 2-1. Многоэкстремальная задача оптимального проектирования.

о допустимой области и изменении в ней критерия оптимальности. В то же время для эффективного поиска глобального оптимума необходимы данные о характере изменения критерия оптимальности во всей допустимой области. Поэтому в основе всех алгоритмов поиска глобального оптимума лежит тот или иной способ выбора конечного набора точек n -мерного пространства конструктивных параметров, по которым можно приближенно оценить расположение всех (или наиболее существенных) локальных оптимумов. Эти точки (выбран-

ные значения параметров) служат начальным приближением для применения методов поиска локального оптимума.

Важным этапом при поиске глобального оптимума является формирование признака окончания поиска. Выполнение этого признака должно с определенной точностью или вероятностью указывать на то, что глобальный оптимум находится среди уже найденных локальных оптимумов.

2-2. МЕТОДЫ ПОСЛЕДОВАТЕЛЬНОСТИ НЕЗАВИСИМЫХ ИСПЫТАНИЙ

Полный перебор на равномерной сетке. Если независимых переменных — конструктивных параметров — немного (не более 3—4) и невелики пределы, в которых они могут изменяться, то поиск на ЭВМ наименьшего (или наибольшего) значения функции $F(x)$ может быть осуществлен обходом в определенном порядке узлов многомерной сетки в пространстве независимых переменных и вычислением в каждой точке значений функций ограничений и критерия оптимальности. При этом ЭВМ отбрасывает те точки, в которых условия-ограничения не выполняются, а среди точек, для которых все ограничения выполняются, выбирает точку с наименьшим (или наибольшим) значением функции $F(x)$.

Возможность использования такого метода поиска основывается на том, что в технических задачах экстремумы обычно пологие.

Алгоритм осуществляет перебор комбинаций значений конструктивных параметров x_1, x_2, \dots, x_n , изменяющихся в заданных пределах:

$$a_j \leq x_j \leq b_j, \quad j=1, 2, \dots, n \quad (2-1)$$

с фиксированным шагом

$$h_j = \frac{1}{N_j} (b_j - a_j), \quad (2-2)$$

где N_j — число различных значений параметров x_j , используемых при поиске. После перебора всех векторов конструктор получает набор значений оптимальных конструктивных параметров с погрешностью, не превышающей $h_j, j=1, 2, \dots, n$, и соответствующее ему значение критерия оптимальности.

Рассмотренный алгоритм весьма прост для программирования, не предполагает проведения поисков локальных экстремумов, но даже при небольшом числе независимых конструктивных параметров может потребовать колоссальных затрат времени счета на ЭВМ.

Наиболее простым приемом, позволяющим ускорить расчеты, является расположение ограничений в порядке постепенного уменьшения вероятности их нарушения. Вначале проверяются те ограничения, которые нарушаются наиболее часто. Тогда уже при проверке небольшого числа ограничений удается исключить из рассмотрения большинство недопустимых точек.

Другой прием состоит в том, что исследование допустимой области производится предварительно с помощью равномерной сетки с достаточно большим шагом. Увеличение длины шага сетки приводит к уменьшению числа точек, в которых производится проверка ограничений и вычисление значений критерия оптимальности. Далее, в окрестности точки с минимальным значением критерия оптимальности производится разбиение области сеткой с меньшим шагом (шаг сетки последовательно уменьшается вдвое, вчетверо и т. д.). Такой прием позволяет значительно сократить общее время вычислений, однако неправильный выбор начального шага сетки может привести к тому, что в действительности локальный экстремум будет принят за глобальный. На практике шаг сетки определяется исходя из опыта решения близких по типу задач (т. е. при оптимальном проектировании целого ряда однотипных конструкций).

Независимые статистические испытания. В отличие от детерминированного перебора значений изменение параметров при статистическом поиске задается с помощью некоторого закона распределения случайных чисел. Поскольку в начальный момент поиска отсутствует информация о возможном расположении глобального оптимума, обычно принимается равномерный закон распределения пробных точек внутри допустимой области.

Как и в предыдущем параграфе, предполагается, что конструктивные параметры ограничены условиями (2-1).

Для нахождения текущего вектора x значений параметров производится генерирование набора из n случайных чисел $\xi_1, \xi_2, \dots, \xi_n$, имеющих равномерное распределение на отрезке $[0, 1]$ и имитирующих последовательность независимых испытаний. В этом случае говорят

о независимом статистическом поиске. Каждая координата вектора x вычисляется по формуле

$$x_j = a_j + \xi_j (b_j - a_j).$$

В точке x вычисляются значения ограничений. При выполнении ограничений вычисляется значение критерия оптимальности. В ходе вычислений так же, как и при полном переборе, отбрасываются точки, оказавшиеся вне области допустимых значений переменных, а среди допустимых точек выбирается и запоминается точка с наименьшим значением критерия оптимальности $F(x)$.

Полученный результат характеризуется вероятностью p того, что при данном числе N случайных проб расположение точки оптимума будет определено с точностью до Δ , где Δ — объем n -мерного параллелепипеда, выраженный в долях от общего объема области поиска, представляющей собой n -мерный параллелепипед, заданный неравенствами (2-1).

В этом случае Δ — вероятность попадания для каждой пробы в заданную область, имеющую объем Δ . Если делается N случайных проб, то вероятность того, что хотя бы одна из этих проб попадает в заданную область Δ (окрестность глобального экстремума), может быть определена по формуле

$$p = 1 - (1 - \Delta)^N, \quad (2-3)$$

так как $1 - \Delta$ — вероятность непопадания в заданную область для одной пробы; $(1 - \Delta)^N$ — вероятность того же события при N пробах.

Из (2-3) легко получается выражение для определения числа случайных проб N , необходимых для того, чтобы с заданной вероятностью p расположение оптимума определялось с точностью Δ :

$$N = \frac{\log(1 - p)}{\log(1 - \Delta)}. \quad (2-4)$$

Эта формула не зависит от размерности пространства конструктивных параметров. Указанное свойство метода независимых статистических испытаний рассматривалось в ряде работ как преимущество, характерное для случайных методов. Оно объясняется тем, что требования к точности определения оптимума выражены через объемную характеристику Δ .

Если точность поиска задана по каждому из параметров отдельно, то ситуация в корне меняется. Пусть относительная точность по каждой координате составляет $\alpha_i / (b_i - a_i)$, $i=1, 2, \dots, n$. Тогда выражение для объема n -мерного параллелепипеда имеет вид:

$$\Delta = \frac{\alpha_1 \alpha_2 \dots \alpha_n}{(b_1 - a_1)(b_2 - a_2) \dots (b_n - a_n)}$$

Отсюда, в частности, следует, что при фиксированной объемной характеристике точности Δ линейная точность (по каждому параметру) снижается с ростом размерности n .

При малом Δ согласно (2-4) число испытаний составляет:

$$N = - \frac{\log(1-p)}{\Delta},$$

а при равной точности по каждому из параметров x_i

$$N = - \frac{(b-a)^n}{\alpha^n} \log(1-p).$$

Следовательно, исходя из заданной точности, необходимое число пробных точек возрастает при увеличении n со скоростью показательной функции.

В связи с этим пробные точки, выбранные в результате статистических испытаний, рассматриваются только как начальные приближения для применения одного из методов поиска локального экстремума.

Из случайно выбранной точки x^0 как из начального приближения производится поиск локального экстремума. Полученная точка обозначается x^{0*} . Затем снова производится случайный выбор (т. е. выбор с помощью метода независимых статистических испытаний) новой начальной точки x^1 и поиск из нее локального экстремума x^{1*} и т. д. После того, как некоторое число локальных экстремумов $x^{0*}, x^{1*}, \dots, x^{s*}$ определено, локальные точки из новых начальных точек с большой вероятностью приводят к окрестности ранее найденных локальных экстремумов. Поскольку для большинства задач оптимального проектирования поиск локального экстремума представляет весьма сложную итеративную процедуру, рассмотренный метод требует значительного времени счета.

Дальнейшее сокращение времени поиска достигается за счет того, что вместо поиска из текущей точки локального экстремума с заданной точностью производится лишь несколько первых итераций из некоторого набора начальных точек. Те последовательности, для которых достигнуто наибольшее улучшение значений критерия оптимальности, запоминаются. Затем дополнительно генерируется несколько новых начальных точек, совершается несколько итераций из этих точек, а также оставленных в результате отбора и осуществляется новый отбор и т. д. Рассмотренный способ отбора наиболее «перспективных» для дальнейшего поиска точек положен в основу целого ряда эффективных методов оптимального проектирования [7, 41]. Оценки скорости сходимости таких методов и достоверности найденного глобального экстремума получают эмпирически в результате накопления опыта расчетного проектирования.

Для того чтобы полностью определить метод поиска, следует указать критерий прекращения вычислений. Если в течение N испытаний (т. е. при генерировании N случайных начальных точек и выполнении из них нескольких шагов поиска) не выявляются новые локальные экстремумы, то частота прихода к уже найденным экстремумам $\nu=1$.

Обозначим через p вероятность того, что случайно выбранная начальная точка приводит к уже известному локальному экстремуму. Вероятность того, что разность

$$|\nu - p| \leq g \sqrt{\frac{p(1-p)}{N}},$$

согласно [54] может быть записана в виде

$$P = P \left\{ |\nu - p| \leq g \sqrt{\frac{p(1-p)}{N}} \right\} = \frac{2}{\sqrt{2\pi}} \int_0^g e^{-\frac{x^2}{2}} dx = 2\beta(g), \quad (2-5)$$

где $\beta(g)$ можно получить из таблиц значений интеграла вероятностей. Границы изменения p называются доверительными границами. Величина P , связанная с границами по формуле (2-5), называется уровнем значимости соответствующего доверительного интервала изменения.

Из (2-5) получаем, что при уровне значимости P доверительные границы изменения p составляют:

$$1 \geq p \geq \frac{N}{N + g^2}. \quad (2-6)$$

Полагая $N / (N + g^2) = 1 - \mu$, получаем соотношение

$$N = \frac{1 - \mu}{\mu} g^2. \quad (2-7)$$

Формулы (2-6) и (2-7) позволяют оценить вероятность пропуска глобального экстремума и число проб N , необходимое для того, чтобы эта вероятность не превышала μ . Так, при $g=2$ и $N=96$ получаются значения $2\beta(g)=0,954$ и $1 \geq p \geq 0,96$. Если задаться уровнем значимости 0,95 и вероятностью пропуска глобального экстремума 0,04, то поиск новых локальных экстремумов прекращается после 96 попыток.

2-3. ИСПОЛЬЗОВАНИЕ АПРИОРНОЙ ИНФОРМАЦИИ О ФУНКЦИИ КРИТЕРИЯ ОПТИМАЛЬНОСТИ

Если известно, что функция, представляющая собой критерий оптимальности, удовлетворяет определенным условиям, то можно ускорить процедуру детерминированного перебора значений проектируемых параметров. Рассмотрим два условия: Липшица и сепарабельности.

Полный перебор на неравномерной сетке (при выполнении условия Липшица). Равномерная сетка с достаточно мелким шагом позволяет определить глобальный экстремум в допустимой области для произвольного критерия оптимальности, выраженного с помощью непрерывной функции или функции, имеющей разрывы первого рода.

Однако такой подход требует слишком большого времени счета. Некоторые априорные сведения о характере критерия оптимальности позволяют выбирать шаг сетки в зависимости от ранее вычисленных значений, т. е. изменять шаг в процессе поиска.

Говорят, что функция $F(x)$ удовлетворяет условию Липшица в области G , если существует такая постоянная величина C , что для любых двух векторов $x^1, x^2 \in G$ выполняется неравенство

$$|F(x^1) - F(x^2)| \leq C |x^1 - x^2|. \quad (2-8)$$

Это условие означает, что $F(x)$ убывает и возрастает не быстрее линейной функции с заданным коэффициентом C .

На рис. 2-2 представлена функция одной переменной, удовлетворяющая условию Липшица на отрезке $0 \leq x \leq a$ с константой C_0 . После того, как найден локальный минимум $x_1^{\text{лок}}$ функции $F(x)$, удается ускорить

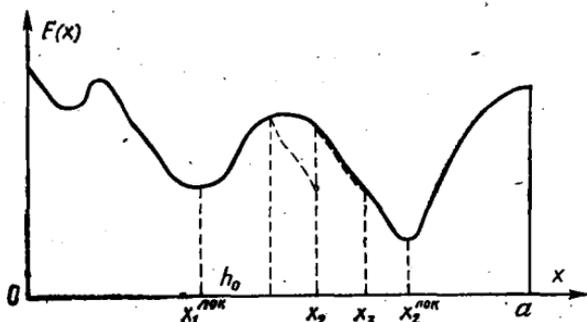


Рис. 2-2. Функция, удовлетворяющая условию Липшица.

поиск остальных локальных минимумов по сравнению с полным перебором на равномерной сетке.

Пусть h_0 — некоторый шаг, осуществленный из точки локального минимума $x_1^{\text{лок}}$. Если в точке $x_1^{\text{лок}} + h_0$ значение функции F не меньше, чем в точке $x_1^{\text{лок}}$, т. е. $F(x_1^{\text{лок}} + h_0) \geq F(x_1^{\text{лок}})$, то, исходя из условия (2-8), можно определить ближайшую точку, в которой функция $F(x)$ может достигнуть значения $F(x_1^{\text{лок}})$ (подозрительного на глобальный минимум).

Действительно, согласно (2-8) функция $F(x)$ убывает не быстрее, чем линейная функция:

$$\varphi(x) = F(x_1^{\text{лок}} + h_0) + C_0 [x - (x_1^{\text{лок}} + h_0)].$$

Отсюда следует, что точка x_2 , в которой необходимо произвести новое пробное вычисление функции $F(x)$, может быть получена из условия $\varphi(x) = F(x_1^{\text{лок}})$, т. е.

$$x_2 = (x_1^{\text{лок}} + h_0) + \frac{F(x_1^{\text{лок}} + h_0) - F(x_1^{\text{лок}})}{C_0}. \quad (2-9)$$

Поскольку $F(x_2) \geq F(x_1^{\text{лок}})$ (рис. 2-2), к ней также можно применить сформулированный выше алгоритм для определения следующей пробной точки x_3 .

В отличие от предыдущей точки x_2 в точке x_3 выполняется равенство $F(x_3) = F(x_1^{\text{лок}})$, поэтому из нее производится поиск локального минимума одним из известных методов. В точке $x_2^{\text{лок}}$ достигается значение, меньшее, чем $F(x_1^{\text{лок}})$, поэтому все дальнейшие шаги конца отрезка определяются исходя из $F(x_2^{\text{лок}})$. Ввиду того что не оказывается новых точек, подозрительных на глобальный оптимум, поиск заканчивается.

Пусть теперь $F(x)$ — критерий оптимальности, зависящий от n -мерного аргумента x , и область изменения независимых переменных определяется формулой (2-1). После того, как определены значения критерия оптимальности в ряде точек (например, при поиске локальных экстремумов), вводится величина

$$f_k = \min\{F(x^1), F(x^2), \dots, F(x^k)\}, \quad (2-10)$$

равная минимальному значению критерия $F(x)$ после k пробных вычислений. Следующий шаг представляет собой формирование области, в которой все функции φ , удовлетворяющие условию Липшица для одной и той же константы C и принимающие в точках x^1, x^2, \dots, x^n значения

$$\varphi(x^i) = F(x^i), \quad i=1, 2, \dots, k, \quad (2-11)$$

имеют значения не меньше $f_k - \varepsilon$. С этой целью в каждой точке $x^i, i=1, 2, \dots, k$ формируется шар, описываемый уравнением

$$C|x - x^i| \leq F(x^i) - F_k + \varepsilon, \quad (2-12)$$

с центром в этой точке и радиусом $R_i = \frac{1}{C} [F(x^i) - F_k + \varepsilon]$.

Если построенные таким способом k шаров полностью покрывают допустимую область, решение получено. В противном случае вычисляются пробные значения вне шаров вида (2-12) и определяются радиусы шаров с центрами в новых точках.

Один из возможных алгоритмов формирования такого покрытия (см. [24]) состоит в разбиении допустимой области на n -мерные кубы и построения содержащих их шаров. При этом связь между длиной грани i -го куба и радиусом R_i описывающего его шара задается формулой

$$x_j^i - \frac{R_i}{\sqrt{n}} \leq x_j \leq x_j^i + \frac{R_i}{\sqrt{n}}, \quad j=1, 2, \dots, n,$$

где

$x^i = (x^i_1, x^i_2, \dots, x^i_n)$ — центр шара с радиусом R_i .

Указанный алгоритм эффективен, если заранее известно не только то, что функция удовлетворяет условию Липшица, но также и значение константы C . Это значение может быть получено в процессе решения ряда однотипных задач оптимального проектирования. В каждом конкретном поиске можно, начав с некоторого $C = C_0$, решать задачу поиска глобального экстремума для C , равного $2C_0$, $4C_0$ и т. д., до тех пор, пока результат будет отличаться от предыдущего не больше чем на ϵ . Хотя этот подход основан на эмпирических соображениях, он дает достаточно точное решение при практических инженерных расчетах.

Кроме того, хорошие результаты дает сочетание методов перебора на неравномерной сетке с методами поиска локальных экстремумов. Действительно, после нахождения локального экстремума радиусы шаров (2-12) можно значительно увеличить, исключив из дальнейшего рассмотрения существенную часть области поиска.

Динамическое программирование (выполняется условие сепарабельности). Исследование допустимой области при определении оптимального решения иногда удается представить как многошаговый процесс управления. На первом шаге этого процесса решается задача выбора одного параметра, затем двух, трех и т. д. С этой целью выделяется вектор параметров ξ_k , который подлежит определению на шаге с номером k . Другой вектор параметров η характеризует состояние системы, он определяет некоторую функцию $\Lambda_k(\eta)$, называемую функцией состояния. Функция $\Lambda_k(\eta)$ определяется как минимальное значение критерия оптимальности при заданном значении вектора η . После выбора η определяется вектор $T(\xi_k, \eta)$ параметров состояния для

остальных $k-1$ шагов. Связь между шагами в методе динамического программирования задается с помощью формулы

$$\Lambda_k(\eta) = \min_k [\varphi_k(\xi_k) + \Lambda_{k-1}(T(\xi_k, \eta))]. \quad (2-13)$$

В частности, динамическое программирование может применяться в тех случаях, когда критерий оптимальности удовлетворяет свойству сепарабельности, т. е. представляет собой сумму функций одной переменной:

$$F(x) = \sum_{j=1}^n \varphi_j(x_j). \quad (2-14)$$

Это возможно, если $\varphi_j(x_j)$ — критерий оптимальности элемента проектируемого изделия, зависящий от единственного параметра x_j .

В простейшем случае ограничение в этой задаче имеет вид линейного неравенства

$$\sum_{j=1}^n a_j x_j \leq b. \quad (2-15)$$

Эту задачу можно представить в виде

$$\Lambda_k(\eta) = \min_{x_k} [\varphi_k(x_k) + \Lambda_{k-1}(\eta - a_k x_k)], \quad k=2, 3, \dots, n, \quad (2-16)$$

поскольку функция состояния $\Lambda_k(\eta)$ записывается с помощью выражения

$$\Lambda_k(\eta) = \min_{x_1, \dots, x_k} \sum_{j=1}^k \varphi_j(x_j),$$

где минимум берется по неотрицательным дискретным значениям, удовлетворяющим условию

$$\sum_{j=1}^k a_j x_j \leq \eta.$$

Функция $\Lambda_1(\eta)$ определяется непосредственно; остальные $\Lambda_k(\eta)$ — с помощью рекуррентных соотношений.

Применение динамического программирования для решения общей задачи оптимального проектирования путем введения на каждом шаге одного нового параметра не всегда оказывается эффективным. Отметим, что такой подход к поиску локального экстремума оказывается эквивалентным методу покоординатного спуска [2].

Оптимальное проектирование с помощью динамического программирования более эффективно, если на каждом шаге вводится определенная группа параметров. Разбиение на группы производится с учетом структуры задачи. В одну группу включаются параметры, тесно связанные общими физическими, техническими и технологическими особенностями. Связь между группами параметров осуществляется по небольшому числу характеристик, которые в некоторых случаях могут изменяться в небольших пределах. Поэтому говорят, что разбиение производится на относительно автономные группы параметров [76, 77].

В качестве примера рассмотрим разбиение переменных (параметров) задачи оптимального проектирования синхронного генератора с принудительным жидкостным трубчатым охлаждением статора и ротора [2].

Критерием оптимальности в этой задаче является масса генератора (с точностью до коэффициента пропорциональности)

$$M(D, b_n, l, h_n, \delta, h_a, b_m, D_B, i_k, d_T), \quad (2-17)$$

где D — диаметр расточки якоря; b_n — ширина зазора паза якоря; l — длина пакета якоря; h_n — глубина паза; δ — рабочий зазор; h_a — высота ярма якоря; b_m — ширина полюсного сердечника; D_B — внутренний диаметр пакета индуктора; i_k — число окон канала корпуса на длине якоря в продольном разрезе генератора; d_T — внутренний диаметр трубки индуктора, расположенной в междуполюсном пространстве.

Задача оптимального проектирования в данном случае состоит в поиске таких значений конструктивных параметров, которые минимизируют выражение (2-17) при определенных ограничениях на коэффициент перегрузочной способности, суммарные гидравлические потери в трубке индуктора, температурные режимы и т. д.

Полная методика электромагнитного, теплового и гидравлического расчетов состоит примерно из 400 фор-

мул. Для решения задачи проектируемые параметры разбиваются на две группы. К первой относятся параметры, позволяющие произвести расчет электромагнитных характеристик конструкции, ко второй — параметры i_k и d_T , лежащие в основе тепловых и гидравлических расчетов. Выделение указанных групп параметров приводит к двум задачам оптимального проектирования для меньшего числа параметров:

1) минимизировать критерий оптимальности при постоянных i_k , d_T ;

2) минимизировать критерий оптимальности путем изменения i_k и d_T при постоянных остальных параметрах в формуле (2-17).

Каждая из задач оптимизации решается с помощью ранее изложенных методов оптимизации, а рекурсивное представление критерия оптимизации в виде (2-13) позволяет реализовать двухшаговый процесс поиска глобального минимума на основе динамического программирования.

2-4. АДАПТАЦИЯ ПРИ ПОИСКЕ ГЛОБАЛЬНОГО ОПТИМУМА

В основе независимого стохастического поиска лежит предположение о равновероятном нахождении экстремума в любой точке допустимой области. Это предположение справедливо лишь для нескольких начальных точек, т. е. тогда, когда отсутствует информация о характере изменения критерия оптимальности. В результате накопления информации о ходе поиска (одного или нескольких) локальных экстремумов удается выделить окрестности, в которых случайные начальные точки с большей вероятностью приводят к одному и тому же локальному экстремуму. Поскольку поиск носит вероятностный характер, можно говорить о наименее и наиболее вероятных районах расположения новых локальных экстремумов. Процесс адаптации заключается в изменении вероятностей генерирования начальных точек по информации, накопленной в процессе поиска.

Формирование вероятностной меры в допустимой области. По результатам поиска локального экстремума формируются вероятностные оценки расположения начальных точек поиска, приводящих к другим локальным экстремумам.

Чтобы проиллюстрировать сказанное, рассмотрим упрощенную постановку задачи оптимального проектирования. Пусть требуется найти минимум критерия оптимальности

$$F=F(x_1, x_2, \dots, x_n) \quad (2-18)$$

при ограничениях

$$a_j \leq x_j \leq b_j, \quad j=1, 2, \dots, n. \quad (2-19)$$

Ограничения (2-19) означают, что переменные изменяются внутри n -мерного параллелепипеда с ребрами, параллельными осям координат. К такому виду задача оптимального проектирования может быть приведена с помощью метода штрафных функций (см. гл. 7).

Поиск локального экстремума с помощью одного из стохастических методов представлен на рис. 2-3. Исследуя расположение этих точек, можно выделить такие области, попадание в которые в процессе поиска приводит к тому, что последующие итерации приближаются к уже известному локальному экстремуму. Большое удаление от них соответствует менее вероятному нахождению уже известного локального экстремума или, что то же самое, более вероятному нахождению нового локального экстремума. Поскольку на расположение промежу-

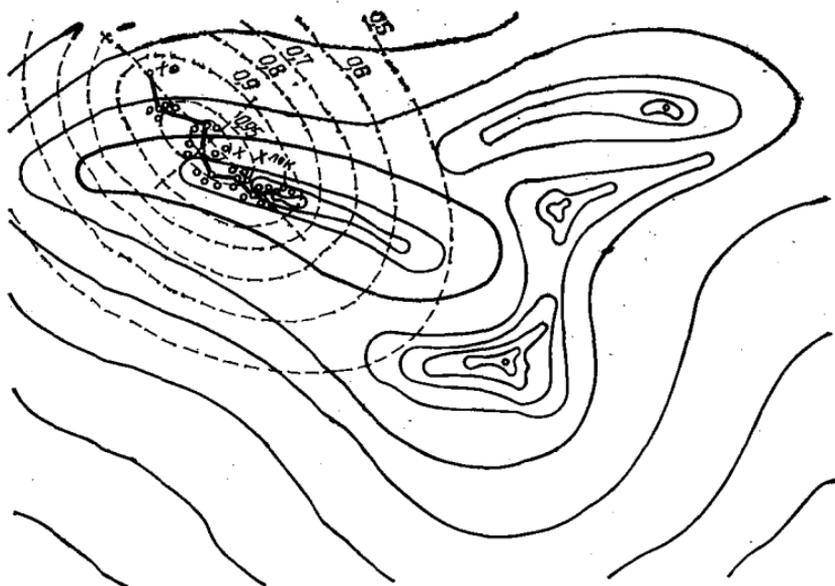


Рис. 2-3. Поиск локального экстремума.

точных точек в процессе поиска локального экстремума, особенно с помощью стохастических методов, оказывают влияние различные неконтролируемые воздействия (неизвестная геометрия критерия оптимальности, погрешности вычислений и т. д.), распределение этих точек можно до известной степени считать нормальным.

Нормальное распределение n -мерной случайной величины — вектора параметров $x = (x_1, x_2, \dots, x_n)$ характеризуется вектором математического ожидания $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$ и матрицей ковариаций C размерности n . В этом случае согласно [54] n -мерные эллипсоиды

$$(x - \bar{x})^T C^{-1} (x - \bar{x}) = r^2 \quad (2-20)$$

представляют собой поверхности постоянной плотности случайной величины x .

Центром этих эллипсоидов служит \bar{x} , а направления главных полуосей определяются собственными векторами матрицы C^{-1} . При этом наибольшему собственному значению соответствует максимальная суммарная проекция векторов $x^k - \bar{x}$, где $x^k (k=1, 2, \dots, N)$ — последовательные итерации, полученные в процессе поиска. Второе по величине собственное значение соответствует максимальной сумме проекций в $(n-1)$ -мерном пространстве (т. е. в ортогональном дополнении к уже выбранному подпространству) и т. д.

В качестве оценки математического ожидания может быть выбрано среднеарифметическое значение параметров, полученных в процессе поиска

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x^k, \quad (2-21)$$

где N — число пробных точек.

В отдельных случаях оценкой математического ожидания служит один из найденных локальных оптимумов $\bar{x} = x^{\text{лок}}$ (в этом случае начальные итерации, находящиеся далеко от экстремума, следует исключить из рассмотрения).

Для оценки матрицы ковариаций вводится матрица \mathcal{X} отклонений пробных точек от среднего значения:

$$\mathcal{X} = (x^1 - \bar{x}, x^2 - \bar{x}, \dots, x^N - \bar{x}). \quad (2-22)$$

Оценкой матрицы ковариаций служит выражение

$$C = \frac{1}{N-1} \mathbb{X} \mathbb{X}^T. \quad (2-23)$$

В соответствии с этим матричным представлением каждый коэффициент ковариации оценивается по формуле

$$c_{ij} = \frac{1}{N-1} \sum_{k=1}^N (x^k_i - \bar{x}^k_i)(x^k_j - \bar{x}^k_j). \quad (2-24)$$

Величина c_{ij} ($i=1, 2, \dots, n; j=1, 2, \dots, n$) характеризует степень связи между параметрами в процессе поиска локального экстремума.

Для определения вероятности события, состоящего в том, что из данной точки алгоритм приводит к уже известному локальному экстремуму, следует оценить, на границе какого эллипсоида вида (2-20) находится исследуемая точка. Поэтому, определив с помощью формул (2-22) и (2-24) уравнение эллипсоида, следует оценить функцию распределения величины r , которую будем называть радиусом эллипсоида. Чтобы получить это распределение, величина r^2 представляется как сумма независимых случайных величин.

Эллипсоид (2-20) ортогональным преобразованием приводится к главным осям. Это означает, что в пространстве проектируемых параметров производится линейное (ортогональное) преобразование, в результате которого оси новой системы координат направлены вдоль главных осей эллипсоида (2-20). В новой системе координат уравнение эллипсоида имеет вид:

$$\frac{1}{\lambda^2_1} u^2_1 + \frac{1}{\lambda^2_2} u^2_2 + \dots + \frac{1}{\lambda^2_n} u^2_n = r^2, \quad (2-25)$$

где $\lambda_1, \lambda_2, \dots, \lambda_n$ — собственное значение матрицы C .

С помощью линейного ортогонального преобразования каждой точке $x^p = (x^p_1, \dots, x^p_n)$ ставится в соответствие точка в новой системе координат $u^p = (u^p_1, u^p_2, \dots, u^p_n)$. Координаты вектора u^p , полученного в результате преобразования, являются независимыми, нормально распределенными случайными величинами с нулевым математическим ожиданием и дисперсиями $\lambda^2_1, \lambda^2_2, \dots, \lambda^2_n$. Отсюда следует, что величина r^2 является сум-

мой квадратов независимых случайных величин, обладающих нулевым математическим ожиданием и единичной дисперсией. Это означает, что r^2 имеет распределение χ^2 с n степенями свободы [55]. На основе известной функции распределения p_{χ^2} определяется вероятность события, состоящего в том, что поиск локального экстремума из точки x , расстояние которой до $x^{\text{лок}}$ не меньше ρ , приведет к точке $x^{\text{лок}}$. Эта вероятность оценивается по формуле

$$\pi = P\{\rho \geq r\} = 1 - P\{\rho \leq r\} = 1 - \int_0^r p_{\chi^2}(\rho) d\rho. \quad (2-26)$$

Обычно для заданной величины π (например, $\pi = 0,9$) определяется радиус r , и дальнейший поиск локальных экстремумов производится из точек, расположенных вне эллипсоида радиусом r . Кроме того, вычислив для произвольной случайной точки x_0 величину r [как результат подстановки вектора x_0 в (2-20)], можно оценить вероятность нахождения нового локального экстремума при поиске из этой начальной точки.

Пусть далее найдены экстремумы $x^{j\text{лок}}$, $j=1, 2, \dots, p$, соответствующие им эллипсоиды (2-20) и вероятности π_j , полученные из (2-26). Тогда вероятность попадания из произвольной точки x^0 в один из ранее найденных экстремумов определяется по формуле

$$P_q(x^0) = \sum_{j=1}^q \pi_j - \sum_{i < j} \pi_i \pi_j + \sum_{i < j < k} \pi_i \pi_j \pi_k - \dots \\ \dots + (-1)^q \pi_1 \pi_2 \dots \pi_q.$$

Таким образом, в процессе поиска глобального экстремума с каждой точкой области x связывается функция $P_q(x)$, определяющая целесообразность поиска локального экстремума из этой точки. Эта величина в дальнейшем уточняется и служит критерием прекращения поиска.

Изменение вероятностной меры и критерий прекращения поиска. В соответствии с полученной вероятностной мерой будем считать, что если $1 - P_q(x_0)$ не превосходит заданную величину $0 \leq \sigma \leq 1$, то с вероятностью $1 - \sigma$ поиск из точки x^0 не приведет к новому локальному

экстремуму. Чтобы определить, найдены ли все содержащиеся в области локальные экстремумы, производится выборочный контроль начальных точек поиска.

Поскольку любой план выборочного контроля не обеспечивает проверку всех точек области, всегда имеется риск закончить вычисления раньше, чем искомая вероятность достигнет $1-\sigma$, или продолжить поиск не смотря на то, что вероятность найти новый экстремум меньше σ . В связи с этим формирование плана выборочного контроля начальных точек предполагает задание величин p_0, p_1, α, β . Величины p_0, p_1 определяются из условия $p_0 \leq 1-\sigma \leq p_1$.

При этом прекращение вычислений рассматривается как ошибка, имеющая практические последствия тогда и только тогда, когда $1-\sigma \geq p_1$. В свою очередь, продолжение вычислений рассматривается как ошибка с практическими последствиями в том и только том случае, когда $1-\sigma \leq p_0$. Условие риска формулируется следующим образом: вероятность продолжать вычисления (генерирование новых начальных точек и поиск локальных экстремумов) не должна превышать некоторой наперед заданной малой величины α , когда $1-\sigma \leq p_0$, и наперед заданной малой величины β , когда $1-\sigma \geq p_1$ [8].

При условии, что проверка производится до первого неудачного испытания, общее число случайным образом выбранных начальных точек вычисляется по формуле

$$N \geq \frac{\ln(1-\alpha) - \ln \beta}{\ln(1-p_0) - \ln(1-p_1)}. \quad (2-27)$$

Испытание считается удачным, если выполнено неравенство

$$1 - P_q(x_0) \leq \sigma. \quad (2-28)$$

Если же для некоторой точки x^0 оно не выполняется, производится поиск локального экстремума с точкой x^0 в качестве начальной итерации. При этом возможны два исхода:

а) поиск приводит к уже известному $x^{j_{\text{лок}}}$, $j=1, 2, \dots, q$;

б) поиск приводит к новому локальному экстремуму $x^{q+1_{\text{лок}}}$.

Первый случай приравнивается к выполнению неравенства (2-28). Во втором — производятся вычисления

функции $P_{q+1}(x)$ на основе $P_q(x)$ и распределение итеративных приближений $x^{i_{q+1}}$, полученных в процессе поиска $x^{q+1, \text{док}}$. При этом выборочные испытания начинаются сначала для функции $P_{q+1}(x)$. Блок-схема рассмотренного алгоритма поиска глобального оптимума представлена на рис. 2-4.

В результате каждая пробная точка используется либо для пересчета вероятностной меры, либо как одно из наблюдений в серии выборочных последовательных испытаний.

Адаптация при выборе направлений движения к оптимуму. В процессе поиска первого из локальных экстремумов движение на начальных итерациях осуществляется на основе локальной информации о критерии оптимальности (по пробным точкам в окрестности текущей итерации). Такие процедуры, как метод сопряженных направлений, метод переменной метрики и др. начинают давать определенный вычислительный эффект лишь после первых n итераций. Поэтому на начальных итерациях очень важно использовать любые априорные сведения для наиболее эффективного выбора направлений поиска. После нахождения одного из локальных оптимумов выбор таких направлений осуществляется на основе матрицы C^{-1} , определенной в процессе поиска.

В качестве первых n направлений используются собственные векторы u^1, u^2, \dots, u^n матрицы C^{-1} . Собственные значения $\lambda_1, \lambda_2, \dots, \lambda_n$, соответствующие этим векторам, позволяют ранжировать их в порядке наиболее быстрого продвижения к оптимуму. В частности, суммарная проекция шагов на направление u^1 , соответствующее наибольшему собственному значению λ_1 , максимальна. Поэтому в направлении u^1 могут производиться достаточно большие пробные шаги. Второе по величине собственное значение λ_2 соответствует ортогональному к u^1 направлению u^2 , в котором производятся наибольшие пробные шаги. Третье направление u^3 позволяет осуществлять наибольшие пробные шаги в ортогональном дополнении к двум предыдущим и т. д.

Поэтому поиск очередного локального экстремума производится путем перебора направлений u^1, u^2, \dots, u^n в порядке убывания собственных значений. В некоторых случаях удается выделить два-три направления, позволяющие продвинуться достаточно близко к очередному локальному экстремуму. Формальное измерение относи-

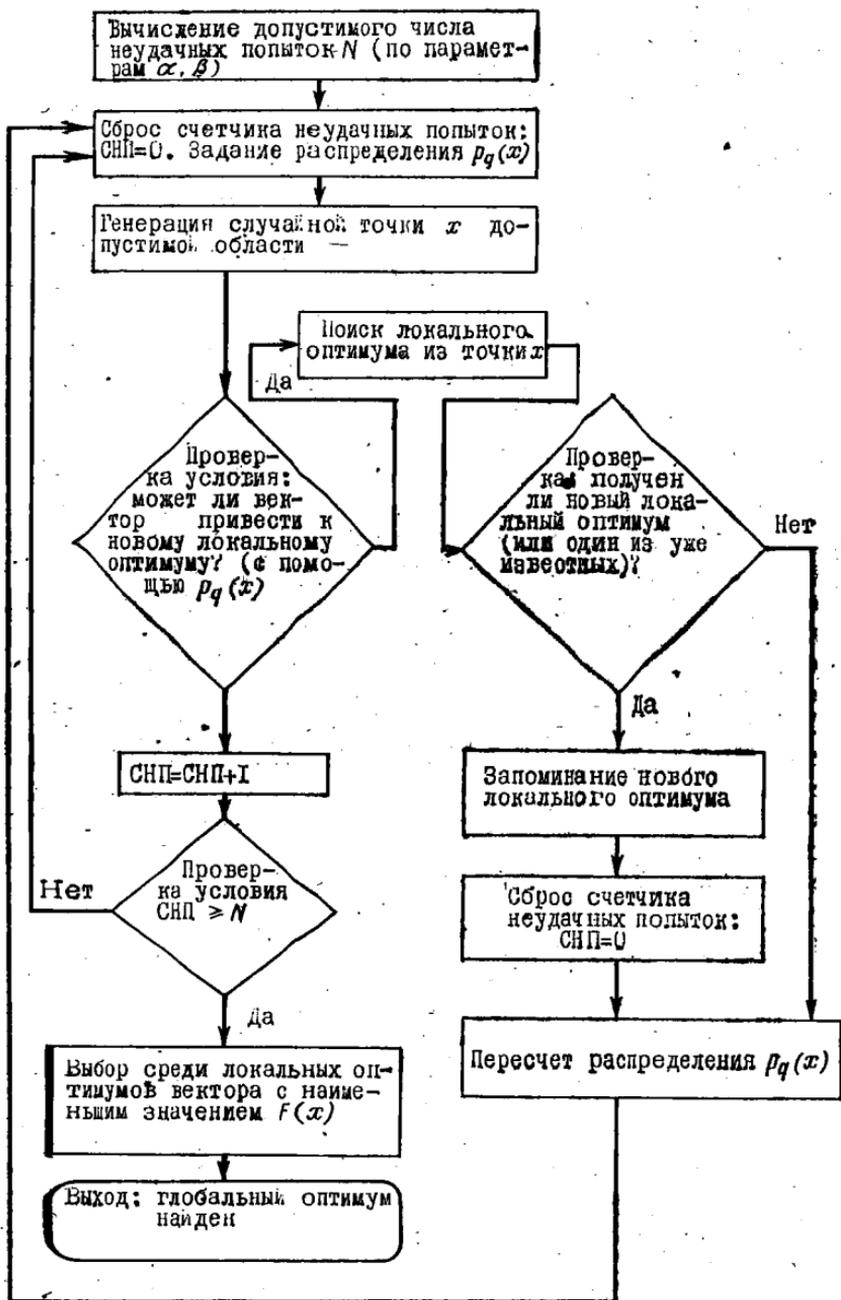


Рис. 2-4. Адаптация при поиске глобального экстремума.

тельного эффекта вдоль выделенных направлений осуществляется с помощью соотношения

$$\rho_q = \frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^n \lambda_i}.$$

Величина ρ_q показывает, какую долю проекции шагов поиска на первые q направлений составляет от их суммарной проекции на все собственные векторы u^1, u^2, \dots, u^n .

Таким образом, все рассмотренные методы поиска глобального оптимума объединяют итеративные процедуры определения локальных экстремумов с накоплением нелокальной информации о критерии оптимальности и допустимой области изменения параметров. При этом могут использоваться разнообразные методы поиска локальных экстремумов.

Глава третья

ОДНОПАРАМЕТРИЧЕСКИЕ МЕТОДЫ ОПТИМИЗАЦИИ

Методы поиска оптимума варьированием всего лишь одного параметра занимают особое место в задачах оптимального проектирования. Они используются для определения длины шага вдоль выбранного направления поиска в многопараметрических задачах оптимального проектирования.

Кроме того, методы однопараметрической оптимизации играют большую самостоятельную роль при исследовании влияния отдельных параметров на критерий оптимальности. Такое исследование необходимо на этапе разработки и оценки адекватности математической модели проектируемого изделия.

Пусть движение к оптимуму осуществляется вдоль вектора $d = (d_1, d_2, \dots, d_n)$, полученного в результате анализа изменения критерия оптимальности $F(x)$ в окрестности точки $x^k = (x^k_1, x^k_2, \dots, x^k_n)$. Тогда задача выбора длины шага поиска ставится следующим обра-

зом: найти значение параметра t , для которого выражение

$$F(x^k + td) = F(x_1^k + td_1, x_2^k + td_2, \dots, x_n^k + td_n) \quad (3-1)$$

достигает минимального значения.

Указанная задача, естественно, решается путем применения однопараметрических методов оптимизации. На рис. 3-1 представлено сечение критерия оптимальности, зависящего от двух параметров, с помощью прямой $x^k +$

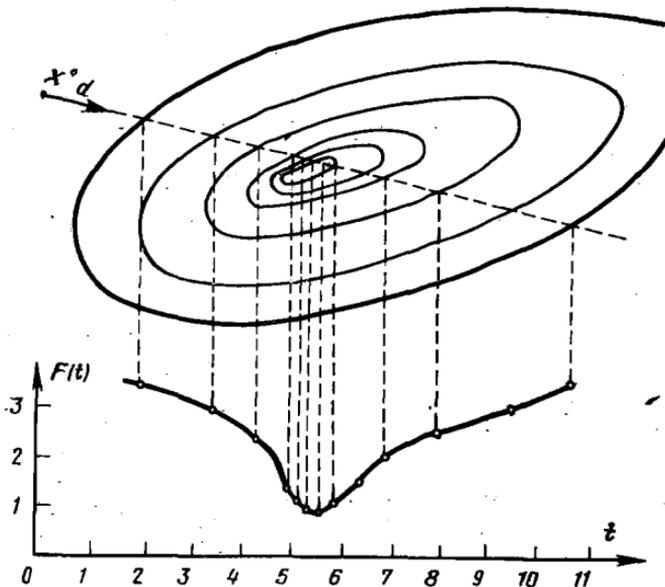


Рис. 3-1. Линии уровня критерия оптимальности $F(x_1, x_2)$ и график сечения вдоль прямой $x^k + td$.

$+td$, где x^k — начальная точка отсчета и t — параметр расстояния.

Однопараметрические методы различаются как по требованиям к степени гладкости критерия оптимальности (непрерывность, дифференцируемость и т. п.), так и по информации, используемой в каждой точке поиска. С этой точки зрения можно выделить два класса алгоритмов. Алгоритмы одного класса учитывают при определении длины шага только признак убывания (возрастания) критерия оптимальности в нескольких последовательно выбираемых точках поиска, а алгоритмы другого — изменения численных значений критерия в одной или нескольких итерациях.

3-1. ПОИСК ПО ПРИЗНАКУ ВОЗРАСТАНИЯ (УБЫВАНИЯ) КРИТЕРИЯ ОПТИМАЛЬНОСТИ

Алгоритмы этого класса позволяют найти оптимум для достаточно широкого класса функций. Они не требуют условия дифференцируемости критерия оптимальности, а только его непрерывности. Допускаются также (в конечном числе точек) разрывы первого рода.

Простейший алгоритм. Простейший метод определения оптимума однопараметрической задачи оптимального проектирования состоит в следующем. Из произвольной начальной точки x_0 осуществляется монотонное изменение параметра x с заданным шагом h в сторону убывания критерия оптимальности, т. е. $x_k = x_0 + kh$, $k = 1, 2 \dots$. После того, как функция $F(x)$ начнет возрастать, т. е. при $F(x_{k_0}) < F(x_{k_0+1})$ для некоторого k_0 , производится смена направления движения с одновременным уменьшением шага вдвое:

$$x_{k_0+2} = x_{k_0+1} - \frac{h}{2}.$$

При следующем изменении направления поиска (возрастании критерия оптимальности) снова производится уменьшение шага и т. д. Общая формула поиска имеет вид:

$$x_{k+1} = x_0 + \sum_{i=1}^{p_k} (-1)^i k_i \frac{h}{2^i}, \quad (3-2)$$

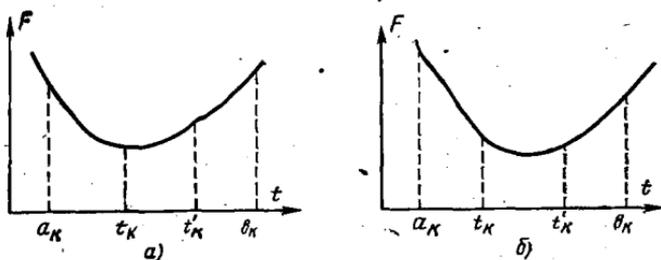
где $k = \sum_{i=1}^{p_k} k_i$; k_i — число шагов после i -й смены направления поиска; p_k — число смен направления поиска до шага k .

Указанный алгоритм довольно прост для программирования на ЭВМ, однако он требует большого числа итераций в тех случаях, когда начальная точка находится далеко от экстремума, а также при пологой функции критерия оптимальности.

Метод Фибоначчи. Повышение эффективности одномерного поиска достигается применением метода Фибоначчи. Пусть интервал поиска экстремума определяется

неравенством $a_0 \leq x \leq b_0$. Для применения метода Фибоначчи должно быть зафиксировано число N точек, в которых производится вычисление критерия оптимальности.

Пусть исходный интервал поиска $[a_0, b_0]$ сократился после k итераций до $[a_k, b_k]$. Тогда для вычисления



Фиг. 3-2. Расположение интервалов неопределенности при поиске экстремума.

a — метод Фибоначчи; *b* — метод «золотого сечения».

нового интервала $[a_{k+1}, b_{k+1}]$ выбираются точки t_k и t'_k (рис. 3-2, *a*) с помощью формул

$$t_k = \frac{\Phi_{N-1-k}}{\Phi_{N+1-k}}(b_k - a_k) + a_k, \quad t'_k = \frac{\Phi_{N-k}}{\Phi_{N+1-k}}(b_k - a_k) + a_k, \quad (3-3)$$

где Φ_k — числа Фибоначчи, которые определяют с помощью рекуррентных соотношений:

$$\Phi_k = \Phi_{k-1} + \Phi_{k-2}; \quad \Phi_0 = \Phi_1 = 1. \quad (3-4)$$

Если $F(t_k) < F(t'_k)$, то в качестве следующего интервала выбирается $[a_{k+1}, b_{k+1}] = [a_k, t'_k]$, и если $F(t_k) > F(t'_k)$, то выбирается $[a_{k+1}, b_{k+1}] = [t_k, b_k]$. Если значения критерия F в точках t и t' равны, то в качестве $[a_{k+1}, b_{k+1}]$ может быть выбран любой из интервалов $[a_k, t'_k]$ или $[t_k, b_k]$, поскольку их длины одинаковы.

Последние точки задаются формулами:

$$t'_{N-1} = \left(\frac{1}{2} + \varepsilon\right)(b_{N-1} - a_{N-1}) + a_{N-1} \quad (3-5)$$

и

$$t_{N-1} = \frac{1}{2}(b_{N-1} - a_{N-1}) + a_{N-1}, \quad (3-6)$$

где $\varepsilon > 0$ — произвольно малое число. Оно вводится на последней итерации, чтобы уменьшить интервал, содержащий минимум.

Если $F(t_{N-1}) < F(t'_{N-1})$, то минимум лежит в интервале $[a_{N-1}, t'_{N-1}]$. В противном случае минимум лежит в интервале $[t_{N-1}, b_{N-1}]$. Отсюда следует, что длина последнего интервала неопределенности равна $(b_{N-1} - a_{N-1})/2$ и относительно исходной длины составляет:

$$b_N - a_N = \frac{1}{2\Phi_N} (b_0 - a_0). \quad (3-7)$$

Таким образом, если задано число обращений к модели (в данном случае, число вычислений критерия оптимальности), то может быть указана точность, с которой производится поиск. В свою очередь, исходя из требуемой точности поиска, пользуясь формулой (3-7), можно определить число вычислений критерия оптимальности.

Метод Фибоначчи обладает наибольшей скоростью сходимости для класса непрерывных функций. Ограничением на его применение является требование наличия на отрезке поиска единственного экстремума (класс унимодальных функций). Выделение такого отрезка может осуществляться с помощью грубых методов оценки экстремума. Например, функция $F(x)$ может вычисляться при значениях $x_0, x_0+h, x_0+2h, x_0+4h$ и т. д. с фиксированным h до тех пор, пока ее значения не начнут увеличиваться. В этом случае три последних значения параметра x определяют наиболее вероятный интервал поиска локального экстремума.

Метод «золотого сечения». Другое весьма значительное неудобство, ограничивающее применение метода Фибоначчи, состоит в том, что стратегия поиска существенно зависит от заранее заданного числа экспериментов. Кроме того, для запоминания чисел Фибоначчи используется память ЭВМ. Поэтому на практике часто пользуются более простым методом «золотого сечения» (рис. 3-2,б).

В этом методе сохраняется постоянным отношение длин двух последовательных интервалов неопределенности, т. е.

$$\tau = L_k / L_{k+1} = 1,680340.$$

Для оценки интервала неопределенности на k -й итерации точки t и t' определяются по формулам:

$$t = a_{k-1} + \frac{1}{\tau} L_{k-1} = b_{k-1} - \frac{1}{\tau} L_{k-1} \quad \text{и} \quad t' = a_{k-1} + \frac{1}{\tau} L_{k-1} = b_{k-1} - \frac{1}{\tau^2} L_{k-1}.$$

Далее выбор одного из отрезков производится так же, как и в методе Фибоначчи. После N итераций длина интервала неопределенности составляет $1/\tau^{N-1}$.

Метод «золотого сечения» обладает несколько меньшей скоростью сходимости, чем метод Фибоначчи, однако при большом N длины интервалов неопределенности, полученные с помощью обоих методов, практически не различаются. Метод «золотого сечения» требует сравнительно небольшого объема памяти ЭВМ и прост в реализации.

3-2. МЕТОД КВАДРАТИЧНОЙ АППРОКСИМАЦИИ

Алгоритмы, позволяющие определить экстремум на прямой по ряду численных значений критерия оптимальности, построены на основе подходящей аппроксимации функции в некоторой окрестности поиска.

Рассмотрим один из таких методов — метод квадратичной аппроксимации, когда скорость одномерного поиска повышается за счет интерполяции функции критерия оптимальности квадратичным полиномом. Экстремум \hat{x}_k этого полинома с некоторой погрешностью дает экстремум критерия оптимальности. Повторяя процесс аппроксимации в окрестности точки \hat{x}_k , удается определить экстремум с любой наперед заданной точностью.

Пусть x_k — некоторая итерация поиска оптимума однопараметрического критерия оптимальности $F(x)$. В ее окрестности осуществляется аппроксимация функции $F(x)$ с помощью выражения

$$F(x) \approx F(x_k) + a_1(x - x_k) + a_2(x - x_k)^2. \quad (3-8)$$

Для определения коэффициентов аппроксимирующего полинома должны быть известны значения критерия оптимальности еще, по меньшей мере, в двух точках (x_{k+1} и x_{k+2}). Значение $F(x_{k+1})$ вычисляется в точке

$x_{k+1} = x_k + \Delta x_{k+1}$. Точка $x_{k+2} = x_k + \Delta x_{k+2}$ определяется по формуле

$$\Delta x_{k+2} = \begin{cases} p_k \Delta x_{k+1}, & p_k \geq 2, \text{ если } F(x_k) > F(x_{k+1}); \\ q_k \Delta x_{k+1}, & q_k \leq -\frac{1}{2}, \text{ если } F(x_k) < F(x_{k+1}). \end{cases}$$

В этом случае аппроксимирующие коэффициенты α_1 и α_2 могут быть определены из системы уравнений:

$$\begin{aligned} F(x_{k+1}) &= F(x_k) + \alpha_1 \Delta x_{k+1} + \alpha_2 \Delta x_{k+1}^2; \\ F(x_{k+2}) &= F(x_k) + \alpha_1 \Delta x_{k+2} + \alpha_2 \Delta x_{k+2}^2. \end{aligned}$$

Для существования минимума параболы (3-7) необходимо, чтобы коэффициент при квадратичном члене был положительным. Это условие запишется в виде

$$\alpha_2 = \frac{F(x_{k+2}) \Delta x_{k+1} - F(x_{k+1}) \Delta x_{k+2} + (\Delta x_{k+1} - \Delta x_{k+2}) F(x_k)}{\Delta x_{k+1} \Delta x_{k+2} (\Delta x_{k+1} - \Delta x_{k+2})}. \quad (3-9)$$

Если точка минимума существует, она может быть вычислена по формуле

$$\begin{aligned} \widehat{x}_{k+1} &= -\frac{\alpha_1}{2\alpha_2} = \frac{1}{2} \times \\ &\times \frac{F(x_{k+1}) \Delta x_{k+2} - F(x_{k+2}) \Delta x_{k+1}}{F(x_{k+2}) \Delta x_{k+1} - F(x_{k+1}) \Delta x_{k+2} + (\Delta x_{k+1} - \Delta x_{k+2}) F(x_k)}. \end{aligned}$$

Упрощение вычислений достигается, если для определения существования минимума аппроксимирующей параболы (3-7) сравнивать с нулем лишь числитель выражения (3-8), который одновременно является знаменателем отношения, определяющего \widehat{x}_{k+1} .

Поиск экстремума производится следующим образом. Пусть задана некоторая константа $\delta > 0$. Если $\alpha_2 < 0$ или $|\widehat{x}_k| > \delta$, то движение осуществляется в направлении убывания функции $F(x)$ с постоянным или монотонно возрастающим шагом. Как только функция начинает возрастать, осуществляется переход к аппроксимации квадратичным полиномом.

Аппроксимация на каждом шаге производится по значениям критерия оптимальности в трех точках. Однако все три значения вычисляются лишь на начальном

этапе поиска. В дальнейшем, как только найдено значение x_k , производится вычисление критерия оптимальности $F(x_k)$ в этой точке, а два других значения выбираются по результатам выполненных ранее вычислений в точках, ближайших к x_k . Процесс поиска повторяется до тех пор, пока не будет найден экстремум с требуемой точностью.

Чтобы осуществить поиск оптимума выражения (3-1) методом квадратичной аппроксимации, формируется полином второго порядка, зависящий от параметра t :

$$F(x^k + td) \approx a_0 + a_1 t + a_2 t^2. \quad (3-10)$$

При этом $t_0 = 0$, а следующие значения выбираются в соответствии с алгоритмом, рассмотренным в предыдущем параграфе. Пусть выбраны значения t_0, t_1, \dots, t_p . Тогда вычисление t_{p+1} осуществляется на основе значений критерия оптимальности $F(x^k + t_{p-2}d)$, $F(x^k + t_{p-1}d)$ и $F(x^k + t_p d)$. Пусть $t_{p-2} = a$, $t_{p-1} = b$, $t_p = c$, а значения в них критерия оптимальности F^a, F^b и F^c соответственно. Тогда точка минимума параболы (3-10) служит очередной итерацией параметра t . Она вычисляется по формуле

$$t_{p+1} = \frac{1}{2} \frac{(b^2 - c^2) F^a + (c^2 - a^2) F^b + (a^2 - b^2) F^c}{(a - b)(b - c)(c - a)}.$$

Условие существования минимума этой параболы имеет вид:

$$a_2 = \frac{(c - b) F^a + (a - c) F^b + (a - b) F^c}{(a - b)(b - c)(c - a)} > 0.$$

Повышая точность одномерного поиска, следует иметь в виду, что достижение общей эффективности алгоритма оптимизации возможно лишь при соответствующей точности выбора направления на экстремум.

3-3. ПРИМЕР. ПРОЕКТИРОВАНИЕ СИСТЕМЫ ОХЛАЖДЕНИЯ КРИОГЕННОГО ЭЛЕКТРОДВИГАТЕЛЯ

При проектировании криогенных электрических машин в первую очередь оцениваются значения теплопритоков в холодную зону для различных вариантов конструкций [9]. Одной из основных составляющих теплопритока в зону криогенных температур является теплоприток за счет теплопроводности по конструктивным элементам: валам, подвескам, тоководам и т. д., которые в дальнейшем обобщенно называются тепловыми мостами.

Расчетная модель теплового моста представляет собой стержень, соединяющий холодную зону, температура T_c которой принимается

равной 4,2 К, и корпус, температура которого T_0 равна 300 К. Часть теплового моста омывается потоком газообразного гелия, имеющего начальную температуру, равную температуре холодной зоны ($t_0 = T_c$).

В процессе прохождения по поверхности теплового моста гелий нагревается до температуры t , забирая при этом тепло Q , поступающее вследствие разности температур между корпусом и холодной зоной. Схема замещения охлаждаемого теплового моста представлена на рис. 3-3.

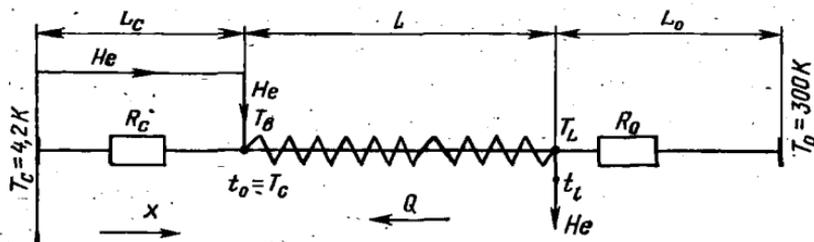


Рис. 3-3. Схема замещения охлаждаемого теплового моста.

Схема состоит из трех участков с длинами L_c , L и L_0 . Тепловые сопротивления «холодного» (R_c) и «теплого» (R_0) участков вычисляются с помощью средних значений коэффициентов теплопроводности λ_c и λ_0 на этих участках:

$$R_c = \frac{L_c}{\lambda_c s}; \quad R_0 = \frac{L_0}{\lambda_0 s}.$$

Тепловой расчет на среднем участке производится путем решения уравнений теплового баланса:

$$\frac{dQ}{dx} = \dot{m} c_p \frac{dt}{dx}; \quad \frac{dQ}{dx} = \alpha(T_c) P(T_c t), \quad (3-11)$$

где t — температура гелия; \dot{m} — расход гелия в единицу времени; c_p — теплоемкость газообразного гелия при постоянном давлении; $\alpha(T_c)$ — коэффициент теплообмена, рассчитываемый по известным формулам; P — отношение поверхности, соприкасающейся с гелием, к длине охлаждаемого участка (омываемый периметр); T_c — температура стенки теплового моста.

В качестве граничных условий принимаются условия второго рода:

$$Q(0) = \frac{1}{R_c} (T_b - T_c) \text{ на «холодном» конце;} \quad (3-12)$$

$$Q(l) = \frac{1}{R_0} (T_0 - T_L) \text{ на «теплом» конце.} \quad (3-13)$$

Принимая во внимание, что теплоприток в холодную зону $Q(0)$ компенсируется за счет теплоты испарения гелия c_L , с учетом коэффициента двухфазного гелия κ получим:

$$Q(0) = \kappa n c_L. \quad (3-14)$$

На основе построенной математической модели определяется область изменения параметра \dot{m} , представляющего собой расход гелия в единицу времени.

Необходимо определить расход гелия, который позволяет поддерживать заданную температуру T_0 в некоторой точке x_0 теплообмена. В качестве меры отклонения температуры теплообменника от заданной принята величина

$$F(\dot{m}) = [T(\dot{m}, x_0) - T_0]^2. \quad (3-15)$$

Задача состоит в поиске значения \dot{m} , для которого выражение (3-15) минимально и, кроме того, удовлетворяются ограничения (3-11), (3-13) и (3-14).

Обращение к модели (вычисление критерия и ограничений для одной точки поиска) включает следующие этапы. Для некоторого значения \dot{m}_k осуществляется решение краевой задачи (3-12), (3-13) для системы уравнений (3-11). В результате получается распределение температур хладагента и теплообменника, в частности, температура в заданной точке x_0 . По значению $T(x_0)$ определяется $F(\dot{m}_k)$.

Уже на этом примере видно, что одно обращение к модели может потребовать значительного времени счета на ЭВМ. Так, в рассматриваемой задаче проектирования тепловых элементов криогенного двигателя система дифференциальных уравнений (3-11) существенно усложняется при дополнительном учете в модели таких факторов, как влияние давления на скорость движения газа, учет конденсирования, гидравлических характеристик и т. п. Оптимизация изделия варьированием только одного параметра дает представление проектировщику, насколько существенно влияние этого параметра как на критерий оптимальности, так и на характеристики, включенные в модель в качестве ограничений.

Для оптимизации используется эффективный и достаточно простой метод «золотого сечения», позволяющий в процессе поиска оценить необходимое число итераций. С его помощью удается также исследовать изменение температурного режима в заданной точке x_0 при небольших колебаниях расхода гелия вокруг оптимального значения \dot{m}^* .

Глава четвертая

ДЕТЕРМИНИРОВАННЫЕ МЕТОДЫ ОПРЕДЕЛЕНИЯ НАПРАВЛЕНИЯ ДВИЖЕНИЯ К ОПТИМУМУ

Простейшими среди детерминированных являются прямые методы. Они позволяют определить направление поиска непосредственно по одному или нескольким значениям критерия оптимальности, без вычисления производных. Более сложными и в ряде случаев более

эффективными являются методы, использующие (приближенные или точные) значения градиента. Наконец, наиболее детальное исследование окрестности текущей итерации проводится с помощью методов Ньютона и методов переменной метрики.

Следует отметить, что излагаемые в этой главе методы, как правило, ориентированы на выбор направления поиска без учета влияния ограничений.

4-1. ПРЯМЫЕ МЕТОДЫ ОЦЕНКИ НАПРАВЛЕНИЯ

Метод покоординатного спуска. В основе метода лежит последовательная оптимизация критерия оптимальности $F(x)$ по каждому параметру.

Пусть $x^0 = (x^0_1, \dots, x^0_n)$ — начальная точка поиска. Итерация x^1 получается из x^0 при выполнении однопараметрической оптимизации по параметру x_1 , т. е.

$$x^1 = (x^1_1, x^0_2, \dots, x^0_n),$$

причем

$$F(x^1_1, x^0_2, \dots, x^0_n) = \min_{x_1} F(x, x^0_2, \dots, x^0_n).$$

Итерация x^2 получается из x^1 путем однопараметрической оптимизации по параметру x_2 и т. п. На n -й итерации из точки x^{n-1} осуществляется однопараметрический поиск по параметру x_n . Далее процесс продолжается из точки x^n снова путем варьирования параметра x_1 и т. д.

Поиск заканчивается, если дальнейшее изменение значений конструктивных параметров не дает улучшения критерия оптимальности.

Этот метод иногда используется и для решения общей задачи оптимального проектирования, включающей наряду с критерием оптимальности систему ограничений. Тогда изменение направления поиска осуществляется:

а) при завершении однопараметрической оптимизации по данному направлению;

б) при выходе на границу допустимой области.

Существенным недостатком метода покоординатного спуска является слишком большое число обращений к модели, т. е. к вычислению значений критерия оптимальности и ограничений. Кроме того, когда линии уров-

ня сильно вытянуты, по координатный спуск может не приводить к улучшению критерия оптимальности, хотя оптимум еще не достигнут. Так, при применении метода по координатного спуска в случае, представленном на рис. 4-1, за оптимум будет принята точка x^k , тогда как на самом деле экстремум находится в точке x^* .

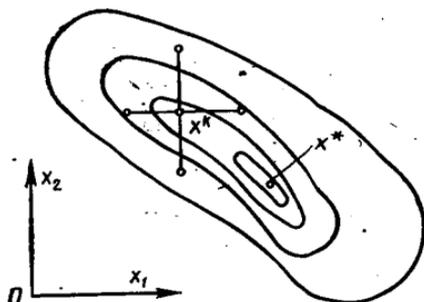


Рис. 4-1. Случай, когда метод по координатного спуска не приводит к минимуму критерия оптимальности.

Возможной модификацией алгоритма является вращение системы координат, которое приводит к методу ортогональных направлений.

Метод ортогональных направлений. Этот весьма эффективный метод определения направления поиска основан на вращении системы координат в соответствии с изменением скорости убывания критерия оптимальности.

Новые направления координатных осей определяются таким образом, чтобы одно из них соответствовало направлению наиболее быстрого убывания (на данной итерации) критерия оптимальности, а остальные находятся из условия ортогональности.

Пусть на итерации с номером k (в точке x^k) направления осей координат обозначены $p^{k_1}, p^{k_2}, \dots, p^{k_n}$. Алгоритм формирования новых направлений $p_1^{k+1}, p_2^{k+1}, \dots, p_n^{k+1}$ на итерации $k+1$ требует n пробных шагов $x^{kl}, l=1, 2, \dots, n$ по одному вдоль каждого из направлений p^{k_l} .

В качестве первой пробной выбирается точка $x^{k1} = x^k + h_1 p^{k_1}$, где h_1 — шаг вдоль вектора p_1 . Следующие пробные шаги выбираются по формуле

$$x^{kl} = x^k + h_l p^{k_l} + \sum_{j=1}^{l-1} \alpha_j h_j p^{k_j}, \quad l=2, \dots, n-1, \quad (4-1)$$

где $\alpha_{l-1} > 1$, если $F(x^{hl-1}) < F(x^{hl-2})$, и $-1 < \alpha_{l-1} < 0$ — в противном случае¹.

¹ Предполагается, что $x^{k_0} = x^k$.

Иными словами, каждая следующая пробная точка x^{kl+1} отстоит от предыдущей на величину $\alpha_{l-1}h_{l-1}p_{l-1}^k + h_l p_l^k$, где $\alpha_{l-1} > 1$, если шаг вдоль l -го направления был удачным, и $-1 < \alpha_{l-1} < 0$, если в точке x^{kl-1} значение функции увеличилось по сравнению с предыдущей пробной точкой (рис. 4-2). В результате

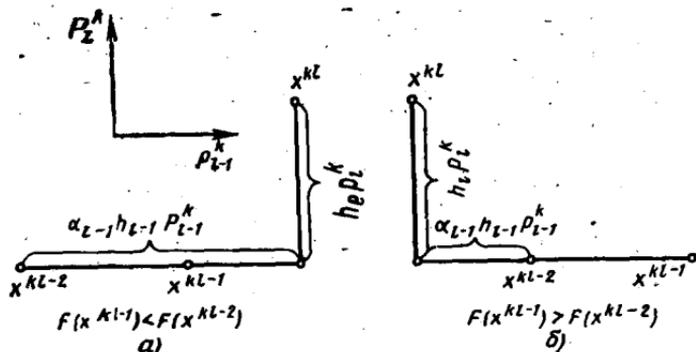


Рис. 4-2. Расположение пробных точек в методе ортогональных направлений.

а — $F(x^{kl-1}) < F(x^{kl-2})$; б — $F(x^{kl-1}) > F(x^{kl-2})$.

$$x^{k+1} = x^k + \sum_{j=1}^n \alpha_j h_j p_j^k.$$

Выбор новых направлений осуществляется следующим образом. В качестве направления p_1^{k+1} выбирается вектор

$$p_1^{k+1} = x^{k+1} - x^k = \sum_{j=1}^n \alpha_j h_j p_j^k. \quad (4-2)$$

Для построения оставшихся направлений формируются набор линейно-независимых векторов

$$q_l^{k+1} = \sum_{j=1}^n \alpha_j h_j p_j^k, \quad l = 1, 2, \dots, n. \quad (4-3)$$

При этом $q_1^{k+1} = p_1^{k+1}$, а остальные направления $p_2^{k+1}, \dots, p_n^{k+1}$ получаются методом ортогонализации.

Согласно этому методу вектор p_2^{k+1} представляется формулой

$$p_2^{k+1} = \beta_1 p_1^{k+1} + q_2^{k+1}, \quad (4.4)$$

где коэффициент β_1 определяется из условия ортогональности:

$$0 = p_1^{(k+1)T} p_2^{k+1} = \beta_1 p_1^{(k+1)T} p_1^{k+1} + p_1^{(k+1)T} q_2^{k+1}.$$

Отсюда для коэффициента β_1 получается следующее выражение:

$$\beta_1 = -p_1^{(k+1)T} q_2^{k+1} / p_1^{(k+1)T} p_1^{k+1}. \quad (4.5)$$

В общем случае, если вычислены векторы p_1^{k+1} , p_2^{k+1} , ..., p_i^{k+1} , следующий вектор p_{i+1}^{k+1} определяется формулой

$$p_{i+1}^{k+1} = \beta_1 p_1^{k+1} + \beta_2 p_2^{k+1} + \dots + \beta_i p_i^{k+1} + q_{i+1}^{k+1}, \quad (4.6)$$

где коэффициенты β_i вычисляются с помощью выражения

$$\beta_i = -p_i^{k+1} q_{i+1}^{k+1} / p_i^{(k+1)T} p_i^{k+1}, \quad i=1, 2, \dots, l. \quad (4.7)$$

Продолжая процесс поиска, в соответствии с размерностью пространства параметров получаем ортогональную систему направлений q_1, q_2, \dots, q_n .

Коэффициенты α_i подбираются эмпирически для различных типов проектируемых изделий. Наиболее эффективные результаты на тестовых примерах получены при $\alpha_i = -1/2$ при неудачных пробах и $\alpha_i = 3$ при удачных.

Некоторая модификация этого метода состоит в том, что формирование новых направлений и их ортогонализация производятся не на каждом шаге, а через некоторое число промежуточных итераций, например для x^m , x^{2m} , x^{3m} и т. д.

Хотя теоретически сходимость этого метода не доказана, практические результаты, полученные с его помощью, вполне удовлетворительны.

Модификацией метода ортогональных направлений является использование при выборе длины шага методов однопараметрической оптимизации. Начальное определение направления поиска осуществляется по формуле

(4-2) с ортогонализацией по остальным координатам. Затем осуществляется поиск минимума критерия оптимальности путем изменения только первой координаты, затем только второй и т. д. Эта операция эквивалентна применению метода покоординатного спуска в новой системе координат.

После $n+1$ итераций x^0, x^1, \dots, x^n выбор новых направлений поиска $p_1^{n+1}, p_2^{n+1}, \dots, p_n^{n+1}$ осуществляется ортогонализацией векторов $q_l^n = x^l - x^0, l=1, 2, \dots, n$, причем в качестве первого направления выбирается $p_1^{n+1} = x^n - x^0$. Процедура повторяется на итерациях $2n, 3n$ и т. д.

Метод параллельных касательных. Другой метод определения направления поиска, предложенный Пауэллом [70], получил широкое развитие как в теоретических работах, посвященных проблеме оптимизации [57, 91], так и при практическом решении задач оптимального проектирования.

Идею этого метода изложим вначале для критерия оптимальности, представленного в виде квадратичной функции от двух переменных (рис. 4-3). В качестве начального приближения выбирается произвольная точка x^0 и производится однопараметрический поиск экстремума вдоль произвольного направления. Полученная в результате точка обозначена через x^1 . Затем выбирается точка x^2 , не лежащая на прямой x^0x^1 , и осуществляется одномерный поиск из точки x^2 вдоль прямой, параллельной x^0x^1 . Точка x^3 , полученная в результате этого поиска, служит начальной для одномерного поиска на отрезке x_1x_3 . Поиск за две итерации приводит к оптимальному значению квадратичного критерия оптимальности, зависящего от двух переменных. При этом прямые x^0x^1 и x^2x^3 могут оказаться как по одну, так и по разные стороны от экстремума.

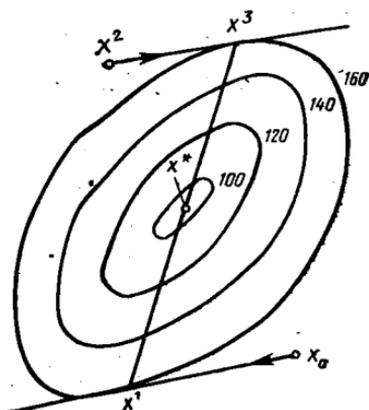


Рис. 4-3. Поиск экстремума квадратичной функции методом параллельных касательных.

В общем случае из начальной точки x^0 совершается n одномерных поисков вдоль ортогональных направлений $q^0_1, q^0_2, \dots, q^0_n$. При этом каждый следующий одномерный поиск производится из точки оптимума, полученной на предыдущем этапе. В качестве следующей итерации x^{n+1} выбирается точка минимума, полученная в результате одномерного поиска вдоль направления $q^1_1 = x^n - x^0$, а новыми направлениями являются $q^1_1, q^0_2, \dots, q^0_n$. На следующем шаге (после $2n$ итераций) производится замена вектора q^0_2 и т. д. Если критерий оптимальности является квадратичной функцией от n переменных, то одномерный поиск после n замен векторов приводит к искомому экстремуму. В общем случае процесс продолжается до тех пор, пока изменения критерия оптимальности существенны.

С помощью метода параллельных касательных минимизация критерия оптимальности, представленного квадратичной функцией от n переменных (при отсутствии ограничений), осуществляется за n итераций. При решении практических задач оптимального проектирования с более сложной функцией критерия оптимальности число итераций, естественно, увеличивается, однако и в этих случаях метод параллельных касательных приводит к высокой скорости сходимости последовательности итераций.

Для иллюстрации вычислительных особенностей методов ортогональных направлений и параллельных касательных рассмотрим задачу поиска минимума функции (1-17), представленной на рис. 1-5. Начальной точкой поиска служит точка $x^0 = (-1; 2; 1,0)$. Результаты вычислений приводятся в табл. 4-1.

Таблица 4-1

Номер итерации	Метод ортогональных направлений			Метод параллельных касательных		
	Значение критерия оптимальности $F(x_1, x_2)$	Число вычислений $F(x_1, x_2)$	Число одномерных поисков	Значение критерия оптимальности $F(x_1, x_2)$	Число вычислений $F(x_1, x_2)$	Число одномерных поисков
0	24,0	1	0	24,0	1	0
5	1,4	43	11	1,1	58	15
10	0,13	91	22	$0,054 \cdot 10^{-1}$	122	29
15	$0,056 \cdot 10^{-1}$	139	32	$0,013 \cdot 10^{-14}$	158	39
20	$0,070 \cdot 10^{-8}$	178	43	—	—	—
21	$0,015 \cdot 10^{-10}$	187	46	—	—	—

Хотя для рассмотренного тестового примера скорость сходимости метода параллельных касательных выше, чем метода ортогональных направлений, оба они могут успешно использоваться при решении практических задач оптимизации.

4.2. ВЫБОР НАПРАВЛЕНИЯ ПОИСКА С ПОМОЩЬЮ ГРАДИЕНТА

Определение градиента. Градиентом функции $F(x)$ векторного аргумента $x = (x_1, x_2, \dots, x_n)$ называется вектор, координатами которого служат частные производные по соответствующим переменным:

$$\nabla F = \left(\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \dots, \frac{\partial F}{\partial x_n} \right).$$

В некоторых случаях градиент ∇F функции F будет для краткости обозначаться g .

Геометрически градиент критерия оптимальности представляется в виде вектора, направленного в сторону наиболее быстрого возрастания функции F . Вектор $-\nabla F$, называемый антиградиентом, направлен в сторону наиболее быстрого убывания критерия оптимальности. Это означает, что продвижение в направлении $-\nabla F$ (в окрестности рассматриваемой точки) на некоторое расстояние Δx позволит достигнуть меньшего значения критерия оптимальности, чем движение на то же расстояние по любому другому направлению. На указанном свойстве градиента основано его использование при поиске оптимального решения.

Из математического анализа известно, что для достаточно широкого класса функций $F(x)$ можно получить аналитическое выражение для частных производных в любой точке допустимой области. В этом случае, составив программу вычислений выражений $\partial F / \partial x_i$, $i = 1, 2, \dots, n$ на ЭВМ, можно определить градиент в любой заданной точке пространства конструктивных параметров. Однако на практике такой подход возможен лишь в исключительных случаях, так как в практических задачах оптимального проектирования формулы для расчета критерия оптимальности настолько сложны, что аналитическое определение частных производных становится невозможным. В этом случае применяются разностные методы для приближенной оценки градиента.

В основе детерминированных методов оценки градиента критерия оптимальности лежит однозначное правило, определяющее расположение пробных точек. Частные производные функции критерия оптимальности, представляющие собой координаты вектора-градиента, аппроксимируются с помощью отношения разности значений функции в пробных точках к разности значений аргумента в этих точках.

Рассмотрим эту аппроксимацию более подробно для однопараметрической задачи оптимального проектирования. Вычисляя значения критерия $F(x)$ в точках x и $x+h$, получаем для аппроксимации производной следующее выражение:

$$\frac{dF}{dx} \approx F_h \equiv \frac{F(x+h) - F(x)}{h}. \quad (4-8)$$

Для оценки точности аппроксимации представим $F(x+h)$ с помощью отрезка ряда Тейлора:

$$F(x+h) = F(x) + F'(x)h + F''(x)\frac{h^2}{2} + F'''(x)\frac{h^3}{6} + O(h^4). \quad (4-9)$$

Отсюда для разности производной и аппроксимирующего выражения получаем:

$$\frac{F(x+h) - F(x)}{h} - F'(x) = F''(x)\frac{h}{2} + O(h^2), \quad (4-10)$$

т. е. с уменьшением шага h точность приближения возрастает со скоростью линейной функции.

Другая формула для приближенного вычисления производной функции $F(x)$ может быть получена по значениям $F(x)$ в точках $x+h$ и $x-h$. В этом случае

$$\frac{dF}{dx} \approx F_{2h} \equiv \frac{F(x+h) - F(x-h)}{2h}. \quad (4-11)$$

Представим $F(x-h)$ с помощью отрезка ряда Тейлора:

$$F(x-h) = F(x) - F'(x)h + F''(x)\frac{h^2}{2} - F'''(x)\frac{h^3}{6} + O(h^4). \quad (4-12)$$

Отсюда на основании (4-9) получаем выражение для погрешности аппроксимации:

$$\frac{F(x+h) - F(x-h)}{2h} - F'(x) = F'''(x) \frac{h^2}{6} + O(h^3). \quad (4-13)$$

Из этой формулы следует, что с уменьшением длины шага h точность аппроксимации производной по формуле (4-11) растет быстрее, чем по формуле (4-9). При заданной точности вычисления производной для формулы (4-9) требуется меньшее значение шага, чем при использовании формулы (4-11).

Следует заметить, что при слишком малом шаге погрешность вычисления производной может оказаться недопустимо большой из-за ошибок округления, которые неизбежны при вычислениях с фиксированным (конечным) числом разрядов.

В общем случае, когда критерий оптимальности есть функция, зависящая от n переменных, для оценки частных производных с помощью конечных разностей служат следующие формулы:

а) при вычислении критерия оптимальности в точках $x = (x_1, \dots, x_i, \dots, x_n)$ и $x + \Delta x^i = (x_1, \dots, x_i + h_i, \dots, x_n)$:

$$\frac{\partial F}{\partial x_i} \approx F_{i, h} = \frac{F(x + \Delta x^i) - F(x)}{h_i}; \quad (4-14)$$

б) при вычислении критерия оптимальности в точках $x + \Delta x^i$ и $x - \Delta x^i$:

$$\frac{\partial F}{\partial x_i} \approx F_{i, 2h} = \frac{F(x + \Delta x^i) - F(x - \Delta x^i)}{2h_i}. \quad (4-15)$$

Как и при минимизации функции одной переменной, точность аппроксимации первых частных производных по формуле (4-15) выше, чем по формуле (4-14).

Однако при достаточно сложном выражении для критерия оптимальности одновременно с точностью вычисления градиента становится существенным также и число арифметических операций, необходимых для его вычисления. Поэтому следует учитывать, сколько раз приходится вычислять критерий оптимальности для определения градиента.

При вычислении градиента по формуле (4-14) требуется $n+1$ раз вычислить критерий оптимальности:

в точке \bar{x} и при поочередных изменениях каждой координаты x_i на величину h_i .

Оценка градиента по формуле (4-15) требует $2n$ вычислений критерия оптимальности. Кроме того, во многих случаях необходимо вычисление значения критерия оптимальности в точке x (как правило, для осуществления одномерного поиска). Поэтому при использовании формулы (4-15) на каждой итерации вычисляются $2n+1$ значений критерия оптимальности. Такое большое число вычислений критерия оптимальности разумно лишь в тех случаях, когда точность вычисления градиента имеет решающее значение для процедуры поиска экстремума. Обычно это целесообразно вблизи экстремума, когда другие способы определения направления движения оказываются неэффективными. Напротив, на начальных стадиях поиска, когда градиент быстро устаревает ввиду успешного продвижения с помощью одномерного поиска вдоль найденных направлений, его можно определять с помощью формулы (4-14). Во многих случаях для вычисления градиента оказываются также полезными методы, основанные на случайном расположении пробных точек.

Градиентные методы. Методы поиска, в которых направление движения от итерации x^k к x^{k+1} определяется градиентом (антиградиентом), вычисленным в точке x^k , носят название градиентных методов. Вдоль направления антиградиента производится шаг определенной длины, и в полученной точке x^{k+1} (новой итерации) снова определяется градиент критерия оптимальности. Градиентные методы отличаются друг от друга способом выбора шага вдоль вектора антиградиента.

Простейший способ — движение из точки x^k по антиградиенту с постоянным шагом $h = \text{const}$, т. е.

$$x^{k+1} = x^k - h \frac{\nabla F}{\sqrt{\nabla F^T \nabla F}}. \quad (4-16)$$

Если окажется, что значение критерия в точке x^{k+1} превосходит величину $F(x^k)$, то поиск возвращается в точку x^k , и из нее вдоль антиградиента производится перемещение с шагом $h/2$. Дробление шага производится до тех пор, пока его значение не станет меньше некоторого малого положительного числа δ .

Поиск прекращается, как только модуль вектора градиента будет меньше заданного $\varepsilon > 0$, т. е. при $\nabla F^T \nabla F \leq \varepsilon$.

В некоторых случаях более эффективным является движение с шагом, пропорциональным модулю градиента:

$$x^{k+1} = x^k - h \nabla F. \quad (4-17)$$

Существует еще один вариант градиентного метода. После определения градиента критерия оптимальности в точке x^k производится движение вдоль прямой $x^k - t \nabla F(x^k)$ до точки, в которой достигается минималь-

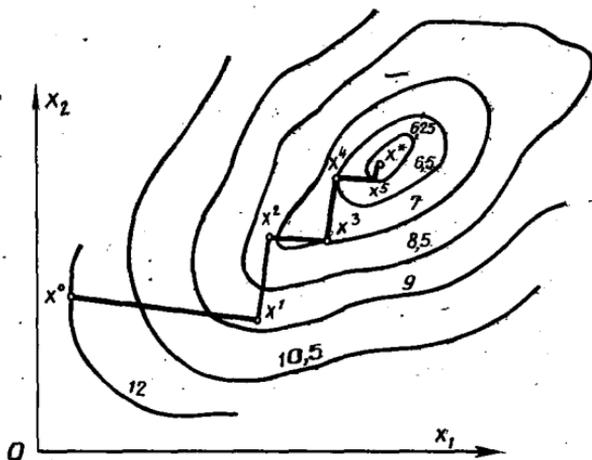


Рис. 4-4. Поиск экстремума методом наискорейшего спуска.

ное значение функции. Затем в этой точке снова определяется градиент, и движение совершается по прямой согласно направлению нового антиградиента и т. д., пока не будет достигнута точка, соответствующая наименьшему значению функции $F(x)$, или не будет пересечена граница допустимой области. Такой метод поиска называется методом наискорейшего спуска. На рис. 4-4 приведен пример движения при поиске методом наискорейшего спуска оптимума для критерия оптимальности, зависящего от двух переменных. Направление $\nabla F(x^k)$ является касательным к линии (поверхности) уровня в точке x^{k+1} , и, следовательно, градиент $\nabla F(x^{k+1})$ в точке x^{k+1} ортогонален $\nabla F(x^k)$.

Как правило, метод наискорейшего спуска требует вычисления градиента в меньшем числе точек, чем гра-

диентные методы поиска с постоянным шагом или с шагом, пропорциональным модулю градиента. Дальнейшее повышение скорости сходимости методов оптимизации достигается за счет уточнения направления движения путем использования результатов вычисления градиента на предыдущих итерациях.

Методы сопряженных направлений. Эффективность градиентных методов поиска существенно снижается, если гиперповерхности постоянного уровня функции F вытянуты в каком-либо направлении. В некоторых случаях длина шага, вычисляемая согласно (4-17), оказывается сравнимой с погрешностями счета, что приводит к заикливанию процедуры поиска.

Для повышения эффективности поиска наряду с градиентом используют также матрицу вторых производных (гессиан) Γ критерия оптимальности. С помощью гессиана удастся выбрать направления поиска, в большей мере соответствующие геометрии критерия оптимальности, чем обычные ортогональные направления. С этой целью вводится понятие сопряженности векторов, обобщающее понятие ортогональности.

Направления s^1, s^2, \dots, s^n называются *сопряженными* относительно симметричной и положительно определенной матрицы Γ , если

$$s^{i^T} \Gamma s^j = 0 \text{ при } i \neq j. \quad (4-18)$$

Хотя в определении сопряженных направлений участвует гессиан Γ , эти направления удастся вычислить, используя только значения критерия оптимальности (и в некоторых случаях градиенты). В частности, сопряженными являются направления, полученные с помощью метода параллельных касательных.

Наиболее эффективным среди методов выбора сопряженных направлений является метод сопряженных градиентов. С его помощью удастся наглядно представить свойство сопряженности как ортогональность вращений вектора градиентов.

Пусть x^k и x^{k+1} — два последовательных приближения к экстремуму критерия оптимальности $F(x)$. Тогда с точностью до слагаемых второго порядка справедливо равенство

$$\nabla F(x^{k+1}) - \nabla F(x^k) \approx \Gamma(x^{k+1} - x^k). \quad (4-19)$$

Если направление s^i является сопряженным направлением $\Delta x^k = x^{k+1} - x^k$, т. е. $s^{iT} \Gamma \Delta x = 0$, то в силу равенства (419) имеем:

$$s^{iT} [\nabla F(x^{k+1}) - \nabla F(x^k)] = s^{iT} \Gamma \Delta x = 0. \quad (4-20)$$

Это и означает, что вектор s^i ортогонален приращению градиента $\nabla F(x^{k+1}) - \nabla F(x^k)$.

Указанное свойство позволяет сформулировать достаточно простой вычислительный алгоритм построения сопряженных направлений (сопряженных градиентов).

Поскольку в дальнейших выкладках часто используется градиент $\nabla F(x^k)$, введем для него более краткое обозначение g^k . Разность градиентов в точках x^{k+1} и x^k будет обозначаться $\Delta g^k = g^{k+1} - g^k$. Начальная итерация поиска при заданной точке x^0 включает вычисление градиента g^0 , критерия оптимальности $F(x)$ в этой точке и поиск минимума в направлении g^0 . Обозначим точку этого минимума x^1 и градиент в этой точке g^1 . Тогда, полагая для первого из сопряженных направлений $s^1 = -g^0$, второе сопряженное направление определяется в виде

$$s^2 = g^1 + \beta_0 s^1,$$

где β_0 — неопределенный коэффициент.

Он вычисляется с помощью условия сопряженности:

$$0 = s^{2T} \Delta g^1 = (g^1 + \beta_0 g^0)^T (g^1 - g^0) = g^{1T} g^1 - g^{1T} g^0 + \beta_0 g^{0T} g^1 - \beta_0 g^{0T} g^0. \quad (4-21)$$

В силу того, что переход от итерации x^0 к x^1 совершается с помощью метода наискорейшего спуска, векторы g^0 и g^1 ортогональны. Поэтому $g^{1T} g^0 = g^{0T} g^1 = 0$. Отсюда получаем:

$$\beta_1 = \frac{g^{1T} g^1}{g^{0T} g^0}. \quad (4-22)$$

В общем случае, представляя очередное направление поиска как линейную комбинацию предыдущих сопряженных направлений:

$$s^{k+1} = g^k + \beta_1 s^1 + \beta_2 s^2 + \dots + \beta_k s^k, \quad (4-23)$$

получаем:

$$\beta_i = 0, \quad i = 1, 2, \dots, k-1; \quad \beta_k = \frac{g^{(k-1)T} g^{(k-1)}}{g^{(k-1)T} g^{(k-1)}}, \quad (4-24)$$

т. е. β_k является единственным, отличным от нуля коэффициентом.

Если критерий оптимальности представлен квадратичной функцией, то $F(x^{k+1})$ является ее минимальным значением на подпространстве, порожденном векторами s^1, s^2, \dots, s^k . Так, например, в точке x^1 достигается минимум вдоль прямой, определяемой направлением s^1 , в x^2 — минимум на плоскости, определяемой векторами x^1 и x^2 , и т. д. Отсюда следует, в частности, что минимум квадратичной функции с помощью метода сопряженных градиентов достигается в точности за n шагов.

В случае критерия оптимальности произвольного вида, не являющегося квадратичной функцией, получаем итеративный процесс поиска оптимума. Он позволяет для заданной погрешности получить приближенное решение быстрее, в ряде случаев, чем это позволяют сделать методы наискорейшего спуска и параллельных касательных.

Недостатком метода является необходимость довольно точного определения минимума на каждой итерации, что приводит к увеличению числа итераций в методах однопараметрического поиска. Указанный недостаток ограничивает возможности метода сопряженных градиентов, особенно в «овражных» ситуациях.

4.3. МЕТОДЫ ПЕРЕМЕННОЙ МЕТРИКИ

Движение к экстремуму по сопряженным направлениям позволяет существенно ускорить поиск, поэтому в работах, направленных на развитие методов оптимизации, значительное внимание уделяется улучшению выбора сопряженных направлений. Наиболее эффективной модификацией является поиск сопряженных направлений с одновременным накоплением информации о гессиане критерия оптимальности. Рассматриваемый класс методов носит название методов переменной метрики или квази-ньютоновских методов.

Метод Ньютона и условие переменной метрики. В основе метода, так же как и в случае сопряженных гра-

диентов, лежит равенство (4-19). Поскольку в точке экстремума градиент равен нулю, необходимым условием того, чтобы экстремум достигался в точке x^{k+1} , является

$$x^{k+1} = x_k - \Gamma^{-1}(x^k) g^k. \quad (4-25)$$

Итеративный процесс, определяемый этой формулой, носит название метода Ньютона. Он обладает значительно более высокой скоростью сходимости по сравнению со всеми ранее рассмотренными методами, но в силу двух существенных недостатков находит ограниченное применение в практических расчетах.

Первый из них — необходимость вычисления гессiana критерия оптимальности в каждой точке. Поскольку критерий обычно имеет довольно сложную форму, гессиан может быть вычислен только с помощью конечных разностей второго порядка. Это означает, что в окрестности точки x^k необходимо произвести $n(n+1)/2$ вычислений критерия оптимальности.

Такое число вычислений значительно увеличивает время, затрачиваемое на получение одной итерации. Если же гессиан изменяется, то информация о нём быстро устаревает и в следующей точке все вычисления производятся заново.

Другим существенным недостатком метода является то, что его применение возможно лишь в небольшой области вблизи экстремума. Это объясняется требованием положительной определенности гессиана, которое может нарушиться при удалении от экстремума.

Несмотря на указанные недостатки, метод Ньютона послужил основой для создания целого класса методов, весьма эффективных, особенно в случае оптимизации при отсутствии ограничений. Этот класс получил название *методов переменной метрики*.

Идея методов переменной метрики состоит в том, чтобы использовать информацию о градиенте критерия оптимальности для приближенного вычисления гессиана. С этой целью формируется последовательность матриц H_k , обладающих следующим свойством:

$$H_j(g^j - g^{j-1}) = x^j - x^{j-1}, \quad j=1, 2, \dots, k. \quad (4-26)$$

Сравнение этого выражения с формулой (4-25) показывает, что для квадратичной функции после n шагов в линейно независимых направлениях матрицы H_n и Γ^{-1} совпадают. В общем случае матрицы H_k аппрокси-

мируют гессиан критерия оптимальности, и эта аппроксимация тем точнее, чем меньше изменяется гессиан в процессе движения к экстремуму.

Выбор матрицы H_k . Все алгоритмы переменной метрики различаются между собой способом вычисления матрицы H_k . Если задана рекуррентная формула пересчета H_k , то один шаг алгоритма может быть представлен в следующем виде:

1) вычислить $p^{k+1} = -\alpha_k H_k g^k$ и изменять α_k до достижения минимума функции $F(x^k - \alpha_k H_k g^k)$, используя одномерный поиск;

2) положить $x^{k+1} = x^k + p^{k+1}$ и вычислить градиент критерия оптимальности в точке x^{k+1} ;

3) вычислить разность градиентов $y^k = g^{k+1} - g^k$ и преобразовать матрицу H_k в H_{k+1} по рекуррентной формуле;

4) положить $k+1$ равным k и повторить все процедуры, начиная с п. 1.

При формировании правила прекращения поиска следует учитывать, что число итераций должно превосходить размерность пространства поиска, поскольку для эффективного приближения матрицы Γ^{-1} необходимо не меньше чем n итераций. Другими критериями могут

быть: малая норма градиента: $g^k, g^k \leq \varepsilon$, или малое значение шага: $(x^{k+1} - x^k)^T (x^{k+1} - x^k) \leq \delta$.

Рекуррентная формула для вычисления H_k может быть получена из следующих соображений. В качестве начальной итерации выбирается единичная матрица $H_0 = E$. Если на k -м шаге выбрана матрица, которая не удовлетворяет равенству (4-26), то формируется матрица $H_{k+1} = H_k + D_k$ так, чтобы выполнялось соотношение

$$(H_k + D_k) y^k = \sigma^k, \quad (4-27)$$

где $\sigma^k = x^{k+1} - x^k$. Отсюда следует, что

$$D_k y^k = \sigma^k - H_k y^k. \quad (4-28)$$

С помощью свойства симметричности гессиана Барнсом и Розеном [37] получена формула для вычисления поправки:

$$D_k = \frac{(\sigma^k - H_k y^k) (\sigma^k - H_k y^k)^T}{(\sigma^k - H_k y^k)^T y^k}. \quad (4-29)$$

Матрица D_k имеет ранг 1, т. е. обладает только одним ненулевым собственным значением.

Другой способ задания матрицы предложен Флетчером и Пауэллом [70]. В этом способе матрица D_k имеет ранг 2 и записывается в виде

$$D_k = \frac{\sigma^k \sigma^k T}{\sigma^k T \sigma^k} - \frac{H_k y^k y^k T H_k T}{y^k T H_k y^k} \quad (4-30)$$

Обе матрицы удовлетворяют условию (4-26). Отличие состоит в том, что первый метод рассматривает правую часть равенства (4-28) как один вектор, а второй — как два независимых вектора.

Реализация этих методов требует порядка $3n^2$ умножений, не считая затрат на определение длины шага и градиента критерия оптимальности. В отличие от метода Ньютона в данном случае производится только вычисление градиента (а не гессиана), поэтому требуется только $n+1$ вычисление значений критерия оптимальности в окрестности очередной итерации.

В [57] доказана сходимость метода переменной метрики [с матрицей в форме (4-30)] для выпуклых, дважды непрерывно дифференцируемых функций, обладающих следующим свойством: существует положительная константа, которая является нижней границей всех собственных значений гессианов функции $F(x)$.

На основе рассмотренных двух типов преобразований матрицы H_k можно определить некоторое семейство поправок, зависящее от одного параметра t . Любая поправка, промежуточная по отношению к обоим методам, может быть записана в виде

$$D_k y^k = t \sigma^k + (1-t) \sigma^k - H_k y^k,$$

который, с одной стороны, эквивалентен равенству (4-28), а с другой — позволяет перейти к параметрическому представлению матрицы D_k . Поправка

$$D_k = t \frac{\sigma^k \sigma^k T}{\sigma^k T \sigma^k} + \frac{((1-t) \sigma^k - H_k y^k) ((1-t) \sigma^k - H_k y^k) T}{((1-t) \sigma^k - H_k y^k) T y^k} \quad (4-31)$$

при любом t удовлетворяет условию (4-26), и потому позволяет минимизировать квадратичную функцию от n переменных за n итераций. При $t=0$ выражение (4-31) представляет собой поправку метода Барнса — Розена, а при $t=1$ — Флетчера — Пауэлла.

Введение параметра t позволяет довольно просто осуществлять выбор метода поиска, а также разрабатывать рекомендации по выбору значения параметра t для однотипных задач оптимального проектирования. Более того, этот выбор может осуществляться автоматически, с помощью ЭВМ, на основе обработки статистики числа итераций, необходимых для поиска оптимального решения. Однако величина t не может принимать произвольные значения. Для некоторых значений t вектор $H_{k+1}g_{k+1}$ при определенных α_k обращается в нуль, не смотря на то что оптимум еще не достигнут. Ограничением на пределы изменения параметра t является условие

$$t > (\alpha_k - 1) / \alpha_k, \quad (4-32)$$

полученное в [87].

При оценке алгоритмов с точки зрения их эффективности в практических расчетах следует учитывать, что оценка градиента производится не аналитически, а с помощью конечных разностей и, кроме того, однопараметрический поиск производится с конечной наперед заданной точностью.

Для проверки эффективности метода в этих условиях использовалась функция Розенброка $R(x_1, x_2)$, определенная формулой (1-17) и имитирующая «овражную ситуацию» при поиске. Параллельно с функцией $R(x_1, x_2)$ применялась так называемая функция Вейбула

$$W(x_1, x_2, x_3) = \sum_{i=1}^{99} \left[\exp\left(-\frac{1}{x_1}(\tau_i - x_2)^{x_3}\right) - y_i \right]^2, \quad (4-33)$$

где y_i и τ_i — известные величины (99 измерений некоторого параметра τ при значениях y_i от 0,1 до 0,99 с шагом 0,01). В табл. 4-2 представлены результаты поиска для различных значений параметра t .

Из таблицы следует, что метод Барнса — Розена неустойчив по отношению к изменению начальных условий. Метод Флетчера — Пауэлла более стабилен, однако требует большего числа обращений к вычислению значений критерия оптимальности, чем метод Барнса — Розена. Хорошая сходимость получена также для случая, когда параметр t изменяется от итерации к итерации и для $t = \infty$, однако число итераций существенно зависит

от вида критерия оптимальности. Поэтому выбор параметра t следует осуществлять на основе опыта решения ряда однотипных задач. Исследования с целью выбора параметра t могут проводиться как на реальных задачах оптимального проектирования, так и с использованием тестовых примеров.

Таблица 4-2

Функция	Начальные точки	$t = 0$		$t = 1$		$t = \frac{1}{2(2\alpha - 1)/\alpha}$		$t = \infty$	
		Число итераций	Число вычислений критерия	Число итераций	Число вычислений критерия	Число итераций	Число вычислений критерия	Число итераций	Число вычислений критерия
$R(x_1, x_2)$	(1; -1, 2)	*	—	13	65	14	56	14	56
	(2; -2)	20	123	20	123	20	78	18	70
	(-3, 635; 5, 621)	**	—	26	213	22	105	21	96
	(0, 639; -0, 221)	**	—	16	72	16	66	16	58
	(1, 489; -2, 547)	*	—	20	134	20	83	18	77
$W_4(x_1, x_2, x_3)$	(5; 0, 15; 2, 5)	18	92	29	239	20	90	19	112
	(250; 0, 3; 5)	**	—	***	—	32	142	35	122
	(100; 3; 12, 5)	41	98	****	—	25	150	28	149

Примечание. * — заданная точность не достигнута после 50 итераций; ** — матрица H_k потеряла свойство положительной определенности; *** — заданная точность не достигнута после 31 итерации и 279 обращений к модели; **** — заданная точность не достигнута после 50 итераций и 499 обращений к модели.

При практических расчетах методы переменной метрики не всегда дают удовлетворительные результаты вследствие ошибок округления, возникающих в процессе счета на ЭВМ, или неудачного начального масштабирования параметров. В результате некоторые из матриц H_k оказываются вырожденными или не удовлетворяют условию положительной определенности.

Выходом в такой ситуации является повышение точности вычислений на ЭВМ путем использования слов двойной длины и изменения масштабов переменных. Кроме того, можно обновить процесс счета, приняв точку x^k , в которой обнаружилась вырожденность матрицы H_k , за новую начальную точку и заменить H_k единичной

матрицей. В течение нескольких следующих итераций она будет сохранять положительную определенность.

Ввиду того что в методах переменной метрики достаточно полно учитывается локальная информация, их целесообразно применять в окрестности оптимального решения.

Глава пятая

СТОХАСТИЧЕСКИЕ МЕТОДЫ ОПРЕДЕЛЕНИЯ НАПРАВЛЕНИЯ ДВИЖЕНИЯ К ОПТИМУМУ

5-1. СТОХАСТИЧЕСКИЙ ПОДХОД В ОПТИМАЛЬНОМ ПРОЕКТИРОВАНИИ

В основе стохастических методов лежит внесение элементов случайности в процедуру формирования пробных точек, используемых для определения направления поиска. На каждой итерации информация о предпочтительных направлениях поиска либо отсутствует, либо выражена в довольно общей форме, отражающей предысторию движения. Поскольку, кроме того, указанная информация быстро устаревает, ее ценность резко снижается, как только сделан некоторый шаг поиска.

В результате напрашивается вывод: вместо однозначного выбора наиболее предпочтительного направления (как это делается в ряде детерминированных методов) сформировать некоторую случайную величину, распределение которой отражает информированность (расчетчика, ЭВМ и т. д.) о степени предпочтительности отдельных направлений поиска. Использование элементов случайности при поиске позволяет не тратить значительных усилий (т. е. времени счета, числа операций и т. д.) на выбор расположения пробных точек, а продвигаться к оптимуму на основе минимальной, достаточно просто получаемой информации.

При стохастическом методе поиска

$$x^{k+1} = x^k + \alpha_k \eta^k, \quad (5-1)$$

где α_k — длина шага и η^k — вектор представляют реализацию нормированной случайной величины с некоторым законом распределения.

Этот закон распределения формируется на основе расчетов критерия оптимальности в пробных точках, расположенных в окрестности точки x^k . Пусть вектор

$\xi^{k1}, \xi^{k2}, \dots, \xi^{kp_k}$ соответствует независимым случайным точкам на сфере радиуса ε с центром в x^k . Если распределение случайной величины η^k зависит только от пробных точек, расположенных в окрестности k -й итерации x^k , то говорят о независимом случайном поиске или поиске без адаптации. Если же на распределение η^k оказывает влияние предыстория (ретроспектива) поиска, то говорят о поиске с адаптацией. Число пробных точек p_k может изменяться от итерации к итерации, как правило, $p_{k+1} \geq p_k$.

Таким образом, внесение элементов случайности в процедуру выбора направления поиска эквивалентно представлению о критерии оптимальности как о некотором «черном ящике», который на случайные приращения независимых переменных реагирует случайным, точнее, неизвестным для исследователя образом.

Процедура, а следовательно, и программа вычисления вектора с помощью стохастических методов поиска значительно проще, чем при использовании детерминированных алгоритмов. Это преимущество обусловило широкую популярность стохастическим методам.

5-2. ПРЯМЫЕ СТОХАСТИЧЕСКИЕ МЕТОДЫ ПОИСКА

Поиск без адаптации. Простейший алгоритм поиска без адаптации включает выбор случайной точки ξ^{k1} на гиперсфере единичного радиуса и вычисление критерия оптимальности в точке $x^k + \delta \xi^{k1}$.

Если $F(x^k) > F(x^k + \delta \xi^{k1})$, то ξ^{k1} выбирается в качестве направления поиска, в противном случае движение к оптимуму осуществляется в противоположном направлении $-\xi^{k1}$. Величина δ может рассматриваться как величина шага поиска, т. е. $x^{k+1} = x^k \pm \delta \xi^{k1}$.

Выбор длины шага выделяется в специальную процедуру, причем длина шага может изменяться в процессе поиска в соответствии с выражением

$$x^{k+1} = x^k \pm \alpha_k \xi^{k1},$$

где $\alpha_k > \delta$.

Таким образом, выбор направления поиска с помощью простейшего алгоритма можно представить в виде

$$\eta^k = \begin{cases} \xi^{k1}, & \text{если } F(x^k + \delta \xi^{k1}) \leq F(x^k); \\ -\xi^{k1}, & \text{если } F(x^k + \delta \xi^{k1}) > F(x^k). \end{cases} \quad (5-2)$$

Такая тактика поиска носит название *линейной*, поскольку она предполагает, что изменение критерия оптимальности в окрестности оптимума может быть хорошо аппроксимировано линейной функцией.

Другой вариант алгоритма получается, если число проб в текущей точке поиска фиксировано и направление движения определяется «по наилучшей пробе». Пусть вектор ξ^k определяется из условия

$$F(x^k + \delta \xi^k) = \min_{1 \leq i \leq p} F(x^k + \delta \xi^{ki}).$$

Тогда

$$\eta = \begin{cases} \xi^k, & \text{если } F(x^k + \delta \xi^k) \leq F(x^k); \\ \text{не определено,} & \text{если } F(x^k + \delta \xi^k) > F(x^k). \end{cases} \quad (5-3)$$

Если η^k не определено при достаточно больших p , то поиск прекращается.

Величина p может возрастать с ростом номера итерации, т. е. $p = p(k)$. Основанием для такого подхода служит то, что вероятность отыскания хотя бы одного благоприятного направления поиска уменьшается по мере приближения к экстремуму.

Весьма эффективным методом поиска является поиск с так называемой *нелинейной* тактикой [46]. Примером алгоритма указанного типа может служить поиск с возвратом к предыдущей итерации в случае неудачной пробы ξ^{k1} , т. е.

$$\eta^k = \begin{cases} \xi^{k1}, & \text{если } F(x^k + \delta \xi^{k1}) \leq F(x^k); \\ -\xi^{k-1,1}, & \text{если } F(x^k + \delta \xi^{k1}) > F(x^k). \end{cases} \quad (5-4)$$

В общем случае все рассмотренные стохастические алгоритмы гарантируют асимптотическую сходимость к локальному экстремуму. Однако скорость сходимости существенно меняется не только для различных алгоритмов, но и для различных критериев оптимальности, начальных точек поиска и т. д.

Практически для каждого алгоритма можно привести примеры медленной сходимости даже при не очень сложных критериях оптимальности.

Обычно применение стохастических алгоритмов дает хорошие результаты лишь на начальных стадиях поиска (вдали от экстремума). Общая стратегия поиска обычно состоит в разумном переборе стохастических алгоритмов в процессе счета на ЭВМ с целью использования наи-

более эффективного из них на каждом этапе поиска. Поэтому возникает необходимость сравнения алгоритмов в процессе поиска.

Для оценки эффективности стохастических методов поиска часто используется локальный критерий (т. е. критерий в точке x^k) относительного приближения к экстремуму в расчете на одну пробу. Эта величина измеряется отношением длины шага, полученного с помощью одномерного поиска вдоль выбранного направления, к числу проб, определяющих это направление:

$$\rho_k = h_k / p_k, \quad (5-5)$$

где h_k — длина шага на k -й итерации и p_k — число пробных точек. Если p_k становится меньше некоторой наперед заданной величины ϵ , производится переход к другому стохастическому методу.

Однако эта оценка может давать формально удовлетворительные результаты даже при весьма медленном продвижении к оптимуму. Так, при малом числе проб могут производиться большие шаги по направлениям, в которых функция убывает очень медленно. Указанная ситуация может возникнуть при движении в направлении, почти ортогональном к градиенту, если линии уровня критерия оптимальности близки к прямым (поиск в линейном поле).

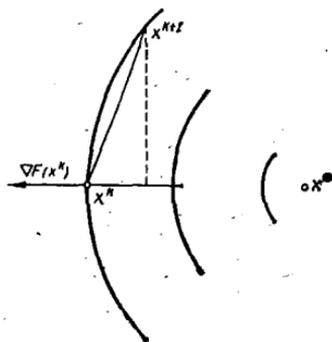


Рис. 5-1. К определению направления к локальному экстремуму.

Другой локальной оценкой эффективности метода служит величина проекции вектора $x^{k+1} - x^k$ на направление градиента

$$\pi_k = \frac{\nabla F(x^k)^T (x^{k+1} - x^k)}{\sqrt{\nabla F(x^k)^T \nabla F(x^k)}}. \quad (5-6)$$

Она позволяет оценить, насколько быстро убывает критерий оптимальности в направлении поиска. Однако и эта оценка не всегда отражает фактическую скорость продвижения к локальному оптимуму. Так, на рис. 5-1

приводится пример, когда переход из точки x^k в x^{k+1} фактически не приближает к экстремуму, хотя $\pi_k \neq 0$. Отсюда следует, что для оценки действительной эффективности стохастических алгоритмов следует производить усреднение выражений (5-5) и (5-6) для целого класса однотипных задач.

Рассмотренные алгоритмы поиска без адаптации просты, устойчивы по отношению к погрешности счета. Однако эти алгоритмы следует применять в тех случаях, когда критерий оптимальности имеет сравнительно простой вид и невелики затраты машинного времени на его вычисление.

Дело в том, что в этих алгоритмах вычисления на предыдущих шагах поиска значения критерия оптимальности не используются для повышения эффективности поиска, например для оценки длины шага или выбора направления на следующей итерации. Иными словами, информация об изменении критерия оптимальности, накопленная на k -й итерации, совершенно не используется при дальнейшем поиске. Более эффективное использование информации предполагает адаптацию, заключающуюся в данном случае в предпочтении тех направлений, которые на предыдущих итерациях позволяли улучшить значение критерия оптимальности.

Адаптация при выборе направления гиперконуса поиска. В процедуре случайного поиска без адаптации используются пробные точки, расположенные на поверхности n -мерной гиперсферы в точке x^k . Выбор точек производится на основе равномерного случайного распределения, а радиус сферы определяет шаг поиска. Процесс адаптации можно представить как выделение на поверхности указанной гиперсферы (или внутри описанного ею шара) областей, в которых наиболее вероятно убывание критерия оптимальности. Внутри этих областей производится выбор пробных точек на основе равномерного распределения.

Одним из способов выделения указанных областей внутри n -мерного гипершара является формирование гиперконуса и изменение его параметров по результатам пробных вычислений [10]. Расположение пробных точек при поиске без адаптации и с адаптацией представлено на рис. 5-2. Поиск итерации x^{k+1} производится внутри n -мерного гиперконуса с вершиной в точке x^k и осью, проходящей через точки x^{k-1} и x^k .

Объем конуса определяется высотой h_k (соответствующей радиусу гиперсферы при поиске без адаптации) и половиной угла при вершине θ_k . Обе величины могут изменяться либо на каждом шаге либо после заданного числа шагов. Алгоритмы изменения параметров h_k и θ_k должны быть определены до начала вычислений. Например, при успешном продвижении к оптимуму h_k сле-

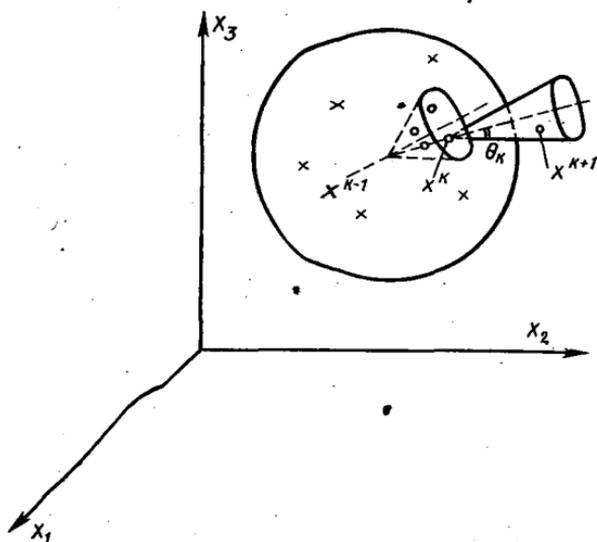


Рис. 5-2. Расположение пробных точек при поиске без адаптации (X) и с самообучением (O).

дует увеличить, а θ_k уменьшить. Простейшим, но достаточно эффективным алгоритмом для этого могут служить линейные зависимости

$$\begin{aligned} h_{k+1} &= \kappa h_k, \quad \kappa > 1; \\ \theta_{k+1} &= \lambda \theta_k, \quad \lambda < 1. \end{aligned} \quad (5-7)$$

На практике могут быть выбраны значения коэффициентов пропорциональности: $\kappa \approx 1,05$; $\lambda \approx 0,95$.

Аналогично при неудачных пробах конус следует расширять, а его высоту уменьшать. Соответствующие линейные зависимости имеют вид:

$$\begin{aligned} h_{k+1} &= \hat{\kappa} h_k, \quad \hat{\kappa} < 1; \\ \theta_{k+1} &= \hat{\lambda} \theta_k, \quad \hat{\lambda} > 1. \end{aligned} \quad (5-8)$$

В общем случае зависимости для h_k и θ_k могут быть представлены функциями:

$$h_{k+1} = r(h_k), \quad \theta_{k+1} = s(\theta_k)$$

после определенного числа удачных проб и

$$h_{k+1} = \hat{r}(h_k), \quad \theta_{k+1} = \hat{s}(\theta_k)$$

после неудачных.

Прямые стохастические методы оказываются весьма полезными на всех стадиях поиска оптимума. При этом на начальных этапах, как правило, применяются методы поиска без адаптации, позволяющие посредством небольшого числа пробных вычислений продвинуться к оптимуму. Затем по мере усложнения геометрии области поиска используется адаптация, а также специальные, «овражные» методы поиска.

5-3. «ОВРАЖНЫЕ» МЕТОДЫ ПОИСКА

В ряде случаев эффективным оказывается локальное моделирование геометрии области поиска вблизи оптимума с помощью достаточно простых математических зависимостей. Такой подход используется, как правило, когда поверхности постоянного уровня критерия оптимальности сильно вытянуты в одних направлениях и сжаты в других (см., например, рис. 1-5). Это означает, что кривизна функции $F(x)$ существенно отличается для различных направлений, т. е. в терминологии [15] имеется «овраг».

Направления малой кривизны соответствуют движению вдоль «оврага», в то время как направления с большой кривизной соответствуют движению поперек «оврага».

Кривизна функции $F(x)$ в заданном направлении $d = (d_1, d_2, \dots, d_n)$ определяется формулой

$$K^2 = \{d^T \Gamma(x) d\}^2 / \{1 + [g(x)^T d]^2\}^3, \quad (5-9)$$

где $\Gamma(x)$ — гессиан в точке x .

Можно считать, что относительно малые собственные значения соответствуют собственным векторам, которые направлены вдоль «оврага», а собственные значения относительно большой величины соответствуют собственным векторам, направленным поперек «оврага».

Кривая, вдоль которой вытянуты линии уровня, называется «дном оврага». Если линию «дна оврага» смещать параллельно самой себе, она попадает на «склон оврага».

«Дно оврага» для функции (1-17) представляет собой кривую второго порядка. «Дно оврага» тестовой функции (1-18) образует спираль. «Овражная» ситуация может иметь место и для критериев оптимальности, зависящих от большего числа конструктивных параметров ($n \geq 3$), однако она уже лишена столь наглядной геометрической интерпретации.

Если направление поиска не совпадает с направлением «дна оврага», шаг оказывается очень малым ввиду значительного изменения из-за функции F при смещении поперек «оврага». Часто направления движения траектории поиска становятся зигзагообразными, что затрудняет продвижение к оптимуму. В то же время направление вдоль «оврага» может быть получено из точки, находящейся на «дне оврага». Этим объясняется неэффективность градиентных методов в «овражной» ситуации. Поэтому при решении многих практических задач в «овражной» ситуации используются эмпирические «овражные» алгоритмы.

«Овражные» алгоритмы предусматривают два этапа. На первом производится поиск из близлежащих начальных точек с помощью простейших стохастических методов. Полученные по простейшим алгоритмам «оптимальные» точки расположены, как правило, по «дну оврага». На втором этапе организуется движение вдоль «дна оврага» с помощью аппроксимации геометрии «дна».

Наиболее разработанными и достаточно наглядными являются алгоритмы, в основе которых лежит предположение, что «дно» имеет форму некоторой кривой. Поиск оптимума производится путем аппроксимации «дна оврага» линейной или квадратичной зависимостью.

Рассмотрим линейную аппроксимацию. Пусть y^0 и y^1 — две точки на «дне оврага», причем $F(y^1) < F(y^2)$. Тогда полагают:

$$\tilde{y}^2 = y^1 + \frac{y^1 - y^0}{|y^1 - y^0|} h,$$

где h — положительная величина, называемая «овражным» шагом.

Точка \tilde{y}^2 находится уже на «склоне оврага», поэтому из нее осуществляется новый спуск на «дно оврага»

с помощью простейших стохастических методов. Результатом этого поиска является точка y^2 , которая так же, как и точки y^0 и y^1 , находится на «дне оврага».

Если уже определены точки y^0, y^1, \dots, y^k , расположенные на «дне оврага», то следующая точка задается формулой

$$\tilde{y}^{k+1} = y^k + \frac{y^k - y^{k-1}}{|y^k - y^{k-1}|} h. \quad (5-10)$$

Она расположена на продолжении отрезка, проходящего через y^k и y^{k-1} . Процедура поиска по указанному алгоритму показана на рис. 5-3,а.

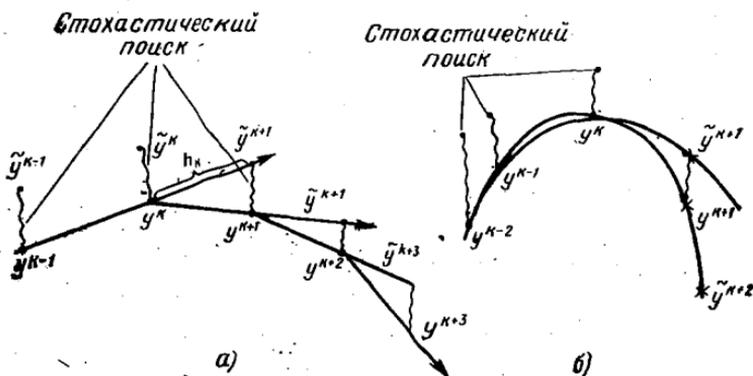


Рис. 5-3. Аппроксимация «дна оврага» линейной (а) и квадратичной (б) зависимостями.

При использовании квадратичной интерполяции итерация \tilde{y}^{k+1} определяется по трем точкам (y^k, y^{k-1}, y^{k-2}) на «дне оврага» (рис. 5-3,б).

Особую роль в процедуре «овражного» поиска играет выбор длины «овражного» шага h . Большой шаг может затруднить поиск, так как не позволяет следить за изгибами «дна оврага»; с малым шагом связаны значительные затраты времени счета на ЭВМ. Поэтому при выборе длины шага, так же как и направления движения, осуществляется адаптация в процессе поиска.

Обычно эффективным оказывается определение длины шага вдоль «дна оврага» по формуле

$$h_{k+1} = K C^{\cos \alpha_k}, \quad (5-11)$$

где α_k — угол между векторами $\tilde{y}^k - y^{k-1}$ и $y^k - y^{k-1}$, а значения параметров $C > 1$ и K выбираются на основе

предыдущего опыта поиска угла α_k между n -мерными векторами $\tilde{y}^k - y^{k-1}$ и $y^k - y^{k-1}$ согласно выражению

$$\cos \alpha = \frac{(\tilde{y}^k - y^{k-1})^T (y^k - y^{k-1})}{|\tilde{y}^k - y^{k-1}| |y^k - y^{k-1}|}.$$

Прямолинейному «дну оврага» соответствует величина $\cos \alpha_k$, близкая к единице. В этом случае может быть выбран достаточно большой «овражный» шаг h_k , что и отражает формула (5-11). Напротив, при искривлении «дна оврага» необходим малый шаг. Соответственно уменьшается величина h_k в выражении (5-11).

Принципиально иной способ движения в «овражной» ситуации предложен в [92]. Он основан на модификации методов переменной метрики в «овражной» ситуации и предполагает вычисление собственных значений гесссиана или аппроксимирующей его матрицы.

Согласно методу Ньютона (см. § 4-3), если известен гесссиан и градиент в точке x^k , переход к итерации x^{k+1} производится по формуле

$$x^{k+1} = x^k - t \Gamma^{-1} g,$$

где t определяется с помощью методов однопараметрического поиска.

Предположим, для простоты, что матрица Γ диагональная. Этого можно достигнуть за счет выбора соответствующей системы координат. Тогда i -й элемент вектора $\Gamma^{-1}g$ имеет вид:

$$\frac{\partial F}{\partial x_i} \bigg/ \frac{\partial^2 F}{\partial x_i^2}.$$

Поскольку диагональные элементы матрицы Γ при «овражной» ситуации могут быть очень малы, координаты вектора $\Gamma^{-1}g$ становятся недопустимо большими. В результате за счет изменения лишь одного коэффициента t не удастся получить достаточно эффективный шаг поиска.

Чтобы избежать этого недостатка, вводится ограничение для малых значений вторых производных. Если гесссиан Γ является диагональной матрицей с элементами λ_i , $i=1, 2, \dots, n$ на диагонали, он заменяется диагональной матрицей $\tilde{\Gamma}$, элементы $\tilde{\lambda}_i$ которой определя-

ются формулой

$$\tilde{\lambda}_i = \max\{\lambda_i, \varepsilon\}, \quad i = 1, 2, \dots, n, \quad (5-12)$$

где ε — заданная, достаточно малая положительная величина.

В общем случае, когда Γ не является диагональной матрицей, существует линейное ортогональное преобразование T такое, что матрица

$$\Gamma^* = T^{-1}\Gamma T \quad (5-13)$$

диагональна.

Поэтому в общем случае «овражный» алгоритм предусматривает: а) вычисление гессиана Γ , нахождение и применение ортогонального преобразования T по формуле (5-13); б) замену преобразованной диагональной матрицы Γ^* на $\tilde{\Gamma}^*$ по формуле (5-12); в) применение обратного преобразования T^{-1} . Выбирая величину ε в зависимости от класса решаемых задач, удается осуществить продвижение к оптимуму вдоль траектории, лежащей на «дне оврага».

При использовании алгоритмов переменной матрицы матрица заменяется матрицей H_k^{-1} , вычисляемой по формуле (4-27).

Рассмотренные методы поиска в «овражной ситуации» используются при решении большинства практических задач оптимального проектирования. Они применяются в сочетании как с методом штрафных функций, так и с методами возможных направлений (см. гл. 6).

5.4. МЕТОДЫ СТОХАСТИЧЕСКОЙ ОЦЕНКИ ГРАДИЕНТА

Стохастический подход используется не только для выбора направления движения к оптимуму, но в ряде случаев также и для оценки градиента критерия оптимальности. Оценка градиента производится на основе выборочных значений критерия оптимальности, вычисленных в окрестности текущей итерации. Она может быть использована в градиентных методах поиска.

В некоторых задачах, особенно на начальных этапах поиска оптимума, небольшого числа проб бывает достаточно для приблизительного определения направления движения к экстремуму. Для грубой оценки градиента не требуется указанных в предыдущей главе $n+1$ пробных точек.

Чтобы оценить точность определения градиента в зависимости от числа пробных точек (объема выборки) и численного метода, оценка $\widehat{\nabla}F^{(k)}$ рассматривается как случайная величина, полученная на основе k наблюдений. Ее свойства характеризуются функцией

распределения и, в частности, величинами математического ожидания и дисперсии.

Для анализа возможности стохастических оценок градиента критерия оптимальности детерминированное выражение для $F(x)$ заменяется его статистическим аналогом. Пусть разложение функции $F(x)$ в ряд Тейлора имеет вид:

$$F(x) = F(x^0) + (x - x^0)^T \nabla F(x^0) + O(|x|^2),$$

где $O(|x|^2)$ включает в себя члены второго и более высоких порядков. Поскольку при дальнейших вычислениях производится оценка градиента по значениям функции $F(x)$, аддитивная составляющая $O(|x|^2)$ вносит в расчеты неконтролируемые помехи.

Относительно этих помех вводится следующее предположение: $O(|x|^2)$ — случайная величина, обладающая нормальным распределением с нулевым математическим ожиданием и дисперсией σ^2 . Стохастические методы оценки градиента применяются до тех пор, пока они приводят к улучшению критерия оптимальности. Такое улучшение имеет место на начальных стадиях поиска, где функцию можно считать приближенно линейной в окрестности точки x (поиск в линейном поле [46]).

Оценка $\widehat{\Delta F}^{(k)}$ градиента ∇F называется несмещенной, если при всех k математическое ожидание $\widehat{\nabla F}^{(k)}$ равно градиенту, т. е. $M\{\widehat{\nabla F}^{(k)}\} = \nabla F$. Очевидно, что только такие оценки представляют интерес для выбора направления движения к экстремуму.

Наилучшая несмещенная оценка градиента должна обладать следующим свойством:

$$D\{\widehat{\nabla F}^{(k)}\} = M\{(\widehat{\nabla F}^{(k)} - \nabla F)^2\} = \min, \quad (5-14)$$

т. е. дисперсия этой оценки минимальна по сравнению со всеми остальными несмещенными оценками. Такая оценка называется *эффективной*.

Примером эффективной оценки может служить вычисление градиента критерия оптимальности по $2n$ пробным точкам:

$$\widehat{\nabla F}^{(2n)} = \sum e_i [F(x + he_i) - F(x - he_i)], \quad (5-15)$$

где $e_i, i=1, 2, \dots, n$ — координатные орты. Указанная оценка является фактически разностной оценкой градиента по формуле (4-15). Ее дисперсия

$$D\{\widehat{\nabla F}^{(2n)}\} = \frac{\sigma^2}{2h^2}, \quad (5-16)$$

т. е. не зависит от размерности пространства параметров, а только от дисперсии случайной составляющей и значения пробного шага.

Обобщение этой оценки на случай произвольного числа $m \leq 2n$ наблюдений приводит к статистическому градиенту, определяемому по формуле

$$\widehat{\nabla F}^{(2m)} = \frac{n}{2mh} \sum e_i [F(x) + h\Xi_i] - F(x - h\Xi_i)], \quad (5-17)$$

где $\Xi_i, i=1, 2, \dots, n$ — независимые случайные векторы единичной длины. Его дисперсия

$$D\{\widehat{\nabla F^{(2m)}}\} = \frac{1}{m} \left(2\nabla F^T \nabla F \frac{n-1}{n-2} + \frac{n\sigma^2}{2h^2} \right). \quad (5-18)$$

Более точная оценка получается, если предположить, что Ξ_i являются случайными единичными векторами, обладающими свойством попарной ортогональности.

Дисперсия полученной таким способом оценки

$$D\{\widehat{\nabla_0 F^{(2m)}}\} = \frac{1}{m} \left[\frac{2(n-m)}{m(n+2)} + \frac{n\sigma^2}{2h^2} \right] \quad (5-19)$$

и при $m=n$ совпадает с (5-16).

Аналогично тому, как построено статистическое обобщение разностной оценки градиента до $2n$ пробам, можно построить такое же обобщение для градиента, вычисленного по $n+1$ пробе. Вычисление градиента по разностной формуле (4-14) может рассматриваться как статистическая оценка. Ее дисперсия

$$D\{\widehat{\nabla F^{(n+1)}}\} = \frac{2\sigma^2}{h^2} \quad (5-20)$$

и для случая $2n$ пробных точек также не зависит от размерности пространства конструктивных параметров. Стохастическим обобщением этой оценки для m пробных точек служит формула

$$\widehat{\nabla F^{(m)}} = \frac{n}{mh} \sum \Xi [F(x + h\Xi_i) - F(x)], \quad (5-21)$$

где $\Xi_i, i=1, 2, \dots, m$, линейно независимые векторы единичной длины.

Таким образом, в основе стохастических методов оценки градиента лежит расчет значений критерия оптимальности в пробных точках, выбранных случайным образом на поверхности n -мерной гиперсферы радиуса h . Они обобщают детерминированные методы вычисления градиента и могут быть использованы при поиске оптимума градиентными методами (см. § 4-2). При этом статистические оценки градиента должны производиться тем точнее, чем ближе к оптимальному решению расположены текущие итерации поиска.

5-5. ПРИМЕНЕНИЕ МЕТОДОВ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

При одном и том же числе пробных точек точность оценки градиента меняется в зависимости от их расположения. В результате возникает проблема определения такого расположения (фиксированного числа) пробных точек, которое позволяет получить наиболее точную оценку. Для ее решения могут быть использованы методы планирования эксперимента.

Методы планирования эксперимента первоначально были разработаны для планирования проводимых в условиях действия помех экспериментов по определению значений управляющих воздействий, обеспечивающих оптимизацию технологического процесса.

Вычисление с помощью ЭВМ значения критерия оптимальности в некоторой точке пространства параметров на основе модели проектируемого объекта может также рассматриваться как некоторый (численный) эксперимент. Если модель проектируемого объекта довольно сложна, этот эксперимент содержит заметную случайную ошибку, возникающую как вследствие использования приближенных численных методов, так и ввиду ограниченности разрядной сети ЭВМ.

В отличие от статистических методов, которые решают вопрос о числе пробных точек, достаточном для оценки градиента с заданной точностью, методы планирования эксперимента предназначены также и для определения расположения пробных точек.

Основное предположение, используемое в теории планирования эксперимента, состоит в том, что критерий оптимальности может быть представлен с помощью линейной модели с аддитивным случайным компонентом. Иными словами,

$$F(x) = a_0 + a^T x + \varepsilon, \quad (5-22)$$

где a_0 — скалярная величина; a — n -мерный вектор коэффициентов и ε — случайная величина, обладающая нулевым математическим ожиданием и дисперсией σ^2 .

В рассматриваемом случае целью планирования эксперимента является построение гиперплоскости, наилучшим образом аппроксимирующей критерий оптимальности к окрестности некоторой точки поиска (текущей итерации). Вектор a , определяющий эту гиперплоскость, принимается в качестве приближенного значения градиента.

Для иллюстрации особенностей применения метода планирования эксперимента рассмотрим задачу определения градиента критерия оптимальности, зависящего от трех переменных: $F(x) \equiv F(x_1, x_2, x_3)$. В этом случае для расчета градиента по формуле (5-17) вычисления производятся в точках:

$$x^0 = (x^0_1, x^0_2, x^0_3), \quad x^1 = (x^1_1, x^1_2, x^1_3),$$

$$x^2 = (x^2_1, x^2_2, x^2_3), \quad x^3 = (x^3_1, x^3_2, x^3_3).$$

Значения критерия в этих точках обозначим соответственно F_0, F_1, F_2, F_3 . Дисперсия результата при использовании модели (5-22)

$$\sigma^2(a_i) = \sigma^2 \left\{ \frac{F_i - F_0}{h} \right\} = \frac{2\sigma^2}{h^2}. \quad (5-23)$$

Такая же оценка дисперсии статистического градиента получается в результате применения формулы (5-20).

Однако существует другой способ вычисления коэффициентов в выражении (5-22), причем полученные оценки обладают меньшей дисперсией.

Заменим вычисление функции в точке x^0 вычислением в точке $x^4 = (x^0_1 + h, x^0_2 + h, x^0_3 + h)$. Формулы для вычисления коэффициентов в этом случае следующие:

$$a_1 = \frac{F_1 - F_2 - F_3 + F_4}{2h}; \quad a_2 = \frac{-F_1 + F_2 - F_3 + F_4}{2h};$$

$$a_3 = \frac{-F_1 - F_2 + F_3 + F_4}{2h}. \quad (5-24)$$

Дисперсия, связанная со случайным компонентом ε , при новой схеме расчета составляет:

$$\sigma^2(a_1) = \sigma^2 \left\{ \frac{-F_1 - F_2 - F_3 + F_4}{2h} \right\} = \frac{4\sigma^2}{4h^2} = \frac{\sigma^2}{h^2}.$$

Аналогичным образом получаем, что $\sigma^2(a_2) = \sigma^2(a_3) = \sigma^2/h^2$, т. е. дисперсия вдвое меньше, чем в предыдущем случае, при том же числе наблюдений. Отличие в расчетах состоит в том, что по новой схеме все пробные точки участвуют в расчете каждого коэффициента линейной модели (5-22).

Если рассматривать вектор $a^T = (a_1, a_2, a_3)$ как градиент функции $F(x^1, x^2, x^3)$, то оказывается, что его координатами являются

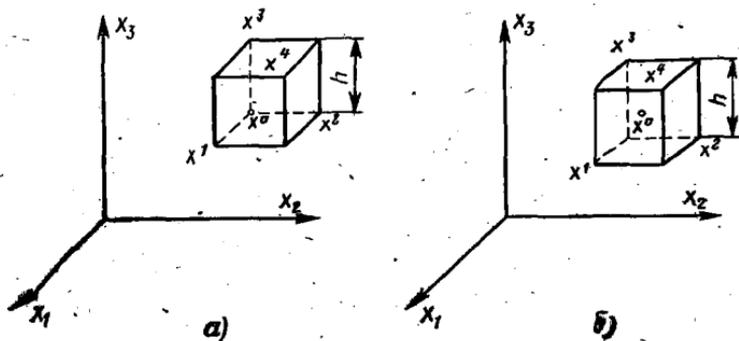


Рис. 5-4. Расположение пробных точек при оценке градиента методом планирования эксперимента.

x_1, x_2, x_3 — точки x^i находятся в вершинах куба; x_0 — точка x^0 находится в центре куба.

частные производные функции F , вычисленные не в исходной точке x^0 , а в промежуточных, лежащих внутри куба с ребром h (рис. 5-4,а). Для того чтобы эти промежуточные точки в дальнейшем были расположены в окрестности точки x^0 , пробные шаги смещаются таким образом, чтобы центр куба совпадал с этой точкой (рис. 5-4,б). При этом получают удобные формы для представления пробных точек (план эксперимента).

Предполагается, что каждая переменная принимает два значения: на верхнем и нижнем уровнях, обозначаемые соответственно $+1$ и -1 . Тогда таблица экспериментов для трех переменных принимает вид:

Таблица 5-1

x_0	a_1	a_2	a_3	F
	x_1	x_2	x_3	
$+1$	$+1$	-1	-1	F_1
$+1$	-1	$+1$	-1	F_2
$+1$	-1	-1	$+1$	F_3
$+1$	$+1$	$+1$	$+1$	F_4

и формулы (5-32) получаются простым умножением последнего столбца на столбец, соответствующий вычисляемому коэффициенту¹.

Если рассматривается, например, модель (5-22), в которой критерий оптимальности зависит от семи параметров, для определения ее коэффициентов достаточно восьми экспериментов. Соответствующий план экспериментов приводится в табл. 5-2.

Таблица 5-2

x_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	F
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	
+1	-1	-1	-1	+1	+1	+1	-1	F_1
+1	+1	-1	-1	-1	-1	+1	+1	F_2
+1	-1	+1	-1	-1	+1	-1	+1	F_3
+1	+1	+1	-1	+1	-1	-1	-1	F_4
+1	-1	-1	+1	+1	-1	-1	+1	F_5
+1	+1	-1	+1	-1	+1	-1	-1	F_6
+1	-1	+1	+1	-1	-1	+1	-1	F_7
+1	+1	+1	+1	+1	+1	+1	+1	F_8

Рассмотрим общие свойства, которым должны удовлетворять планы экспериментов для определения градиента. Пусть производится N экспериментов для оценки n координат вектора a . Обозначив через x_{ij} элемент i -го столбца и j -й строки в таблице, аналогичной табл. 5-1 и 5-2 для N экспериментов, получим, что планы должны обладать следующими свойствами:

а) симметричность:
$$\sum_{j=1}^N x_{ij} = 0, \quad i = 1, 2, \dots, n;$$

б) ортогональность:
$$\sum_{j=1}^N x_{ij} x_{kj} = 0, \quad i = 1, 2, \dots, n, \quad k = 1, 2, \dots, n;$$

в) нормировка:
$$\sum_{j=1}^n x_{ij}^2 = N, \quad i = 1, 2, \dots, n.$$

В этом случае коэффициенты a_1, a_2, \dots, a_n и их дисперсии вычисляются с помощью достаточно простых формул:

$$a_i = \frac{\sum_{j=1}^N x_{ij} F_j}{N}; \quad \sigma^2(a_i) = \frac{\sigma^2}{N}, \quad i = 1, 2, \dots, n.$$

¹ Переменная x_0 носит формальный характер. Она используется при проверке адекватности модели и при независимой оценке свободного члена a_0 .

Заметим, что в рассмотренных выше примерах справедливо соотношение $N=n+1$. Планы, удовлетворяющие условиям пп. а—в и содержащие $n+1$ эксперимент, называются *насыщенными*. Однако насыщенные планы существуют лишь для векторных пространств, для которых $n+1$ кратно четырем. В остальных случаях следует либо использовать другие методы оценки градиента, либо увеличивать число пробных точек. Использование методов планирования эксперимента дает близкие к насыщенным оптимальные симплекс-планы [4].

Планирование эксперимента позволяет путем дополнительных расчетов критерия оптимальности в одной или нескольких пробных точках проверить гипотезу о допустимости линейной модели (5-22) на рассматриваемом участке. Обычно для этой цели используется эксперимент в центре гиперкуба (рис. 5-4б), полученный в результате одномерного поиска экстремума на предыдущей итерации.

Метод планирования эксперимента позволяет также сократить объем вычислений за счет отсеивания тех параметров, изменение которых не дает существенного критерия оптимальности. С этой целью производятся так называемые отсеивающие эксперименты. Если значение критерия оптимальности мало меняется при изменении отдельных параметров на уровнях ± 1 или некоторые координаты вектора \hat{y} близки к нулю, то соответствующие конструктивные параметры временно исключаются из дальнейшего рассмотрения. В течение нескольких последующих итераций им присваивается одно и то же постоянное значение. В результате уменьшается размерность пространства поиска, что ускоряет движение к оптимуму. При этом после определенного числа шагов необходимо проверить, следует ли снова расширить модель введением ранее исключенных переменных.

В заключение отметим, что конструктивные параметры x_i , принятые за независимые переменные, являются различными несоизмеримыми физическими величинами. При использовании оценок градиента статистическими методами или методами планирования эксперимента на скорость сходимости существенно влияет выбор единиц измерения конструктивных параметров. Посредством удачного масштабирования (выбора единиц измерения) параметров иногда удается повысить скорость продвижения к оптимуму.

Глава шестая

МЕТОДЫ ВОЗМОЖНЫХ НАПРАВЛЕНИЙ

В теории оптимального проектирования значительное внимание уделяется методам, в которых при выборе направления поиска в явном виде учитываются ограничения, определяющие допустимую область изменения независимых переменных. Движение к оптимуму в ограниченной области производится с помощью методов возможных направлений. Они позволяют выбрать в фиксированной (неоптимальной) точке x допустимой области такое направление, вдоль которого существует отрезок, целиком состоящий из точек области G с лучшими, чем в точке x , значениями критерия оптимальности.

В этой главе рассматриваются методы определения возможных направлений, выбора из них наиболее предпочтительных по тем или иным признакам, а также вычисления длины шага поиска для перехода к следующей итерации.

6-1. ФОРМИРОВАНИЕ НАПРАВЛЕНИЙ ПОИСКА В ДОПУСТИМОЙ ОБЛАСТИ

Определение возможных направлений. Напомним, что общая задача оптимального проектирования заключается в поиске значений конструктивных параметров $x = (x_1, x_2, \dots, x_n)$, при которых достигается минимальное (максимальное) значение критерия оптимальности $F(x)$ при ограничениях

$$R_i(x) \leq 0, \quad i=1, 2, \dots, m \quad (6-1)$$

и

$$x_j \geq 0, \quad j=1, 2, \dots, n. \quad (6-2)$$

Предполагается, что критерий оптимальности $F(x)$ и функции ограничений достаточно гладкие, т. е. в каждой точке допустимой области G можно вычислить их градиенты.

Задача выбора в точке x^h области G направления поиска s^h , улучшающего значения критерия оптимальности и приводящего к новому допустимому вектору параметров x^{h+1} , для которого $F(x^{h+1}) < F(x^h)$ является основным (и наиболее трудоемким) этапом поиска в ограниченной области.

Для того чтобы выделить те направления, которые позволяют приблизиться к оптимуму, введем следующие определения.

Направление s^h называется *возможным* в точке $x^h \in G$, если существует такое $\lambda > 0$, что отрезок $[x^h, x^h + \lambda s^h]$ также принадлежит области G . Возможное направление называется *подходящим*, если $\nabla F(x^h)^T s^h < 0$, т. е. критерий оптимальности убывает в направлении s^h .

Методы возможных направлений при решении задач оптимального проектирования учитывают ограничения в явном виде. Эти методы реализуют следующую схему вычислений:

1) в качестве начальной точки x^0 выбирается произвольная точка допустимой области G ;

2) среди возможных направлений определяется подходящее направление s^k ;

3) для подходящего направления определяется значение шага λ_k , не выводящего за пределы допустимой области и обеспечивающего выполнение условия $F(x^k + \lambda_k s^k) < F(x^k)$. Точка $x^{k+1} = x^k + \lambda_k s^k$ принимается в качестве следующей итерации;

4) процесс повторяется до тех пор, пока не удовлетворится некоторый критерий прекращения поиска. Таким критерием может служить либо разность значений критерия оптимальности в двух последовательных точках поиска, либо значение шага λ_k .

Выбор подходящего направления представляет наибольшие трудности при выходе на границу допустимой области. Граница может иметь участки, состоящие из одной гиперповерхности или пересечения двух или нескольких гиперповерхностей.

Если при подстановке точки x^k в (6-1) и (6-2) эти ограничения обращаются в равенство, то это означает, что точка x^k расположена на пересечении l ограничивающих гиперповерхностей. В дальнейшем именно эти гиперповерхности оказывают влияние на выбор допустимого направления.

Обозначим символом $I(k)$ множество индексов тех ограничений из (6-1), которые выполняются в точке x^k как строгие равенства, т. е.

$$I(k) = \{i | R_i(x^k) = 0, i = 1, 2, \dots, m\} \quad (6-3)$$

Символом $J(k)$ будем обозначать множество индексов, соответствующих нулевым значениям проектируемых параметров в точке x^k , т. е. тех ограничений вида (6-2), которые выполняются как строгие равенства

$$J(k) = \{j | x_j^k = 0, j = 1, 2, \dots, n\}. \quad (6-4)$$

Согласно методу возможных направлений точка $x^{k+1} = x^k + \lambda_k s^k$ лежит внутри допустимой области. Поэтому условия на выбор вектора s формируются, исходя из того, что должны выполняться ограничения с индексами из множеств $I(k)$ и $J(k)$:

$$R_i(x^{k+1}) = R_i(x^k + \lambda_k s) \approx R_i(x^k) + \lambda_k s^T \nabla R_i \leq 0$$

для $i \in I(k)$;

$$x^{k+1} = x^k + \lambda_k s \geq 0 \quad \text{для } j \in J(k).$$

Отсюда получаем следующие ограничения на выбор направления s :

$$s^T \nabla R_i < 0 \quad \text{для } i \in I(k); \quad (6-5)$$

$$-s_j \leq 0 \quad \text{для } j \in J(k). \quad (6-6)$$

Формирование подходящего направления поиска (т. е. направления, позволяющего улучшить значения критерия оптимальности без выхода за пределы допустимой области) осуществляется путем проектирования некоторого направления N , вдоль которого происходит уменьшение значений критерия оптимальности $F(x)$, на многогранник, определяемый ограничениями (6-5) и (6-6).

Проекцией точки x векторного пространства E^n на замкнутое множество $G \subseteq E^n$ называется точка $P_G(x)$, удовлетворяющая условию

$$|x - P_G(x)| = \inf_{y \in G} |x - y|.$$

Если $G = E^n$, то $P_G(x) = x$. В остальных случаях проекция точки x зависит от выбора множества G и способа введения меры (метрики) в пространстве E^n . В частности, если

$$|x| = \sqrt{\sum_{j=1}^n x_j^2}$$

— обычная метрика евклидова пространства E^n , то оператор P осуществляет ортогональное проектирование.

Таким образом, очередная итерация поиска локального оптимума может быть представлена формулой

$$x^{k+1} = x^k + \lambda_k P[N_F(x^k)]. \quad (6-7)$$

В зависимости от выбора оператора N может осуществляться построение проекции для градиента, направлений переменной метрики стохастического градиента и т. д. Поскольку процедура проектирования направления N на область, определяемую ограничениями (6-5) и (6-6), а также определение длины шага λ_k являются довольно сложными, выбор направления N должен быть достаточно обоснован. В большинстве рассматриваемых

в этом параграфе методов в качестве направления N выбирается направление антиградиента критерия оптимальности $F(x)$.

После того, как направление N выбрано, построение его проекции на допустимую область G формулируется как задача оптимизации:

$$\min_{x \in G} |x - N_F(x^k)|.$$

Эта задача эквивалентна нахождению некоторого вектора s единичной длины, т. е. $s^T s = 1$, доставляющего минимум выражению $s^T N_F(x^k)$ при ограничениях (6-5) и (6-6). Она решается методами квадратичного программирования [1]. Если $N_F(x^k) = -\nabla F(x^k)$, то вектор s соответствует направлению наиболее быстрого убывания критерия оптимальности $F(x)$ в точке x^k среди всех подходящих направлений.

Однако решение задачи квадратичного программирования на каждой итерации приводит к значительным затратам машинного времени, а полученное с его помощью направление поиска оптимально лишь в небольшой окрестности точки x^k . Поэтому применение методов квадратичного программирования может быть эффективно лишь вблизи оптимального решения, когда поиск должен осуществляться с высокой точностью.

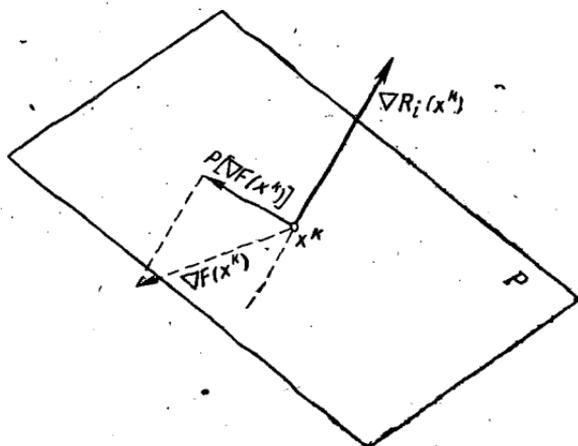


Рис. 6-1. Построение проекции градиента на гиперплоскость, касательную в точке x^k к ограничивающей поверхности $R_i(x) = 0$.

Если направление движения к оптимуму N_F , вычисленное в точке x^k без учета влияния ограничений, таково, что для него выполняются условия (6-5) и (6-6), то направление N_F является подходящим. Необходимость построения проекции вектора N_F возникает, если эти условия в точке x^k не выполняются.

Ограничение $R_i(x) \leq 0$ ($x_j \geq 0$) называется существенным в точке x^k , если $i \in I(k)$ ($j \in J(k)$) и $N_F^T \nabla R_i(x^k) > 0$ ($N_F^T e_j > 0$), если e_j — j -й единичный координатный вектор.

Очевидно, что именно существенные ограничения определяют необходимость изменения направления движения к оптимуму. Так, представленная на рис. 6-1 гиперплоскость, касательная в точке x^k ограничивающей поверхности $R_i(x) = 0$, препятствует движению по направлению антиградиента $N_F = -\nabla F(x^k)$. В этом случае движение производится в плоскости P , ортогональной направлению ∇R_i в рассматриваемой точке.

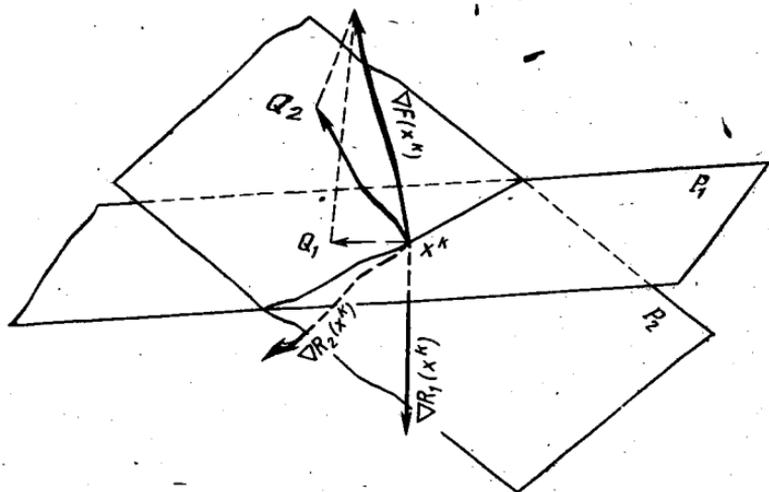


Рис. 6-2. Выбор направления движения при нескольких существенных ограничениях.

Выбор направления движения при наличии нескольких существенных ограничений значительно сложнее (рис. 6-2). Задача состоит в определении такого набора существенных ограничений, чтобы проекция вектора $N_F = -\nabla F(x)$ на их пересечение давала подходящее направление.

Ввиду того, что проектирование направления N_F на допустимую область G сопряжено с значительными трудностями, при решении практических задач используются проекции не на всю область G , а лишь на некоторые подмножества ограничивающих ее гиперповерхностей. Рассмотрим некоторые подходы к построению таких проекций и более подробно остановимся на проекционном градиентном методе [84] и его модификации [20], предусматривающей адаптацию при поиске.

Еще один способ построения проекции направления поиска осуществляется в условиях постоянного изменения оператора проектирования в процессе поиска. Этот подход реализуется с помощью методов переменной метрики, модифицированных для ограниченной допустимой области.

6.2. ПРОЕКЦИОННЫЙ ГРАДИЕНТНЫЙ МЕТОД И ЕГО МОДИФИКАЦИИ

Проекционный градиентный метод. Наиболее простой подход реализован в проекционном градиентном методе, когда проектирование производится на пересечение всех ограничений, которые выполняются в точке x^k как равенства. Проверка, какие из ограничений являются несущественными, производится лишь в том случае, если нельзя продвинуться вдоль их пересечения. Тогда одно или несколько несущественных ограничений исключается на данном шаге.

Направления, выбранные с помощью проекционного градиентного метода, не всегда оказываются наилучшими из всех подходящих направлений, однако простое правило формирования проекций делает его удобным средством для решения практических задач оптимального проектирования.

Выполнение условий, положенных в основу определения множеств $I(k)$ и $J(k)$, приводит к следующим ограничениям на выбор направления:

$$a_i^T s = 0, \text{ где } a_i = \nabla R_i, i \in I(k); \quad (6-8)$$

$$s_j = 0, \text{ где } j \in J(k). \quad (6-9)$$

Геометрически эти ограничения соответствуют требованию ортогональности вектора s векторам $a_i, i \in I(k)$ и $e_j^T = (0, \dots, 0, 1, 0, \dots, 0), j \in J(k)$. Пусть далее множество векторов $a_i, i \in I(k)$ и $e_j, j \in J(k)$ являются линейно-

независимыми (выполнение этого условия достигается исключением линейно-зависимых векторов). Тогда они образуют базис некоторого векторного пространства M_k .

Вектор s^k определяется как проекция антиградиента на пространство, ортогональное к M_k . Чтобы вычислить эту проекцию, рассмотрим матрицу Q_k , отображающую пространство E^n на подпространство M_k :

$$Q_k = \begin{pmatrix} a_{i_1 1} & a_{i_1 2} & \dots & a_{i_1 n} \\ a_{i_2 1} & a_{i_2 2} & \dots & a_{i_2 n} \\ \dots & \dots & \dots & \dots \\ a_{i_l 1} & a_{i_l 2} & \dots & a_{i_l n} \\ 0 & \dots & -1 & 0 \dots 0 \\ 0 & \dots & 0 & -1 \dots 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 \dots -1 \end{pmatrix} = \begin{pmatrix} a_{i_1}^T \\ a_{i_2}^T \\ \dots \\ a_{i_l}^T \\ -e_{j_1}^T \\ -e_{j_2}^T \\ \dots \\ e_{j_p}^T \end{pmatrix}, \quad (6-10)$$

где $i_1, i_2, \dots, i_n \in I(k)$; $j_1, j_2, \dots, j_n \in J(k)$.

Произвольный вектор m пространства M_k может быть представлен в виде линейной комбинации векторов базиса, т. е. с помощью линейной комбинации строк матрицы Q_k :

$$m = Q_k^T \alpha, \quad (6-11)$$

причем $\alpha^T = (\alpha_1, \alpha_2, \dots, \alpha_l, \alpha_{l+1}, \dots, \alpha_{l+p})$ — вектор коэффициентов в разложении m по базисным векторам пространства M_k . Направление антиградиента может быть представлено в виде суммы двух векторов:

$$-\nabla F(x^k) = s^k + m^k, \quad (6-12)$$

где $m^k \in M_k$ и s^k ортогонален любому вектору пространства M_k . В силу ортогональности s^k базисным векторам из M_k имеем:

$$Q_k s^k = 0. \quad (6-13)$$

Для определения s^k из равенства (6-12) следует вычислить вектор m^k . Умножая вектор (6-12) слева на Q_k и используя свойство (6-13), получаем:

$$-Q_k \nabla F(x^k) = Q_k m^k.$$

Подставляя выражение для вектора m^k из (6-11), получаем:

$$\alpha^k = -(Q_k Q_k^T)^{-1} Q_k \nabla F(x^k), \quad (6-14)$$

откуда

$$m^k = Q^T R \alpha^k = -Q^T R (Q_k Q_k^T)^{-1} Q_k \nabla F(x^k).$$

Подставляя эту формулу в (6-12), получаем:

$$s^k = -(\nabla F(x^k) + m^k) = -\nabla F(x^k) + \\ + Q_k^T (Q_k Q_k^T)^{-1} Q_k \nabla F(x^k) = [E - Q_k^T (Q_k Q_k^T)^{-1} Q_k] [-\nabla F(x^k)], \quad (6-15)$$

где E — единичная матрица размерности n . Матрица

$$P_k = E - Q_k^T (Q_k Q_k^T)^{-1} Q_k \quad (6-16)$$

называется *матрицей проектирования*.

Если вектор s^k , полученный по формуле (6-15), отличен от нуля, следующая итерация определяется по формуле $x^{k+1} = x^k + \lambda s^k$. Однако в тех случаях, когда $s^k = 0$, точка x^k не всегда является оптимальным решением. Равенство ξ^k нулю означает, что вектор $\nabla F(x^k)$ находится в подпространстве M_k , базисом которого служат нормали к плоскостям, ограничивающим допустимую область.

Для дальнейшего продвижения к оптимуму следует исключить из матрицы Q_k одно из ограничений. Это ограничение определяется на основе разложения антиградиента по векторам, входящим в матрицу Q_k . Из условия $s^k = 0$ с помощью (6-11) и (6-12) получаем:

$$-\nabla F(x^k) = Q^T R \alpha. \quad (6-17)$$

Это выражение представляет собой систему из n уравнений для определения коэффициентов α . Поскольку размерность вектора равна $l+p$, то для вычисления его координат достаточно выбрать первые $l+p$ уравнений. Остальные уравнения удовлетворяются автоматически в силу условия $-\nabla F(x^k) \in M_k$. Дальнейший выбор направления s осуществляется на основе анализа коэффициентов α_i , $i=1, 2, \dots, l+p$.

Если среди значений α_i имеются отрицательные, удастся определить некоторое подходящее направление. Пусть, например, $\alpha_1 < 0$. Тогда ограничение, соответствующее α_1 , можно ослабить, допустив, что оно может выполняться, как неравенство (до этого предполагалось, что оно выполняется, как строгое равенство, что соответствовало движению вдоль ограничивающей гиперплоскости). При этом проекция s^k антиградиента на

подпространство M_k , полученное путем исключения из базиса вектора с коэффициентом α_i , отлична от нуля. Следовательно, движение в направлении \hat{s}^k позволяет уменьшить значение критерия оптимальности.

Если же все α_i неотрицательны, то в точке x^k подходящие направления отсутствуют и, следовательно, в ней достигается оптимум критерия $F(x)$.

Модификация отдельных процедур метода при расчетах на ЭВМ. Условие $s^k=0$, подлежащее проверке на каждой итерации, не может выполняться точно при расчетах с помощью ЭВМ из-за погрешностей вычислений. Поэтому ограничиваются приближенной проверкой этого равенства. Так, относительная устойчивость метода по отношению к погрешностям вычислений достигается при использовании в качестве контрольного параметра величины проекции градиента на направление s^k . Величина проекции в точке x^k вычисляется по формуле

$$g(s^k, x) = \frac{s^{kT} \nabla F(x)}{s^{kT} s^k}. \quad (6-18)$$

В точке x^k вычисляется величина $g(s^k, x^k)$ для направления $s^k = -P_k \nabla F(x^k)$. Если $g(s^k, x^k) < \epsilon$, производится выбор $\min \alpha_i$. Пусть $\alpha_{i_0} = \min \alpha_i$, тогда матрица Q_k получается исключением из матрицы Q_k строки, соответствующей ограничению i_0 , и матрица P_k получается из Q_k с помощью формулы (6-16).

Для вектора $\hat{s}^k = -P_k \nabla F(x^k)$ также производится вычисление величины $g(\hat{s}^k, x^k)$. Движение к экстремуму происходит по тому из двух направлений s^k и \hat{s}^k , для которого проекция градиента принимает большее значение. Процесс вычислений заканчивается, если проекция градиента на каждое из направлений с $\alpha_i < 0$ не превосходит ϵ .

Применение проекционного градиентного метода приводит иногда к необоснованному увеличению числа итераций. Чтобы этого избежать, используют обычно два приема, каждый из которых обладает своими достоинствами и недостатками.

Первый прием состоит в том, что из матрицы Q_k исключаются все ограничения, для которых коэффициенты α_i принимают отрицательное значение. Полученная в результате матрица проектирования P_k позволяет бо-

лее быстро продвигаться к экстремуму. Недостаток такого приема в том, что матрица Q_{k+1} , вычисляемая в точке x_{k+1} , никак не связана с матрицей Q_k . В результате матрицу Q_{k+1} приходится вычислять заново, что значительно усложняет процесс вычислений. Этот недостаток становится особенно заметным при наличии большого числа ограничений.

Второй прием основан на предположении, что начальная точка поиска выбирается по возможности, внутри области, а не на ее границе. В этом случае матрица Q_k имеет вначале небольшую размерность, которая постепенно возрастает с ростом числа итераций. Однако выбор начальной точки внутри области сопряжен со значительными трудностями. Обычно он осуществляется на основе предварительных пробных расчетов проектируемых конструкций.

Модификации метода. С целью ускорения поиска оптимума в [35] предлагается при выходе на границу допустимой области проектировать направление $-\nabla F(x^k)$ на одно из существенных ограничений и проверять, является ли такая проекция допустимым направлением. Если это условие не выполняется ни для одного из существенных направлений, то строится проекция антиградиента на пересечение всех существенных в данной точке ограничений.

Такой алгоритм, иногда приводит к преждевременному прекращению поиска, хотя оптимум еще не найден. Это иллюстрируется в следующих примерах [20].

Пример 1. Пусть $\nabla F(x^k) = (1; 1/8; -1)$; $\nabla R_1(x^k) = (-\sqrt{2}/2; -\sqrt{2}/2; 0)$; $x_j \geq 0$, $j = 1, 2, 3$ и в точке x^k имеет место $x^k_2 = 0$, $R_1(x^k) = 0$. Поскольку ограничение $R_1(x) \leq 0$ является существенным, то по алгоритму [35] предлагается двигаться по проекции антиградиента на гиперплоскость, касательную к $R_1(x)$ в точке x^k . Это направление $s^k = (+7/16; -7/16, -1)$, однако, не является возможным, так как противоречит условию $(s^k, e_2) \geq 0$. Между тем в этом случае существует подходящее направление, например $s^k = (0, 0, -1)$.

Пример 2. Пусть $\nabla F(x^k) = (1; 0; 1; 1)$; $R_i(x^k) = 0$, $i = 1, 2, 3$; $\nabla R_1(x^k) = (-4/5; 3/5; 0; 0)$; $\nabla R_2(x^k) = (-4/5; -3/5; 0; 0)$; $\nabla R_3(x^k) = (-5/6; -1/2; 1/6; 1/6)$.

Все ограничения являются существенными, и ни одна из проекций на касательные гиперплоскости ограничений не является возможным направлением. В то же вре-

мя проекция на пересечение ограничений не является подходящим направлением. Однако подходящее направление существует: $s^k = (0; 0; -1; -1)$. Оно может быть получено проектированием направления антиградиента на пересечение ограничений $R_1(x) \leq 0$ и $R_2(x) \leq 0$.

Более целесообразным представляется следующий алгоритм перебора проекций на существенные ограничения. Сначала производится проектирование вектора $-\nabla F(x)$ на каждое из существенных ограничений в отдельности и проверяется, является ли полученное направление возможным. Если среди них возможных не оказалось, то вектор $-\nabla F(x)$ проектируется на все попарные пересечения существенных ограничений, затем на все пересечения, состоящие из трех существенных ограничений, из четырех и т. д. Последней рассматривается проекция на пересечение всех существенных ограничений.

Рассмотренный подход полезен при решении очень сложных задач оптимального проектирования, когда объем вычислений при расчете модели проектируемого изделия в одной точке пространства параметров сравним с затратами на вычисления при выборе подходящего направления.

6-3. МЕТОД ПЕРЕМЕННОЙ МЕТРИКИ В ОГРАНИЧЕННОЙ ОБЛАСТИ

Выбор направления движения к оптимуму N_F может осуществляться на основе метода переменной метрики (см. § 4-3). Поскольку формирование направления N_F с помощью этого метода сопровождается изменением метрических свойств пространства параметров, естественно также и оператор проектирования $P[N_F]$ рассматривать в этой новой метрике.

По методу переменной метрики направление движения к оптимуму определяется формулой $s^k = -H_k \nabla F(x^k)$, где матрица H_k формируется по некоторому рекуррентному соотношению.

Пусть в точке x^k некоторое ограничение выполняется как строгое равенство, т. е.

$$R_i(x^k) = 0.$$

Тогда в x^k осуществляется преобразование матрицы H_k для того, чтобы новое направление было возможным.

Новая матрица H_{k1} , используемая для определения направления s^k , вычисляется по формуле

$$H_{k1} = H_{k0} - \frac{H_{k0} \nabla R_i \nabla R_i^T H_{k0}^T}{\nabla R_i^T H_{k0} \nabla R_i}, \text{ где } H_{k0} = H_k. \quad (6-19)$$

Если в точке x^k несколько ограничений выполняются как строгие равенства, для них формируется матрица Q_k в соответствии с формулой (6-10). Для выбора направления поиска вычисляется матрица H_{k1} по формуле

$$H_{k1} = H_{k0} - H_{k0} Q_k (Q_k^T H_{k0} Q_k)^{-1} Q_k^T H_{k0},$$

и для определения направления s^k имеем: $s^k = -H_{k1} \times \nabla F(x^k)$.

Если на некотором шаге $H_{k1} \nabla F(x^k) = 0$, то число строк в матрице следует уменьшить. Так же, как и в проекционном градиентном методе, производится вычисление вектора α путем решения системы линейных уравнений. Соответствующая система имеет вид:

$$(Q_k^T Q_k) \alpha = Q_k^T g^k. \quad (6-20)$$

Если все координаты вектора α положительны, оптимальное решение достигнуто. В противном случае из матрицы Q_k исключается строка, соответствующая наибольшему по модулю отрицательному элементу α_p . Это означает, что ограничение вида (6-1) или (6-2), коэффициенты которого составляют p -ю строку q^T_p матрицы Q_k , выполняется на следующей итерации как строгое неравенство. Полученную в результате исключения p -й строки матрицу обозначим через Q_k . Исключенная из матрицы Q_k строка имеет вид: $q^T_p = a^T$, если $p \in I(k)$, и $q^T_p = -e^T_j$, если $p \in J(k)$.

Для преобразованной матрицы H_{k1-1} получаем следующее выражение:

$$H_{k1-1} = H_{k1} + \frac{P_{l-1} q_p q_p^T P_{l-1}^T}{q_p^T P_{l-1} q_p},$$

где

$$P_{l-1} = E - Q_k (Q_k^T Q_k)^{-1} Q_k^T.$$

Таким образом, основные этапы рассмотренного алгоритма соответствуют основным шагам проекционно-градиентного метода. Отличие заключается в том, что про-

ексионно-градиентный метод использует исключительно ортогональную проекцию градиента на допустимую область. Такое проектирование соответствует единичной матрице H_k . В методе переменной метрики отношение ортогональности полностью определяется матрицей H_k и поэтому меняется от итерации к итерации. Изложенный метод позволяет быстрее продвинуться к экстремуму как за счет более эффективного выбора направления на экстремум критерия оптимальности, так и за счет способа построения проекции этого направления на допустимую область.

6.4. ВХОЖДЕНИЕ В ДОПУСТИМУЮ ОБЛАСТЬ. ВЫБОР ДЛИНЫ ШАГА

Для полноты изложения метода возможных направлений необходимо указать способ получения начальной точки x^0 , лежащей внутри допустимой области, а также способ вычисления длины шага λ_k .

Вхождение в допустимую область. С целью универсализации процедур поиска задача вхождения в допустимую область формулируется как некоторая задача оптимизации. Пусть \tilde{x} — произвольная точка n -мерного пространства проектируемых параметров, в которой часть из ограничений (6-1) не удовлетворяется. Тогда множество всех ограничений (6-1) можно разбить на две группы:

$$I_1 = \{i \mid R_i(\tilde{x}) \leq 0, i = 1, 2, \dots, m\};$$

$$I_2 = \{i \mid R_i(\tilde{x}) > 0, i = 1, 2, \dots, m\}.$$

Ограничения с индексами из множества I_1 выполняются в точке \tilde{x} , а с индексами из I_2 — не выполняются. Задача вхождения в допустимую область, иначе говоря, поиска допустимого решения формулируется следующим образом. Найти $(n+1)$ -мерный вектор (ξ, x) , для которого критерий оптимальности вида

$$\Phi(\xi, x) = \xi$$

достигает своего минимального значения и удовлетворяются ограничения

$$R_i(x) \leq 0 \text{ для } i \in I_1; R_i(x) - \xi \leq 0 \text{ для } i \in I_2.$$

При достаточно большом значении ξ вектор (ξ, \tilde{x}) является допустимым решением этой задачи¹. Минимизация критерия $\Phi(\xi, x)$ приводит к постепенному уменьшению максимального из значений функций ограничений множества I_2 . Как только достигнуто значение $\Phi(0, x^0)$, точка x^0 оказывается допустимым решением исходной за-

¹ Например, при $\tilde{\xi} = \max_{i \in I_2} \{-R_i(\tilde{x})\}$.

дачи. Она и принимается в качестве начальной точки для метода возможных направлений.

Выбор длины шага поиска. После выбора s^k следующая итерация задается формулой

$$x^{k+1} = x^k + \lambda_k s^k, \quad (6-21)$$

где величина шага λ_k определяется из условия

$$F(x^k + \lambda s^k) = \min_{\lambda} \{F(x^k + \lambda s^k) \mid x^k + \lambda s^k \in G\}. \quad (6-22)$$

При этом, если в задачах безусловной оптимизации одномерный поиск экстремума производится на всей прямой, при наличии огра-

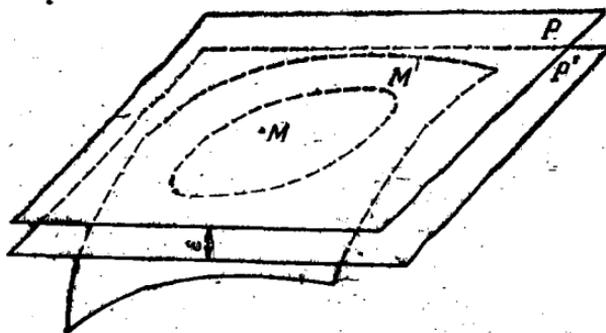


Рис. 6-3. Аппроксимация нелинейного ограничения касательной и текущей гиперплоскостями.

ничений он осуществляется лишь на некотором ее отрезке, соответствующем допустимым значениям переменных. Поэтому необходимо контролировать, не нарушено ли при выбранной длине шага какое-либо из ограничений.

Предположим, что в окрестности текущей итерации осуществляется линейная аппроксимация ограничений. Тогда допустимая область в окрестности точки x^0 ограничена гиперплоскостями

$$\sum_{i=1}^n \frac{\partial R_i(x^0)}{\partial x_j} (x_j - x_j^0) \leq 0, \quad i = 1, 2, \dots, m, \quad (6-23)$$

касательными к ограничениям в точке x^0 . Они аппроксимируют исходную допустимую область в некоторой окрестности точки x^0 .

Погрешность аппроксимации тем существенней, чем больше кривизна поверхности ограничения в окрестности этой точки. Чтобы уменьшить эту погрешность в тех случаях, когда известно направление изгиба поверхности, производится параллельный перенос ограничивающей гиперплоскости. Так, например, если функция $R_i(x)$

является выпуклой (рис. 6-3), то она аппроксимируется линейным ограничением

$$\sum_{j=1}^n \frac{\partial R_i(x^0)}{\partial x_j} (x_j - x_j^0) \leq -\varepsilon, \quad (6-24)$$

где ε — некоторое положительное число. Соответствующая гиперплоскость P' параллельна гиперплоскости P , касательной к ограничению, но смещена внутрь допустимой области. Аппроксимация допустимой области гиперплоскостями вида (6-24) определяет так называемую ε -область. Хотя при таком подходе часть точек оказывается вне многогранника линейных ограничений, точность аппроксимации оказывается несколько выше. Если же функция $R_i(x)$ вогнута, то смещение ограничения производится в противоположную сторону.

Линейные ограничения, соответствующие аппроксимации (6-23) или (6-24), будем для краткости обозначать

$$\sum a_{ij} x_j \leq b_i, \quad i=1, 2, \dots, m. \quad (6-25)$$

Чтобы выполнялись ограничения (6-2), величина шага должна удовлетворять следующему условию:

$$\lambda < \min_j \left\{ -\frac{x_j^k}{s_j^k} \right\}, \quad (6-26)$$

где минимум берется по тем координатам вектора s^k , которые принимают отрицательные значения в точке x^k .

Чтобы выполнялись ограничения (6-25), шаг λ должен удовлетворять условию

$$\lambda < \min_i \left\{ \frac{b_i - \sum_{j=1}^n a_{ij} x_j^k}{\sum_{j=1}^n a_{ij} s_j^k} \right\}, \quad (6-27)$$

где минимум берется по всем ограничениям $i=1, 2, \dots, m$, для которых $\sum_{j=1}^n a_{ij} s_j^k < 0$.

Поэтому для проведения одномерного поиска из точки x^k в направлении s^k определяются две величины:

$$p_k = \begin{cases} \min_j \left\{ -\frac{x_j^k}{s_j^k} \right\} & \text{по всем отрицательным координатам} \\ & \text{вектора } s^k; \\ \infty, & \text{если все координаты вектора } s^k \text{ положительны,} \end{cases} \quad (6-28)$$

и

$$\gamma_k = \begin{cases} \min_j \left\{ \frac{b_i - \sum_{j=1}^n a_{ij} x_j^k}{\sum_{j=1}^n a_{ij} s_j^k} \right\} & \text{по всем ограничениям } i=1, 2, \dots \\ & \dots, m, \text{ для которых } \sum_{j=1}^n a_{ij} s_j^k > 0; \\ \infty, & \text{если для всех ограничений справедливо неравенство} \\ & \sum_{j=1}^n a_{ij} s_j^k \leq 0, i = 1, 2, \dots, m. \end{cases} \quad (6-29)$$

Поиск минимума осуществляется на отрезке

$$0 \leq \lambda \leq \min \{\lambda_k, \rho_k, \gamma_k\}. \quad (6-30)$$

Однако приведенные условия позволяют выбрать однозначно параметр λ , когда на отрезке, определенном формулой (6-30), имеется не один, а несколько локальных минимумов. Такая ситуация встречается как в многоэкстремальных, так и в одноэкстремальных задачах оптимального проектирования. Поэтому для определенности в алгоритмах одномерного поиска производится выбор минимального значения λ , для которого выполнено условие (6-30).

6-5. МЕТОД ЛИНЕЙНОГО ЛОКАЛЬНОГО МОДЕЛИРОВАНИЯ

Использование проекционного градиентного метода эффективно, когда функции ограничений являются линейными. Поэтому при решении практических задач используется локальная линеаризация области (а иногда и критерия оптимальности). Такой подход носит название *линейного локального моделирования* [20].

Локальное моделирование заключается в том, что в окрестности опорной точки x^0 исходная математическая модель упрощается, например линеаризуется, и, таким образом, получается некоторая локальная модель. Устанавливаются вид и размеры окрестности, в пределах которой локальная модель с достаточной точностью заменяет исходную. Пусть это будет, например сфера S_ρ радиусом ρ с центром в точке x^0 . Область, которая на локальной модели соответствует допустимой области, обозначается G .

На локальной модели определяется оптимальная точка x^0 , т. е. обладающая наилучшим значением критерия

оптимальности по сравнению со всеми другими точками, принадлежащими $S_p \cap \tilde{G}$. В точке \tilde{x} прогнозируются значения критерия оптимальности F и функций ограничений R_i . Кроме того, в ней ставится эксперимент, т. е. вычисляются значения F и R_i . Если прогноз, полученный на основе локальной модели, подтверждается, т. е. точка \tilde{x} оказывается лучше опорной точки x^0 , центр локальной модели переносится в точку \tilde{x} и строится новая локальная модель. Затем отыскивается оптимальная точка новой локальной модели и т. д. до тех пор, пока $|x^0 - \tilde{x}|$ не станет достаточно малой. Если прогноз не подтверждается, то центр x^0 остается прежним, однако радиус сферы S_p уменьшается и повторяются заново все операции.

При локальном моделировании легко осуществляется адаптация при поиске. Адаптация заключается в том, что по рассогласованию прогноза и результатов эксперимента корректируется радиус ρ либо при достаточно малом рассогласовании локальная модель в той или иной степени сохраняется и для нового центра x^0 .

Рассмотрим вариант метода — *линейное локальное моделирование с адаптацией*. Построение локальной модели эквивалентно вычислению функций ограничений, которые переносятся в новый центр при хорошем совпадении прогноза и результатов эксперимента.

Будем называть *линейной моделью* совокупность центра x^0 , радиуса ρ , значений критерия оптимальности F , функций ограничений R_i и их градиентов. Критерий оптимальности и функции ограничений в пределах сферы S_p могут быть представлены в виде

$$\tilde{F}(x) = F(x^0) + F'(x^0)^T (x - x^0); \quad (6-31)$$

$$\tilde{R}_i(x) = R_i(x^0) + \nabla R_i(x^0)^T (x - x^0), \quad i = 1, 2, \dots, m. \quad (6-32)$$

Поиск оптимума на линейной модели осуществляется с помощью проектирования градиента критерия оптимальности.

Для изложения алгоритма вводятся следующие обозначения:

G_s — s -область, т. е. множество точек, принадлежащих допустимой области G и удаленных от ее границ

не более чем на ε . Соответствующая область на модели обозначается через \tilde{G}_ε ;

$F(\alpha)$ и $R_j(\alpha)$ — соответственно результаты прогноза значений функции F и R_j в точке α , вычисленные по формулам;

Γ — граница допустимой области G ;

Γ_ε и $\tilde{\Gamma}_\varepsilon$ — границы соответственно областей G_ε и \tilde{G}_ε ;

N^0 — единичный вектор в направлении N ;

$H_{0_\varepsilon}(\psi, \xi) = \{j | \tilde{R}_j(\xi) - \varepsilon | R_j(\psi) | = 0\}$ — множество индексов плоскостей, ограничивающих и проходящих через точку ξ ;

$H_1(N) = \{j | N, R_j(\psi) < 0\}$ — множество индексов ограничений, существенных по отношению к направлению N .

Радиус ρ устанавливается так, чтобы $\tilde{G}_\varepsilon \cap S_\rho \subset G_{1/2}$. Приращения аргументов при вычислении градиентов в центре новой модели $\Delta x_i = \varepsilon/2$, $i=1, 2, \dots, n$. При этом пробные точки при вычислении градиентов остаются в области G_ε .

Для определения оптимальной точки модели прежде всего вычисляется оптимальное направление N из центра модели ψ . Двигаясь по нему, можно либо выйти на границу линейной модели и тогда оптимальная точка модели найдена, либо в пределах сферы s_ρ встретить границу Γ_ε и тогда, обозначив точку встречи через ξ , отыскивать на локальной модели новое оптимальное направление $N(\xi)$. Итак:

1) принимаем $\xi = \psi$;

2) определяем оптимальное направление N .

Если $H_{0_\varepsilon}(\psi, \xi) \neq \emptyset$, полагаем $N = F(\psi)$. Пусть $H_{0_\varepsilon}(\psi, \xi) = \emptyset$ и существует μ подмножеств H^v , $v=1, 2, \dots, \mu$ множества $H_{0_\varepsilon}(\psi, \xi)$ таких, что

$$H^v \cap H_1(F) \neq \emptyset \quad (6-33)$$

и

$$(P(\bigcap_{j \in H^v} \lambda_j F(\psi), R_j) \geq 0 \text{ для всех } H_{0_\varepsilon}(\psi, \xi). \quad (6-34)$$

Находим $M(\xi) = \max(F(\omega), P(\bigcap_{j \in H^v} \lambda_j)F(\omega))$. Пусть максимум $M(\xi)$ достигается при $v = v_0$, т. е. $(F(\omega), P(\bigcap_{j \in H^{v_0}} \lambda_j)F(\omega)) = M(\xi)$. Тогда полагаем $N = (P(\bigcap_{j \in H^{v_0}} \lambda_j)F(\omega))$.

Очевидно, что если подмножества H^{v_1} и H^{v_2} удовлетворяют условиям (6-33) и (6-34) и $H^{v_1} < H^{v_2}$, то

$$(F(\omega), P(\bigcap_{j \in H^{v_1}} \lambda_j)F(\omega)) \leq (F(\omega), P(\bigcap_{j \in H^{v_2}} \lambda_j)F(\omega)).$$

Поэтому для определения $M(\xi)$ не требуется полного перебора подмножества H^v , $v = 1, 2, \dots, \mu$.

Если не существует подмножества, удовлетворяющего (6-33) и (6-34), или оказывается, что $|N| < \varepsilon_1$, где ε_1 — некоторое наперед заданное положительное число, то точка ξ является искомой оптимальной точкой линейной модели.

3. Если $|N| > \varepsilon_1$, то вычисляем длину пути из точки ξ по направлению N до границы Γ_s :

$$l_{min} = \min_{j \in H_1(N)} l(N, j), \text{ где } l(N, j) = \left| \frac{\tilde{R}_j(\xi) - \varepsilon |R_j(\xi)|}{(N, R_j(\xi))} \right|,$$

и запоминаем индекс j_l , при котором достигается этот минимум; затем вычисляем длину пути l' из точки ξ до поверхности сферы S_p в том же направлении N :

$$l' = (N^0, (\psi - \xi)) + \sqrt{\rho^2 + (N^0, (\psi - \xi))^2 - (\psi - \xi)^2}.$$

4. Вычисляем $\alpha = \xi + N^0_{min}\{l', l_{min}\}$.

5. Прогнозируем значения критерия и функций ограничений для точки α по формулам (6-31) и (6-32).

При $l_{min} \geq l'$ точка α является оптимальной точкой линейной модели. Если $l_{min} < l'$, то принимаем $\xi = \alpha$ и заметив, что $H_{0_s}(\psi, \xi) = H^{v_0} \cup \{j_l\}$, возвращаемся к п. 2.

Адаптация при поиске и переход к следующей модели осуществляется следующим образом. Если по результатам вычислений $F(\alpha)$ и $R_j(\alpha)$ оказывается, что $F(\alpha) < F(\psi)$ или $\alpha \notin G_s$, то радиус ρ уменьшается и повторяются пп. 1-5.

При $F(\alpha) > F(\psi)$ вычисляются рассогласования:

$$\Delta_0 = \frac{|F(\alpha) - \tilde{F}(\alpha)|}{|F(\alpha)| + |\tilde{F}(\alpha)|}; \quad \Delta_j = \frac{|R_j(\alpha) - \tilde{R}_j(\alpha)|}{|R_j(\alpha)| + |\tilde{R}_j(\alpha)|}, \quad j=1, 2, \dots, k,$$

и если хотя бы одно из них превышает заданное допустимое значение ε_2 , то пересчитываются градиенты $F(\alpha)$ и $R_j(\alpha)$, а радиус ρ уменьшается. При $\Delta_j < \varepsilon_2, j=1, 2, \dots, k$ градиенты не пересчитываются, а радиус увеличи-

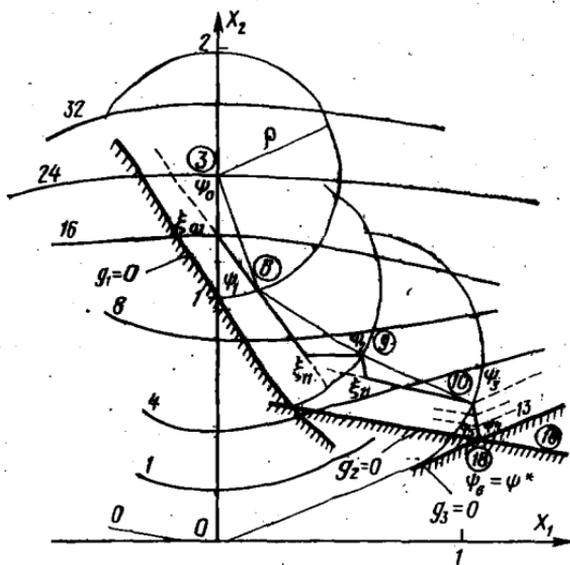


Рис. 6-4. Поиск оптимума методом линейного локального моделирования.

вается. Построение локальной модели завершается переносом $\psi = \alpha$.

Если оптимальная точка находится внутри локальной модели, уточнение положения оптимума производится по тому же алгоритму, но при уменьшенных ε и ρ . Поиск заканчивается при $\rho < \varepsilon_{min}$, $\varepsilon < \varepsilon_{min}$.

Движение в допустимую область, а при необходимости возврат в нее осуществляется с помощью минимизации вспомогательной функции (см. § 6-4).

Ход поиска с помощью изложенного алгоритма представлен на рис. 6-4. Минимизируется функция

$$F = x_1^4 + 4x_1^2 x_2^2 - 4x_1^2 x_2 - 3x_1^2 + 8x_2^2 + 4x_2$$

при ограничениях

$$R_1 = x_2 - (x_1 - 1)^2 \geq 0; R_2 = 6x_2^2 - 2 + x_1 \geq 0;$$

$$R_3 = 6x_2^2 + 2 - 3x_1 \geq 0.$$

Принятые значения параметров: $\rho_0 = 0,5$; $\epsilon_0 = 0,1$; $\epsilon_{\min} = 0,02$; $\Delta j = 10\%$, $j = 1, 2, \dots, k$. В кружках показано число вычислений функций F, R_1, R_2, R_3 с начала поиска.

Пунктиром обозначены границы $\tilde{\Gamma}_j$ локальных моделей.

Алгоритм линейного локального моделирования позволяет более полно использовать информацию, полученную в ходе поиска. Благодаря локализации, т. е. определению положения точки поиска по отношению к границе допустимой области, и введению ϵ -области, поиск не выходит за пределы допустимой области после каждой итерации.

Выбор на линейной модели наилучшего возможного направления движения среди проекций градиента критерия оптимальности на граничные поверхности выгоден в тех практических задачах, где сравнительно велико время вычисления критерия оптимальности и функций ограничений.

Большая экономия затрат на поиск в рассмотренном алгоритме достигается благодаря адаптации. В частности, градиенты критерия оптимальности и функций ограничений пересчитываются при переходе от одной локальной модели к другой лишь при достаточно большом рассогласовании прогноза и результатов эксперимента; с помощью адаптации корректируется и величина ρ , характеризующая размер локальной модели. Это дает заметные преимущества в тех случаях, когда невелика кривизна ограничивающих поверхностей и поверхностей уровня критерия оптимальности.

6-6. ПРИМЕР. ОПТИМАЛЬНОЕ ПРОЕКТИРОВАНИЕ ДИСКА ТУРБИНЫ

Рассмотрим постановку задачи проектирования диска паровой турбины, имеющего минимальную массу [64]. Для ее решения используется модификация проекционно-градиентного метода, позволяющая в ряде случаев при выходе на границу допустимой области находить новую начальную точку поиска внутри области, не ухудшая значения критерия оптимальности.

Математическая модель объекта. Диск турбины можно рассматривать как вращающийся круглый диск переменной толщины (рис. 6-5). Масса такого диска выражается с помощью формулы

$$W(r) = \int_{a_1}^{a_m} 2\pi r h(r) dr, \quad (6-35)$$

где a_1 и a_m — внутренний и внешний радиусы соответственно; ρ — плотность; $h(r)$ — толщина диска на расстоянии r от оси вращения.

Минимум достигается подбором функции $h(r)$. Однако эту функцию нельзя изменять произвольно, так как должны быть вы-

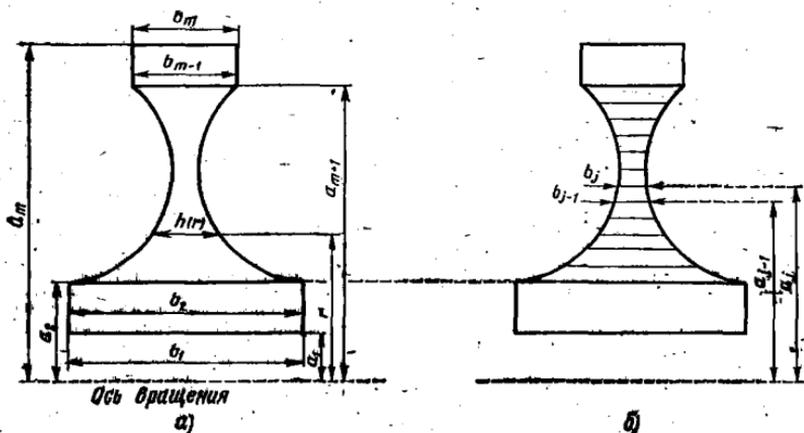


Рис. 6-5. Поперечное сечение диска паровой турбины (а) и его кусочно-линейная аппроксимация (б).

полнены требования к прочности диска при воздействии сил, возникающих во время его вращения.

Уравнение равновесия сил для вращающегося диска имеет вид:

$$\frac{d}{dr} (h\sigma_r) + \frac{h}{r} (\sigma_r - \sigma_\theta) + \rho\omega^2 r h = 0, \quad (6-36)$$

где σ_r и σ_θ — радиальное и тангенциальное давления соответственно, а ω — частота вращения диска. Это уравнение справедливо в предположении радиальной симметрии сил в плоскости, ортогональной к оси вращения.

Силы могут быть выражены через величину радиального смещения $u(r)$ с помощью следующих соотношений:

$$\sigma_r = \frac{E}{1-\nu^2} (\epsilon_r + \nu\epsilon_\theta); \quad \sigma_\theta = \frac{E}{1-\nu^2} (\nu\epsilon_r + \epsilon_\theta); \quad (6-37)$$

$$e_r = du/dr; e_\theta = u/r, \quad (6-38)$$

где e_r и e_θ — радиальная и тангенциальная деформации; E — модуль Юнга и ν — отношение Пуассона.

Подставляя уравнения (6-37) и (6-38) в (6-36), получаем:

$$\frac{d^2u}{dr^2} + \left(\frac{1}{r} + \frac{1}{h} \frac{dh}{dr} \right) \frac{du}{dr} - \left(\frac{1}{r} - \frac{\nu}{h} \frac{dh}{dr} \right) \frac{u}{r} + \frac{\rho\omega^2(1-\nu^2)}{E} r = 0, \quad (6-39)$$

Отсюда следует, что для определения $u(r)$ необходимо знать геометрию диска, т. е. $h=h(r)$, $a_1 \leq r \leq a_m$. В этом случае $u(r)$ определится как решение краевой задачи для уравнения (6-39) с граничными условиями:

$$\sigma_r|_{r=a_1} = s_1 \text{ и } \sigma_r|_{r=a_m} = s_m. \quad (6-40)$$

В результате решения краевой задачи и последующего применения формул (6-37) и (6-38) можно определить давления σ_r и σ_θ .

Кроме условия равновесия сил при вращении диска, должны удовлетворяться требования по прочности. Соответствующее ограничение имеет вид:

$$\max \{ |\sigma_r - \sigma_\theta|, |\sigma_r|, |\sigma_\theta| \} \leq 2\sigma_0, \quad (6-41)$$

где σ_0 — предельно допустимое давление.

Наконец, ограничения на пределы изменения толщины диска заданы следующим образом:

$$h(r) \begin{cases} = b_1 & \text{при } a_1 \leq r \leq a_2; \\ \geq \varepsilon & \text{при } a_2 \leq r \leq a_{m-1}; \\ = b_m & \text{при } a_{m-1} \leq r \leq a_m, \end{cases} \quad (6-42)$$

где ε — нижний предел толщины, заданный из технологических соображений; $a_m - a_{m-1} = \text{const}$.

Полученная в результате задачи минимизации выражения (6-35) при ограничениях (6-39)–(6-42) может быть решена методами оптимального проектирования.

Постановка задачи оптимального проектирования. Значения функций ограничений и критерия оптимальности полностью определены, если задана зависимость $h(r)$. Чтобы представить задачу оптимального проектирования как задачу варьирования многомерного вектора конструктивных параметров, производится аппроксимация зависимости $h(r)$ некоторой кусочно-линейной функцией. С этой целью производится разбиение радиуса диска на m частей:

$$a_1 < a_2 < \dots < a_{m-1} < a_m. \quad (6-43)$$

На каждом отрезке строится линейная функция, аппроксимирующая соответствующий участок изменения $h(r)$ (см. рис. 6-5).

Если толщина диска задана в точках разбиения величинами

$$h(a_j) = b_j, \quad j = 1, 2, \dots, m, \quad (6-44)$$

то кусочно-линейная аппроксимация полностью определена формулой

$$h(r) \approx \begin{cases} b_1 & \text{при } a_1 \leq r \leq a_2; \\ b_{j-1} + \left(\frac{b_j - b_{j-1}}{a_j - a_{j-1}} \right) (r - a_{j-1}) & \text{при } a_{j-1} \leq r \leq a_j, \\ & j = 3, \dots, m-1; \\ b_m & \text{при } a_{m-1} \leq r \leq a_m. \end{cases} \quad (6-45)$$

Подставляя формулу (6-45) в критерий оптимальности (6-35), получаем следующее выражение:

$$W = \frac{1}{3} \pi r \left\{ b_1 (-3a^2_1 + a^2_2 + a^2_3 + a_2 a_3) + \sum_{j=3}^{m-2} b_j (a_{j+1} - a_{j-1}) (a_{j+1} + a_j + a_{j-1}) + b_m (3a^2_m - a^2_{m-1} - a^2_{m-2} - a_{m-1} a_{m-2}) \right\}. \quad (6-46)$$

где $\varepsilon_1, \varepsilon_2, \varepsilon_3$ — заданные величины. В векторной форме эти ограничения могут быть представлены неравенством

$$l \leq x \leq u, \quad (6-48)$$

где $l^T = (\varepsilon_1, \dots, \varepsilon_1, a_1 + \varepsilon_3)$ и $u^T = (\infty, \dots, \infty, a_3 - \varepsilon_2)$.

Чтобы представить ограничение (6-39) в соответствии с принятым разбиением, запишем его отдельно для каждого участка. При этом будем предполагать разбиение таким, что изменением $h(r)$ внутри каждого участка можно пренебречь.

Уравнение (6-39) для каждого участка разбиения принимает более простую форму:

$$\frac{d^2 u}{dr^2} + \frac{1}{r} \frac{du}{dr} - \frac{u}{r^2} + \frac{\rho \omega^2 (1 - \nu^2)}{E} r = 0,$$

что позволяет получить общее решение в виде

$$u = c_1 r + \frac{c_2}{r} - \frac{\rho \omega^2 (1 - \nu^2)}{8E} r^3,$$

где c_1 и c_2 — произвольные постоянные. Отсюда с помощью (6-37) и (6-38) получаются выражения для сил вращения:

$$\begin{aligned} \sigma_r &= \alpha - \frac{\beta}{r^2} - \frac{\rho \omega^2 (3 + \nu)}{8} r^2; \\ \sigma_\theta &= \alpha + \frac{\beta}{r^2} - \frac{\rho \omega^2 (1 + 3\nu)}{8} r^2, \end{aligned} \quad (6-49)$$

где α , β — постоянные внутри каждого кольца, полученного в результате разбиения.

Для проектируемой конструкции определяются также силы σ_r и σ_θ в точках r_1, r_2, \dots, r_n . Ограничение (6-41) записывается отдельно для каждого интервала $[r_{j-1}, r_j]$. Имеем:

$$\tau_1 = \frac{1}{2} |\sigma_r - \sigma_\theta|, \quad \tau_2 = \frac{1}{2} |\sigma_r|, \quad \tau_3 = \frac{1}{2} |\sigma_\theta|,$$

откуда ввиду (6-41) получаем:

$$\max \{\tau_1, \tau_2, \tau_3\} \leq \tau_0. \quad (6-50)$$

Вектор функций ограничений может быть записан в виде

$$y^T = (\tau_{r_1}, \tau_{r_2}, \dots, \tau_{r_n}).$$

Пределы его изменения в соответствии с условием (6-50) записываются с помощью неравенства

$$L \leq y(x) \leq U, \quad (6-51)$$

где $L = (0, 0, \dots, 0)$ и $U = (\tau_0, \tau_0, \dots, \tau_0)$.

Таким образом, задача оптимального проектирования диска паровой турбины сводится к поиску минимума критерия оптимальности (6-46) при ограничениях (6-50) и (6-51) путем варьирования вектора $x = (b_3, \dots, b_{n-2}, a_2)$.

Движение к оптимуму внутри допустимой области. Поскольку на начальных этапах поиск осуществляется внутри допустимой области, используется метод наискорейшего спуска. Рекуррентная формула метода наискорейшего спуска имеет вид:

$$x^{k+1} = x^k + t_k \psi^k, \quad (6-52)$$

где $x^k = (b_3^k, b_4^k, \dots, b_{m-2}^k, a_2^k)$; $\psi^k = -\frac{\nabla w(x^k)}{|\nabla w(x^k)|}$; t_k — длина шага.

Частные производные критерия оптимальности в данном случае могут быть представлены в аналитической форме:

$$\frac{\partial w}{\partial b_j} = \frac{\pi \rho}{3} (a_{j+1} - a_{j-1}) (a_{j+1} + a_j + a_{j-1}) \quad \text{при } j = 3, \dots, m-2;$$

$$\frac{\partial w}{\partial a_2} = \frac{\pi \rho}{3} (2a_2 + a_3) (b_1 - b_3).$$

Поэтому соотношение (6-52) сводится к следующим выражениям:

$$b_j^{k+1} = b_j^k - \frac{\pi\rho}{3} (a_{j+1} - a_{j-1}) (a_{j+1} + a_j + a_{j-1}) \frac{t_k}{N_k};$$

$$a_2^{k+1} = a_2^k - \frac{\pi\rho}{3} (2a_2 + a_3) (b_1 - b_3) \frac{t_k}{N_k}, \quad (6-53)$$

где нормализующий множитель N_k имеет вид:

$$N_k = \frac{\pi\rho}{3} \left[\sum_{j=3}^{m-2} (a_{j+1} - a_{j-1})^2 (a_{j+1} + a_j + a_{j-1})^2 + \right. \\ \left. + (2a_2 + a_3)^2 (b_1 - b_3)^2 \right]^{1/2}.$$

Согласно методу наискорейшего спуска t_k выбирается в результате решения однопараметрической задачи минимизации. Кроме того, с учетом ограничений величина t_k подбирается таким образом, чтобы x^{k+1} было допустимым решением.

Следует установить, какое максимально возможное значение t_k не выводит за пределы допустимой области.

Для ограничений на конфигурацию изделия (6-48) с помощью соотношений (6-53) получаем:

$$t_j = \frac{3(b_{jk} - \varepsilon_j) N_k}{\pi\rho (a_{j+1} - a_j)(a_{j+1} + a_j + a_{j-1})} \quad \text{при } j = 3, \dots, m-2;$$

$$\hat{t}_1 = \frac{3(a_{2k} - a_1 - \varepsilon_1) N_k}{\pi\rho (2a_2 + a_3) (b_1 - b_3)};$$

$$\hat{t}_2 = \frac{3(a_3 - \varepsilon_2 - a_{2k}) N_k}{-\pi\rho (2a_2 - a_3) (b_1 - b_3)}.$$

Следовательно,

$$t_k \leq \min_{1 \leq j \leq m-2} \{\hat{t}_j \mid \hat{t}_j \geq 0\},$$

т. е. t_k выбирается минимальным среди всех положительных значений \hat{t}_j .

Чтобы сформулировать условия для выбора t_k , исходя из ограничений (6-51), рассмотрим эти ограничения на каждом из элементов разбиения. Обозначим через h_{ki} толщину диска при радиусе r_i ; τ_{r_i} — максимальное значение деформирующего усилия в слое r_i . Будем считать, что каждое h_{ki} может изменяться независимо без учета распределения сил на соседних участках разбиения.

Тогда расстояние до i -го участка разбиения

$$t_{ki} = \left(\frac{r_0 - r_{r_i}}{r_0} \right) \frac{h_{ki}}{\varphi_{ki}}, \quad (6-54)$$

где

$$\varphi_{ki} = \frac{\psi_{kj}(r_i - a_{j-1}) + \psi_{k|j-1}(a_j - r_j)}{a_j - a_{j-1}}$$

и

$$a_{j-1} \leq r_i \leq a_j \text{ при } j=3, \dots, m-2,$$

откуда следует, что для выполнения (6-54) необходимо, чтобы

$$t_k \leq \min_{3 \leq i \leq m-2} t_{ki}.$$

Таким образом, t_k убывает по мере приближения к ограничению. До тех пор, пока t_k имеет положительное значение, движение осуществляется внутри области; при выходе на границу t_k принимает нулевое значение.

Движение вдоль границы допустимой области. Способы движения вдоль границы допустимой области различаются в зависимости от типа ограничений. Так, при нарушении геометрических ограничений (6-47), линейных относительно конструктивных параметров, используется проекционно-градиентный метод.

Движение с помощью проекционно-градиентного метода, как и раньше, определяется формулой

$$x^{k+1} = x^k - t_k \psi^k,$$

где ψ^k вычисляется, исходя из следующих условий.

Случай 1. x^k лежит на ограничении $b_i = e_i$, $3 \leq i \leq m-2$. Тогда

$$\psi_j^k = \begin{cases} -\frac{1}{N} \left(\frac{\partial w}{\partial b_{j+2}} \right) & \text{при } j = 1, \dots, m-4, j \neq i; \\ 0 & \text{при } j = i; \\ -\frac{1}{N} \left(\frac{\partial w}{\partial a_2} \right) & \text{при } j = m-3, \end{cases}$$

где

$$N = \left[\sum_{j \neq i} \left(\frac{\partial w}{\partial b_{j+2}} \right)^2 + \left(\frac{\partial w}{\partial a_2} \right)^2 \right]^{1/2}.$$

Случай 2. x^k лежит на ограничении $a_2 = a_1 + e_3$ или $a_2 = a_3 - e_2$.

Тогда

$$\psi_j^k = \begin{cases} -\frac{1}{N} \left(\frac{\partial w}{\partial b_{j+2}} \right) & \text{при } j = 1, \dots, m-4; \\ 0 & \text{при } j = m-3, \end{cases}$$

где

$$N = \left[\sum_{j=1}^{m-4} \left(\frac{\partial w}{\partial b_{j+2}} \right)^2 \right]^{1/2}.$$

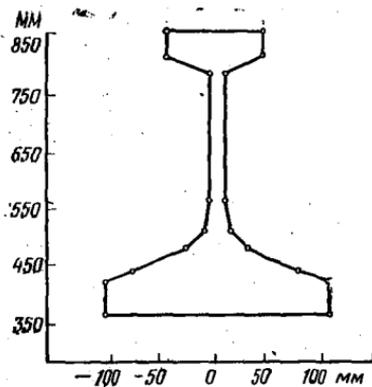


Рис. 6-6. Пример. Определение оптимального профиля диска турбины.

результаты не будут отличаться меньше, чем на заданное значение.

На рис. 6-6 представлен оптимальный вариант диска турбины, полученный в результате применения рассмотренного алгоритма.

Глава седьмая

МЕТОДЫ ШТРАФНЫХ ФУНКЦИЙ

7-1. ФОРМИРОВАНИЕ ШТРАФНЫХ ФУНКЦИЙ

Основная идея, положенная в основу методов штрафных функций, состоит в замене критерия оптимальности исходной задачи (масса изделия, приведенные затраты, мощность и т. п.) некоторым обобщенным критерием, значения которого совпадают с исходным критерием оптимальности внутри допустимой области. При выходе из допустимой области, а в некоторых случаях уже при приближении к границе обобщенный критерий оптимальности резко возрастает за счет штрафных функций, зависящих от ограничений.

Штрафные функции обеспечивают либо быстрое возвращение в допустимую область, либо невозможность выхода из нее. Вследствие того, что при использовании штрафных функций ограничения не присутствуют в явной форме, для поиска оптимального решения могут применяться методы безусловной оптимизации (т. е. оптимизации при отсутствии ограничений). Таким образом, методы штрафных функций позволяют свести задачу оптимального проектирования к последовательности задач, не содержащих ограничения.

Указанные формулы получены преобразованием проекций общего вида для конкретных ограничений.

При выходе на границу области, соответствующую нелинейным ограничениям (6-51), используются методы линейного локального моделирования этих ограничений.

В целом стратегия поиска оптимума должна строиться следующим образом. Для некоторого разбиения (вначале достаточно грубого) производится поиск оптимального варианта, затем шаг разбиения уменьшается вдвое и снова производится поиск оптимума.

Процесс продолжается до тех пор, пока при двух последовательных разбиениях полученные ре-

Весьма эффективно применение методов штрафных функций в тех случаях, когда гиперповерхности, ограничивающие допустимую область значений параметров, заданы нелинейными функциями. В этом случае движение к оптимуму вдоль границы допустимой области осуществляется автоматически в результате оптимизации обобщенного критерия оптимальности. В зависимости от способа формирования штрафных функций различают *метод штрафных функций и метод барьерных функций*.

Метод штрафных функций. Рассмотрим задачу поиска локального минимума критерия оптимальности $F(x)$ в области, ограниченной системой неравенств

$$R_i(x) \leq 0, \quad i=1, 2, \dots, m. \quad (7-1)$$

Введение обобщенного критерия оптимальности по методу штрафных функций производится с помощью некоторой непрерывной функции $Q(x)$, удовлетворяющей следующим условиям:

1) $Q(x) = 0$ внутри и на границе допустимой области, т. е. при значениях параметров x_1, x_2, \dots, x_n , удовлетворяющих ограничениям (7-1);

2) $Q(x) > 0$, если не выполнено хотя бы одно из ограничений (7-1).

Примерами функций $Q(x)$ могут служить выражения:

$$Q_1(x) = \sum_{i=1}^m [\max R_i(x), 0]^a, \quad a \geq 1, \quad (7-2)$$

а также

$$Q_2(x) = \sum_{i=1}^m \left[\frac{R_i(x) + |R_i(x)|}{2} \right]^a. \quad (7-3)$$

Обобщенным критерием оптимальности согласно методу штрафных функций является выражение

$$T(x, t) = F(x) + tQ(x), \quad (7-4)$$

где t — некоторое положительное число, называемое коэффициентом штрафа.

Метод оптимизации с помощью последовательности штрафных функций состоит в следующем. Рассматривается некоторая неограниченная, монотонно возрастающая последовательность $\{t_k\}$, $k=1, 2, \dots$ положительных чисел. Для первого элемента этой последовательности

с помощью методов, изложенных в предыдущих главах, отыскивается безусловный локальный минимум функции $T(x, t_1)$. Пусть этот минимум достигается в точке x^1_* .

Вектор x^1_* используется как начальное приближение для решения задачи поиска минимума функции $T(x, t_2)$, где $t_2 > t_1$ и т. д. Таким образом, решается последовательность задач минимизации функций $T(x, t_k)$; $k=1, 2, \dots$, причем результат предыдущей оптимизации x^k_* используется в качестве начального приближения для поиска x^{k+1}_* .

Поскольку для бесконечно возрастающей последовательности t_k локальные минимумы приближаются к допустимой области (далекне от допустимой области минимумы погашаются за счет роста штрафного члена), последовательность x^k_* , $k=1, 2, \dots$ сходится к локальному оптимуму, расположенному внутри или на границе допустимой области.

В качестве примера рассмотрим задачу поиска минимума критерия оптимальности, представляющего собой квадратичную функцию от двух переменных:

$$F(x_1, x_2) = 0,8x_1^2 + 1,7x_2^2 - 1,2x_1x_2 - 12,3x_1 + 11,4x_2 + 48,2$$

при ограничениях

$$R_1(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 2)^2 - 9 \leq 0; \\ x_1 \geq 0; x_2 \geq 0.$$

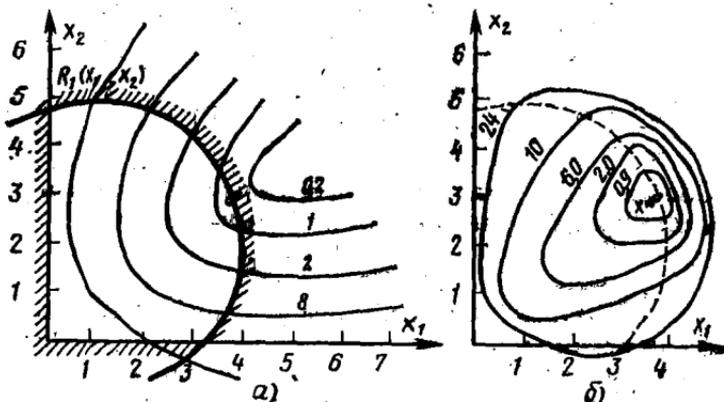


Рис. 7-1. Преобразование задачи оптимального проектирования с помощью штрафных функций.

а — исходная задача оптимизации; б — обобщенный критерий оптимальности.

Линии уровня функции $F(x_1, x_2)$ и допустимая область значений конструктивных параметров x_1 и x_2 представлены на рис. 7-1,а. Обобщенный критерий оптимальности сформирован с помощью функции $Q_2(x)$, определяемой формулой (7-3).

Линии уровня обобщенного критерия оптимальности при значении коэффициента штрафа $t=2$ приводятся на рис. 7-1,б. Результаты вычислений приводятся в табл. 7-1.

Таблица 7-1

k	t _k	x ₁ ^k	x ₂ ^k	Число итераций	Точность ∇T(x ^k)	Число вычислений	
						функции	градиента
1	2	4,127	3,435	4	0,06	15	5
2	4	3,994	3,196	5	0,02	17	6
3	10	3,941	3,045	5	0,04	19	6
4	25	3,907	2,974	6	0,02	22	7
5	50	3,876	2,912	7	0,01	28	8

Промежуточные итерации $x^{k*} = (x^{k_{1*}}, x^{k_{2*}})$ расположены вне допустимой области. Это свойство метода сохраняется при решении более сложных задач оптимизации. Если локальный минимум расположен на границе допустимой области, то последовательность безусловных оптимумов в методе штрафных функций сходится к оптимальному решению, все время находясь снаружи относительно допустимой области. По этой причине метод носит также название *метода внешней точки* [53].

Обобщенный критерий оптимальности представляет собой сложную функцию даже для допустимых областей достаточно простого вида. На рис. 7-2 приводится пример возникновения «овражной» ситуации при переходе к обобщенному критерию оптимальности. На рис. 7-2,а представлены линии уровня критерия оптимальности в исходной задаче, а на рис. 7-2,б — линии уровня обобщенного критерия. В последнем случае имеем дело с «овражной» ситуацией. При этом один из «склонов оврага» соответствует линиям уровня функции ограничения, а другой — линиям уровня критерия оптимальности. Однако трудности, порождаемые сложностью обобщенного критерия оптимальности, компенсируются возможностью использования хорошо разработанных

процедур поиска оптимума при отсутствии ограничений.

При наличии ограничений в виде равенств

$$g_l(x) = 0, \quad l = 1, 2, \dots, k \quad (7-5)$$

слагаемые функции $Q_1(x)$, соответствующие ограничениям (7-5), имеют вид:

$$|q_l(x)|^a, \quad l = 1, 2, \dots, k.$$

Обобщенный критерий оптимальности, построенный с помощью функции $Q_1(x)$ для случаев допустимой

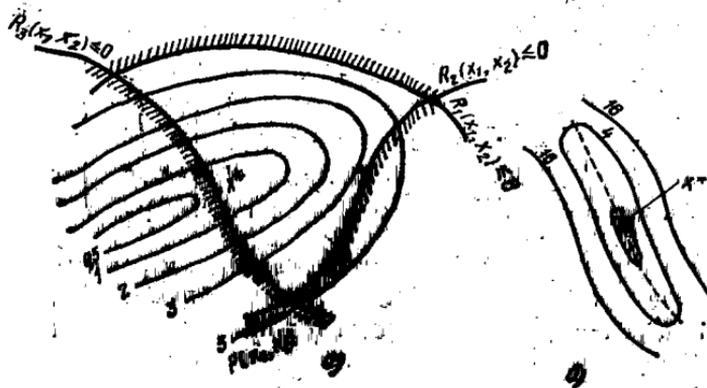


Рис. 7-2. Возникновение «отраженной» ситуации при формировании штрафной функции.

a — исходная задача; *b* — обобщенный критерий оптимальности.

области, заданной ограничениями (7-1) и (7-5), записывается в форме

$$T_1(x, t) = F(x) + t \left\{ \sum_{i=1}^k [\max(R_i(x), 0)]^a + \sum_{l=1}^k |g_l(x)|^a \right\}. \quad (7-6)$$

Коэффициент a позволяет избежать резкого возрастания критерия оптимальности при выходе из допустимой области. Это особенно важно на начальной стадии поиска.

Для квадратичной функции штрафа $Q_2(x)$ [см. (7-3)] и ограничений (7-1), (7-5) формула обобщенного

критерия оптимальности имеет вид:

$$T_2(x, t) = F(x) + t \left\{ \sum_{i=1}^m \left[\frac{R_i(x) - |R_i(x)|}{2} \right]^2 + \sum_{i=1}^m g_i^2(x) \right\}. \quad (7-7)$$

При применении метода штрафных функций любая точка векторного пространства параметров может быть выбрана в качестве начальной. Это значительно упрощает программирование алгоритмов, использующих штрафные функции.

Метод барьерных функций. В этом методе для построения обобщенного критерия оптимальности используются специальные функции, называемые *барьерными*. Значения, принимаемые барьерной функцией, неограниченно возрастают при приближении к границе допустимой области. Барьерные функции применяются в тех случаях, когда ограничения заданы только в виде неравенств.

Для построения обобщенного критерия оптимальности вводится функция $I(x)$, непрерывная на множестве внутренних точек G^0 допустимой области G , т. е. тех точек, для которых ограничения (7-1) выполняются как строгие неравенства. Она должна обладать следующим свойством: если $\{x^k\}$, $k=1, 2, \dots$ — последовательность внутренних точек, сходящаяся к граничной точке области G , то последовательность значений функции $I(x^k)$, $k=1, 2, \dots$ неограниченно возрастает. Функция $I(x)$ называется *барьерной функцией*.

Примером функции $I(x)$ является

$$I_1(x) = - \sum_{i=1}^m \ln [-R_i(x)]. \quad (7-8)$$

Эта функция существует только внутри допустимой области. Вне области G и на ее границе функция $I(x)$ не определена. При приближении к границе она неограниченно возрастает.

Другим примером барьерной функции служит:

$$I_2(x) = - \sum_{i=1}^m [R_i(x)]^{-1}, \quad (7-9)$$

которая несколько проще, чем $I_1(x)$, и определена всюду, за исключением границы области G . Обе барьерные

функции используются при решении практических задач оптимизации.

После того, как определена вспомогательная функция $I(x)$, обобщенный критерий оптимальности записывается в виде

$$U(x, r) = F(x) + rI(x), \quad (7-10)$$

где r — некоторое положительное число.

В частности, для рассмотренных примеров барьерных функций (7-8) и (7-9) получаем обобщенные критерии оптимальности:

$$U_1(x, r) = F(x) + rI_1(x) \quad (7-11)$$

и

$$U_2(x, r) = F(x) + rI_2(x). \quad (7-12)$$

В алгоритме оптимизации используется последовательность положительных чисел $\{r_k\}$, $k=1, 2, \dots$, монотонно сходящаяся к нулю. В качестве начальной точки выбирается произвольная внутренняя точка x^0 области D . Она является исходной для поиска минимума обобщенного критерия оптимальности $U(x, r_1)$. Минимум x^1 функции $U(x, r_1)$ выбирается в качестве начального приближения для поиска минимума функции $U(x, r_2)$ и т. д. Последовательность полученных таким образом безусловных минимумов x^k сходится к оптимальному решению исходной задачи.

В качестве примера рассмотрим задачу поиска минимума критерия оптимальности, заданного квадратичной функцией

$$F(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2) - x_1 - 2x_2$$

Таблица 7-2

k	r_k	x_1^k	x_2^k	Число итераций	Точность $ \nabla U(x^k) $	Число вычислений	
						функции	градиента
1	0,5	0,8238	0,6848	4	0,0002	28	5
2	0,25	0,7876	0,8246	5	0,00001	35	6
3	0,1	0,7652	0,9549	3	0,0007	30	4
4	0,01	0,7646	1,0483	4	0,0041	30	5
5	0,001	0,7635	1,0578	5	0,0004	32	6

при наличии линейных ограничений

$$R_1(x_1, x_2) \equiv 2x_1 + 3x_2 - 6 \leq 0; R_2(x_1, x_2) \equiv x_1 + 4x_2 - 5 \leq 0, x_1 \geq 0, x_2 \geq 0.$$

Геометрическое представление задачи приводится на рис. 7-3. Для решения задачи методом барьерных функций воспользуемся критерием оптимальности (7-12). Результаты поиска оптимума с помощью метода переменной метрики представлены в табл. 7-2.

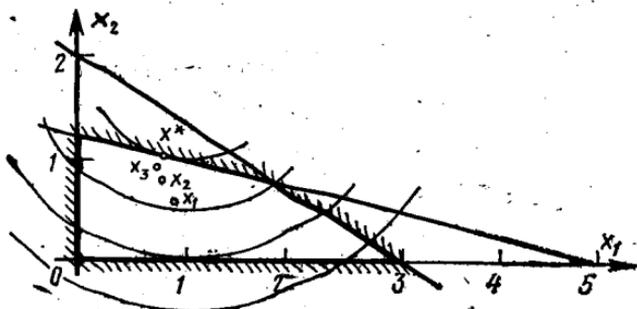


Рис. 7-3. Пример поиска экстремума с помощью метода барьерных функций.

Оптимальное значение $x^* = (0,764; 1,05)$ достигается после решения пяти задач безусловной оптимизации при различных значениях параметра r .

Указанный метод гарантирует, что каждая следующая итерация не выйдет за пределы допустимой области. При уменьшении r уменьшается влияние штрафного члена и возрастает влияние исходного критерия оптимальности. Поэтому последовательность функций $U(x, r_k)$ дает сколь угодно точное (для достаточно большого номера k) приближение к локальному экстремуму. Если искомый экстремум лежит внутри допустимой области G , то решение может быть получено после нескольких первых значений параметра r .

Наибольшие сложности при решении задачи возникают в тех случаях, когда оптимум расположен на границе допустимой области. Поскольку при приближении к границе значения обобщенного критерия оптимальности, по существу, определяются барьерной функцией, экстремум не всегда может быть вычислен с заданной степенью точности. Чтобы избежать этот недостаток, производится так называемая регуляризация допусти-

мой области. Она основана на ослаблении ограничений на начальных этапах поиска. Иными словами, ограничения вида (7-1) заменяются неравенствами:

$$R_i(x) \leq r, \quad i=1, 2, \dots, m, \quad (7-13)$$

где r — некоторое положительное число.

Построенный для этих ограничений обобщенный критерий оптимальности с $I(x) = I_2(x)$ имеет вид:

$$U_s(x, r) = F(x) - r^2 \sum_{i=1}^m [f_i(x) - r]^{-1}. \quad (7-14)$$

С его помощью удастся избежать трудностей поиска вблизи границы, однако некоторые ограничения при этом могут быть нарушены.

Решение задач безусловной оптимизации для сходящейся к нулю последовательности $\{r_k\}$, $k=1, 2, \dots$ приводит к постепенному уменьшению этих нарушений. После достаточного числа итераций (которое определяется в процессе решения задачи) значение погрешности, с которой выполняются ограничения, оказывается в пределах требуемой точности.

Основная трудность применения метода барьерных функций обусловлена тем, что начальная точка x^0 должна находиться внутри допустимой области. Поиск этой точки иногда представляет собой задачу, по сложности сравнимую с задачей определения экстремума. В следующем параграфе рассматривается один из подходов к ее решению.

В тех случаях, когда критерий оптимальности или функции ограничений определены не для всех, а лишь для некоторых значений переменных, барьерные функции представляют собой единственный способ исключения ограничений.

При практических расчетах иногда используется комбинированный подход, учитывающий особенности барьерных и штрафных функций. Обобщенный критерий оптимальности включает в себя слагаемые в виде барьерных функций для ограничений типа неравенств, штрафных функций — для равенств. Например, обобщенный критерий оптимальности может иметь вид логарифмически квадратичной функции

$$W(x, r) = F(x) - r \sum_{i=1}^m \ln[-R_i(x)] + r^{-1} \sum_{i=1}^k g_i^2(x), \quad (7-15)$$

где первая сумма получена методом барьерных, а вторая — методом штрафных функций. Функция $W(x, r)$ удобна также тем, что последовательность задач безусловной оптимизации зависит от одного параметра, который удовлетворяет требованиям, предъявляемым к последовательности значений штрафных коэффициентов $\{r_k\}$ и $\{t_k\}$.

Построенные в результате применения методов штрафных функций обобщенные критерии оптимальности объединяют исходный критерий оптимальности и ограничения модели. Поэтому сложность вычисления каждого значения обобщенного критерия оптимальности существенно возрастает. Это приводит к необходимости эффективного использования каждого вычисленного значения критерия оптимальности для приближения к оптимуму.

7-2. ОСОБЕННОСТИ ПОИСКА С ИСПОЛЬЗОВАНИЕМ ШТРАФНЫХ ФУНКЦИЙ

Полученная в результате преобразования с помощью методов штрафных функций последовательность задач безусловной оптимизации обладает рядом свойств, которые позволяют ускорить решение общей задачи. Рассмотрим лишь некоторые особенности методов штрафных функций, позволяющие осуществить:

вхождение в допустимую область, когда обобщенный критерий оптимальности формируется с помощью барьерных функций;

экстраполяцию последовательности безусловных оптимумов x^k ;

выбор последовательности штрафных t_k и барьерных r_k коэффициентов.

Использование специальных приемов, разработанных для осуществления перечисленных этапов, позволяет повысить скорость сходимости методов и сократить время счета на ЭВМ.

Вхождение в допустимую область. При достаточно сложной геометрии области допустимых значений параметров даже отыскание допустимого решения (т. е. набора значений параметров, удовлетворяющего всем ограничениям) представляет значительные трудности. Чтобы отыскать допустимое решение, необходим алгоритм, позволяющий из произвольной точки n -мерного пространства попасть в допустимую область.

Будем предполагать, что ограничениями задачи служат неравенства (7-1). Случайно выбранная начальная точка удовлетворяет лишь части из них. Пусть $T(x^0)$ — множество тех ограничений, которые выполняются в точке x^0 как строгие неравенства, т. е.

$$T(x^0) = \{i | R_i(x^0) < 0, i=1, 2, \dots, m\} \quad (7-16)$$

и $S(x^0)$ — множество тех ограничений, которые либо не удовлетворяются в точке x^0 , либо удовлетворяются как равенства

$$\hat{S}(x^0) = \{i | R_i(x^0) \geq 0, i=1, 2, \dots, m\}. \quad (7-17)$$

Поиск должен осуществляться таким образом, чтобы, не нарушая уже выполняющихся ограничений, перейти к некоторой другой точке, в которой удовлетворяются хотя бы некоторые из ограничений, входящих в множество $S(x^0)$. Это достигается путем минимизации функций $R_i, i \in S(x^0)$ при условии выполнения неравенств (7-16).

Обобщенный критерий оптимальности, соответствующий сформулированной задаче, имеет вид:

$$U_k(x, r) = R_i(x) + rI_T(x), \quad (7-18)$$

где $R_i(x)$ — одно из ограничений, входящих в множество $S(x^0)$, а $I_T(x)$ — барьерная функция вида (7-8) или (7-9) относительно ограничений из множества $T(x^0)$. Минимизация выражений $U(x, r_k)$ для последовательности $\{r_k\}, k=1, 2, \dots$, сходящейся к нулю, приводит к выполнению ограничения $R_i(x) < 0$ в некоторой точке \bar{x} . Эта точка снова принимается за x^0 , а множество $T(x^0)$ дополняется ограничением R_i , исключенным из $S(x^0)$.

Возможно даже одновременное уменьшение всех функций $R_i, i \in S$. Для этого формируется обобщенный критерий оптимальности

$$U_S(x, r) = \sum_{i \in S} R_i(x) + rI_T(x). \quad (7-19)$$

В процессе его оптимизации производится проверка знаков функций $R_i, i \in S$. Как только одна из функций меняет знак, она переходит во второе слагаемое, т. е. формируется соответствующий компонент барьерной функции $I_T(x)$. Оптимизация нового критерия оптимальности производится из последней вычисленной точ-

ки, которая служит начальным приближением для поиска. Метод гарантирует вхождение в допустимую область, если в ней есть хотя бы одна внутренняя точка.

Недостатком вспомогательного критерия $U_B(x, r)$ является игнорирование влияния критерия оптимальности исходной задачи оптимального проектирования при поиске допустимого решения вхождения в допустимую область. Функцию $F(x)$ можно включать в обобщенный критерий оптимальности (7-19), однако в случае большого числа ограничений такой подход не может привести к «блужданию» вокруг допустимой области, что приводит к возрастанию общего времени счета.

Процедура вхождения в допустимую область не предусматривает проверки факта существования хотя бы одного допустимого решения. Однако на практике заранее предполагается существование хотя бы одного изделия с параметрами, находящимися в заданных пределах.

Экстраполяция по штрафным коэффициентам. В процессе решения последовательности задач безусловной оптимизации для обобщенных критериев оптимальности иногда удается по нескольким начальным задачам для параметров r_1, r_2, \dots, r_m довольно точно определить расположение оптимума для обобщенного критерия, зависящего от r_{p+1} . Эта идея положена в основу некоторого приема, позволяющего ускорить поиск оптимального решения с помощью методов штрафных функций.

Пусть $W(x, r)$ — обобщенный критерий оптимальности, определенный формулой (7-15), при $r_1 > r_2 > \dots > r_m > 0$ имеет минимумы в точках x^1, x^2, \dots, x^m , причем на каждом r , точка минимума единственна. Экстраполяция минимальных значений может преследовать две цели: построить предполагаемое значение минимума для критерия $W(x^1, r_{m+1})$ или предсказать расположение предела последовательности x^1, x^2, \dots, x^m .

Наиболее естественный способ достижения этих целей — построение прогнозирующего многочлена, зависящего от скалярной переменной r . Для каждой координаты вектора $x = (x_1, x_2, \dots, x_n)$ строится полином

$$x_i = \sum_{j=0}^l a_{ij} r^j, \quad (7-20)$$

степень которого определяется номером итерации и точностью прогноза. При этом возможны два способа по-

строения экстраполяционного полинома (7-20). Один — по всем предыдущим значениям r_k , когда степень в полиноме равна номеру итерации без единицы, и другой — по нескольким последним итерациям с сохранением степени полинома $l = \text{const}$ (например, экстраполяция с помощью кубического полинома).

Неопределенные коэффициенты a_{ij} вычисляются с помощью систем линейных уравнений (по числу координат вектора x в пространстве параметров). Порядок каждой системы превосходит степень аппроксимирующего полинома на единицу. Для координаты x_i соответствующая система записывается в виде

$$\begin{aligned} x_i(r_{m-l}) &= a_{i0} + a_{i1}r_{m-l} + \dots + a_{il}r_{m-l}^l \\ x_i(r_{m-l+1}) &= a_{i0} + a_{i1}r_{m-l+1} + \dots + a_{il}r_{m-l+1}^l \\ x_i(r_m) &= a_{i0} + a_{i1}r_m + \dots + a_{il}r_m^l \end{aligned} \quad (7-21)$$

На начальных этапах решения задачи (при малых m) более целесообразно прогнозировать оптимум критерия $W(x, r_{m+1})$ и выбирать прогнозируемую точку в качестве начального приближения поиска. После достаточного числа итераций удастся довольно точно экстраполировать оптимум исходной задачи.

Подстановка значения $r=0$ приводит к тому, что в (7-20) остается только свободный член a_{i0} . Поэтому задача экстраполяции оптимального решения фактически сводится к вычислению коэффициента a_{i0} . Отсюда следует, что, определив такой порядок вычисления коэффициентов полинома, который позволяет оценить в первую очередь коэффициент a_{i0} , удастся получить удобную формулу экстраполяции. Формула также упрощается при дополнительном предположении, что последовательность r_k представляет собой убывающую геометрическую прогрессию с начальным членом r_1 и знаменате-

лем $q < 1$. Она может быть получена путем подстановки в систему (7-21) выражения

$$r_k = r_1 q^{k-1} \quad (7-22)$$

для коэффициентов обобщенного критерия оптимальности (7-15).

Ускорение сходимости методов штрафных функций. Наименее разработанной проблемой, возникающей при решении задачи оптимального проектирования с помощью штрафных функций, является выбор последовательностей $\{r_k\}$ и $\{t_k\}$. В (7-15) для обобщенного критерия оптимальности задача оптимизации решается путем выбора некоторой последовательности $\{r_k\}$, $k=1, 2, \dots$, сходящейся к нулю. В некоторых случаях последовательность можно задать в форме геометрической прогрессии (7-22). При этом проблема определения последовательности $\{r_k\}$, $k=1, 2, \dots$ сводится к выбору знаменателя прогрессии.

Пусть по некоторому признаку выделен класс задач оптимального проектирования, например задачи проектирования асинхронных двигателей. Для этого конкретного класса можно сформулировать задачу выбора оптимального, в некотором смысле, алгоритма с обобщенным критерием $W(x, r)$ [см. (7-15)]. Для заданной точности поиска ε и знаменателя геометрической прогрессии q решение задачи оптимального проектирования достигается за N -итераций. Поэтому каждой паре величин ε и q можно поставить в соответствие целое число

$$N = N(\varepsilon, q). \quad (7-23)$$

Эта величина является средним числом итераций для задач рассматриваемого класса при фиксированных ε и q .

Задавая требуемую точность поиска ε_0 , можно определить значение q , при котором достигается минимум функции $N(\varepsilon_0, q)$.

Отличие этой задачи от задач, решаемых методами однопараметрической оптимизации, состоит в следующем. Критерий оптимальности $F(x)$, зависящий от скалярной переменной x , является детерминированной функцией, и поэтому каждому значению аргумента x соответствует единственное значение функции. В свою очередь $N(\varepsilon_0, q)$ является случайной функцией от аргумента q , т. е. при каждом q возможен целый набор значе-

ний *N*. Оптимизация таких функций осуществляется методами стохастической аппроксимации и составляет самостоятельный предмет исследования [33].

7.3. МОДИФИКАЦИЯ ШТРАФНЫХ ФУНКЦИЙ ДЛЯ ДИСКРЕТНОЙ ОБЛАСТИ ПОИСКА

Методы штрафных функций дают хорошие результаты, если переменные задачи — конструктивные параметры — являются непрерывными. При наличии параметров, изменяющихся дискретно, значительно усложняются алгоритмы поиска. Поэтому поиск экстремума в дискретной области производится обычно путем предварительного решения задачи в предположении непрерывности параметров. Полученное решение округляется и интерпретируется на основе «здравого смысла». При этом фактически производится округление до ближайшего дискретного значения.

В целом ряде задач такой подход может привести к результатам, весьма далеким от оптимума. Существуют также задачи, которые просто не удается математически сформулировать без предположения дискретности (или даже целочисленности) некоторых переменных; поскольку с ошибкой округления может быть потеряно не только эффект оптимизации, но и нарушение условия допустимости решения.

Формирование штрафной функции. При дискретных конструктивных параметрах вектор x в задаче оптимального проектирования с дискретными параметрами может быть представлен в виде

$$x = \begin{pmatrix} x^c \\ x^d \end{pmatrix}, \quad x^c \in R^c, \quad x^d \in R^d,$$

где R^c — пространство изменения непрерывных параметров, а R^d — пространство изменения дискретных параметров.

Обобщенный критерий оптимальности имеет вид:

$$V(x, r, \rho) = F(x) + rI(x) + \rho I(x^d), \quad (7-24)$$

где $I(x)$ является барьерной функцией, например (7-9), а выражение $I(x^d)$ является штрафной функцией, соответствующей условию дискретности. Она должна обладать следующим свойством:

$$I(x^d) = \begin{cases} 0, & \text{если } x^d \in R^d; \\ \mu > 0, & \text{если } x^d \notin R^d. \end{cases} \quad (7-25)$$

Таким образом, задача формирования штрафной функции $V(x, r, \rho)$ сводится к построению такого выражения для $I(x^d)$, которое удовлетворяет равенству (7-23). Очевидно, что функция $I(x^d)$ многоэкстремальна, если имеется несколько возможных дискретных значений вектора x^d . Отсюда следует, что штрафная функция $V(x, r, \rho)$ также многоэкстремальна и ее оптимизация заключается в поиске глобального экстремума.

Определим:

$$I(x^d, \beta) = \sum_{j \in D} 4q_j (1 - q_j)^3, \quad (7-26)$$

где

$$q_j = (x_j - z_{jl}) / (z_{ju} - z_{jl}) \quad (7-27)$$

и $z_{jl} \leq x_j$, $z_{ju} \geq x_j$ являются двумя соседними дискретными точками. Максимальное значение каждого слагаемого в (7-26) равно единице, и функция $I(x^d, \beta)$ непрерывна в точках дискретности для всех $\beta \geq 1$ вместе со своими первыми производными.

Преобразованный по формуле (7-23) критерий задачи оптимального проектирования с частично дискретными переменными минимизируется для последовательности коэффициентов β_k , r_k и ρ_k так, чтобы полученная последовательность безусловных экстремумов сходилась к решению исходной задачи. На практике такая последовательность содержит от 5 до 10 элементов.

Применение метода штрафных функций для поиска экстремума в дискретной области связано с решением проблемы формирования процедуры обхода локальных экстремумов, которые не являются глобальными.

Процедура обхода локальных экстремумов. В процессе поиска оптимального решения задачи с дискретными переменными возникают локальные оптимумы двух типов. Одни соответствуют некоторому допустимому, но не оптимальному решению, другие — недопустимым значениям параметров (рис. 7-4). Такая ситуация возникает обычно, когда одна из допустимых дискретных точек расположена близко к ограничивающей поверхности (7-29). Чтобы из такой точки перейти ближе к глобальному оптимуму, следует увеличить коэффициент r_k .

В общем случае, если обобщенный критерий оптимальности на k -й итерации имеет вид:

$$V(x, r_k, \rho_k) = F(x) + r_k I(x) + \rho_k J_k(x),$$

процедура выхода из локального оптимума состоит из двух циклов. В первом цикле значения r_k увеличиваются, а значения ρ_k уменьшаются. Это позволяет осуществить выход из окрестности точки локального оптимума. В следующем цикле уменьшается значение r_k и увеличи-

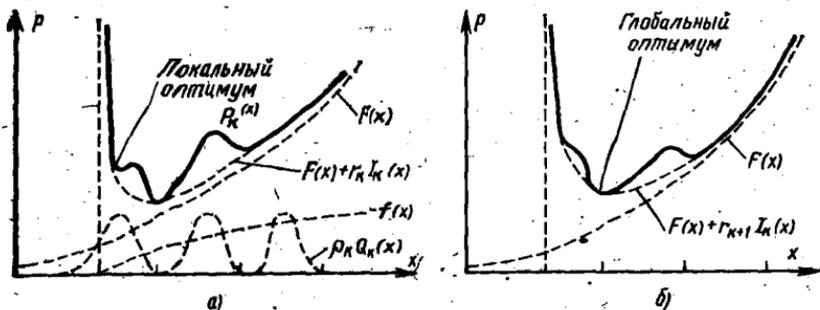


Рис. 7-4. Процедура обхода локальных экстремумов при применении метода штрафных функций для поиска глобального экстремума в дискретной области.

Локальный экстремум в допустимой (а) и недопустимой (б) области значений параметров.

вается ρ_k , чтобы достигнуть хорошей точности результата.

Пусть локальный оптимум достигнут при значениях параметров ρ_k и r_k . Тогда:

1. В первом цикле первого шага процедуры обхода локального экстремума (R1.1) следует положить $r_{R1.1} = r_{k-2}$ и $\rho_{R1.1} \approx 2\rho_{k-1}$.

2. Во втором цикле первого шага процедуры обхода (R1.2) следует положить $r_{R1.2} = r_{k-1}$ и $\rho_{R1.2} \approx 2\rho_k$.

Если это не приводит к новой точке оптимума, первый и второй шаг повторяются для следующих этапов процедуры обхода локальных экстремумов. В частности, на втором шаге имеем:

$$r_{R2.1} = r_{k-3}, \rho_{R2.1} \approx 2\rho_k \text{ и } r_{R2.2} = r_{k-2}, \rho_{R2.2} \approx 2\rho_{k+1}.$$

После достаточного числа таких процедур удастся найти точку дискретной допустимой области с лучшим значением обобщенного критерия оптимальности.

Указанная процедура обхода дает хорошие результаты на тестовых примерах и в практических расчетах. При неблагоприятной геометрии допустимой области

число шагов может колебаться от четырех до шести, но обычно достаточно от одного до трех шагов. Однако без полного перебора невозможно определить, соответствует ли полученный результат глобальному или локальному оптимуму.

Пример. Проектирование щита для люка. В качестве примера применения метода штрафных функций рассмотрим задачу оптимального проектирования щита люка [17]. Отверстие прямоугольной формы шириной $l_0 = 600$ м (рис. 7-5,а) должно быть покрыто пол-

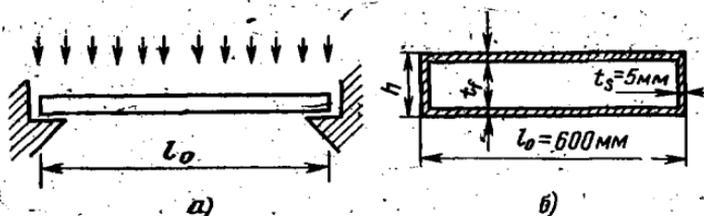


Рис. 7-5. Пример. Проектирование щита.

ми брусками из алюминия длиной l_0 и шириной в 60 см (рис. 7-5,б). Покрывтие должно выдерживать максимальную удельную нагрузку до 1000 кг/м^2 .

В качестве переменных задачи (конструктивных параметров) приняты следующие: $x_1 = t_s$ — толщина стенок и $x_2 = h$ — высота бруска.

Приняты также следующие предположения:

а) коэффициент Пуассона $\nu = 0,3$;
 б) материал обладает линейной упругостью с модулем Юнга $E = 7 \cdot 10^5 \text{ кг/см}^2$;

в) остаточная деформация материала должна отсутствовать.

При построении модели формируются следующие ограничения:

1. На максимальное касательное напряжение τ_{max} . Полагая, что максимум поперечной силы $Q = 1800 \text{ кг}$, для максимального касательного напряжения имеем:

$$\tau = \frac{Q}{2t_s h} \leq \tau_{max} = 450 \text{ кг/см}^2.$$

2. На максимальное изгибающее напряжение $\sigma_{b_{max}}$. На основе

предварительного анализа конструкции получен максимальный изгибающий момент $M = 2,7 \cdot 10^4 \text{ кг/см}$. Для изгибающего напряжения, действующего на стенки щита, имеем:

$$\sigma_b = \frac{M}{bt_s h} \leq \sigma_{b_{max}} = 700 \text{ кг/см}^2.$$

3. На максимальный прогиб $\delta_{max} \rho l_0$, где $\rho = 0,025$. Величина прогиба определяется по формуле

$$\delta = 0,0158 \frac{q_1 l_0^4}{EI} \leq \delta_{max},$$

где $q_1 = q_0 b$, и для величины l принята аппроксимация

$$l = \frac{1}{2} b t_f / h^2.$$

4. На изгибающее напряжение в стенках. Изгибающее напряжение в стенках выражается формулой

$$\sigma_k = \frac{4n^2 E}{12(1-\nu^2)} \left(\frac{t_f}{b} \right)^2 \approx \frac{E t_f^2}{1000} \geq \sigma_b.$$

В результате задача оптимального проектирования формируется с учетом следующих ограничений:

$$f_1 = 1,0 - \frac{\tau_{max}}{\tau} \leq 0; \quad f_2 = 1,0 - \frac{p l_0}{\delta} \leq 0;$$

$$f_3 = 1,0 - \frac{\sigma_{b_{max}}}{\sigma} \leq 0; \quad f_4 = 1,0 - \frac{\sigma_k}{\sigma_b} \leq 0;$$

$$h \geq 0; \quad t_f \geq 0.$$

Критерием оптимальности является масса щита на единицу длины¹

$$F(t_f, h) = h + 120 t_f.$$

Графическое решение этой задачи без учета дискретности переменных представлено на рис. 7-6. Минимальная масса щита составляет $F=101$ кг. Она достигается при $h=86$ см и $t_f=0,026$ см.

Для решения этой же задачи в дискретной постановке, когда переменные t_f и h принимают значения:

$$t_f = 0,1; 0,2 \text{ см} \dots$$

$$h = 15; 25; 40; 60 \text{ см} \dots$$

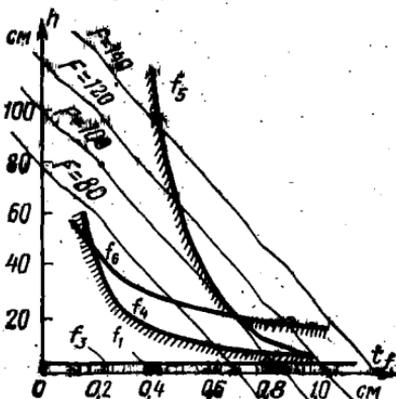


Рис. 7-6. Допустимая область и критерий оптимальности задачи проектирования щита для люка.

используется штрафная функция (7-24). Поиск оптимума состоит из пяти последовательных итеративных процедур, соответствующих различным значениям коэффициентов при штрафных функциях. Кроме того, 3 раза была применена процедура выхода из локального экстремума, причем в первых двух локальных экстремумах оказалось

¹ Строго говоря, это выражение представляет собой массу с точностью до коэффициентов пропорциональности, соответствующего плотности используемого материала.

достаточно одношаговой процедуры, а в третьем — была использована двухшаговая процедура. Последовательность итераций приводится в табл. 7.3. Оптимальное решение: $t_f = 0,7$ см и $h = 25$ см.

Таблица 7.3

Номер итерации	r_k	p_k	Результаты расчетов				Число вычислений критерия
			t_f	h	$\min V$	$F(x)$	
0	—	—	0,700	40,0	130,20	124,00	—
1	1,9746	0	0,644	32,09	121,35	109,37	37
2	$0,9873 \cdot 10^{-1}$	0,17	0,634	26,78	104,91	102,90	79
3	$0,4937 \cdot 10^{-2}$	0,78	0,630	25,98	102,25	101,86	37
4	$0,2468 \cdot 10^{-3}$	3,49	0,628	25,98	102,82	102,73	23
5	$0,1234 \cdot 10^{-4}$	15,72	0,634	25,10	108,22	108,19	34
R1. 1	$0,4937 \cdot 10^{-2}$	7,29	0,627	26,26	105,14	104,65	149
R2. 1	$0,9873 \cdot 10^{-1}$	14,58	0,646	25,09	112,21	109,85	101
R3. 1	1,9746	29,16	0,700	25,25	124,57	109,50	85
R3. 2	$0,9873 \cdot 10^{-1}$	131,24	0,700	25,00	109,79	109,00	36

СПИСОК ЛИТЕРАТУРЫ

1. Аоки М. Введение в методы оптимизации. — М.: Наука, 1977. — 344 с.
2. Аветисян Д. А., Бертинов А. И. Динамическое программирование расчета оптимальных электрических машин на ЦВМ. — Электричество, 1966, № 11, с. 45—51.
3. Аветисян Д. А., Соколов В. С., Хан В. Х. Оптимальное проектирование электрических машин на ЭВМ. — М.: Энергия, 1976. — 206 с.
4. Адлер Ю. П., Маркова Е. В., Грановский Ю. В. Планирование эксперимента при поиске оптимальных условий. — М.: Наука, 1976. — 279 с.
5. Баргелъ И., Маркс Л. Оптимальное проектирование механических систем при противоречивых критериях. — Конструирование и технология машиностроения, 1974, № 1, с. 341—348.
6. Вайнштейн Д. И. Методы оптимального проектирования. — М.: Советское радио, 1976. — 216 с.
7. Визин М. Стохастическая аппроксимация. — М.: Мир, 1979. — 275 с.
8. Вилд А. Последовательный анализ. — М.: Физматгиз, 1966. — 365 с.
9. Об одной модели теплового расчета элементов криогенных электрических машин в режиме криостатирования. / Воскресенский А. С., Геминтерн В. И., Михайлова С. П., Скляр Е. С. — Труды ВНИИЭМ, 1976, т. 47, с. 27—33.
10. Вознич М., Хейдт Г. Гиперконический случайный поиск. — Динамические системы и управление, 1972, № 1, т. 94, с. 87—96.
11. Воронский Г. П. Комплексный подход к созданию единых серий электромеханических изделий. — Электротехника, 1978, № 9, с. 1—3.
12. Критерий оптимальности при проектировании серий асинхронных двигателей общего назначения / Воскресенский А. П., Емельянова Г. Н., Череванов В. Е., Стрельбицкий Э. К. Тезисы докладов на международной конференции «Исследование и развитие асинхронных двигателей». — Владимир: Информэлектро, 1978, с. 24—25.
13. Воскресенский А. П., Мазия Л. В., Сорокер Т. Г. Основные принципы системы автоматизированного проектирования асинхронных двигателей. — Электротехника, 1978, № 9, с. 14—17.
14. Гаспарский В. Практикологический анализ проектно-конструкторских разработок. — М.: Мир, 1978. — 172 с.
15. Гельфанд И. М., Цеглин М. Л. Принцип нелокального поиска в системах автоматической оптимизации. — Доклады АН СССР, т. 137, 1961, № 2, с. 295—298.

16. Гемингер В. И., Каган Б. М. Методы оптимального проектирования электротехнических изделий. — М.: Информэлектро, 1975. — 55 с.
17. Гисволд К., Мое И. Метод смешанного целочисленного программирования и его применение к задачам проектирования. — Динамические системы и управление, 1972, № 2, т. 94, с. 43—58.
18. Глушков В. М. О диалоговом методе решения оптимизационных задач. — Кибернетика, 1975, № 4, с. 3—7.
19. Гринев В. Б., Филиппов А. П. Оптимизация элементов конструкций по механическим характеристикам. — Киев: Наукова думка, 1975. — 294 с.
20. Данилейко С. Е., Каган Б. М., Шахуняц Т. Г. Проекционный градиентный метод решения задач оптимального проектирования. — Автоматика и вычислительная техника, 1976, № 3, с. 31—36.
21. Данилейко С. Е., Каган Б. М., Шахуняц Т. Г. О дискретных алгоритмах в задачах оптимизации. — Автоматика и вычислительная техника, 1977, № 2, с. 39—42.
22. Дворянкин А. М., Половинкин А. И., Соболев А. Н. Методы синтеза технических решений. — М.: Наука, 1977. — 103 с.
23. Диалоговые системы моделирования/ Под ред. В. В. Пирогова. — Рига: Зинатне, 1971. — 165 с.
24. Евушенко Ю. Г. Методы поиска глобального экстремума. Исследование операций. — М.: ВЦ АН СССР, 1974, — вып. 4, с. 76.
25. Зенер К. Геометрическое программирование и техническое проектирование. — М.: Мир, 1973. — 111 с.
26. Каган Б. М. Решение задач нелинейного программирования на ЦВМ. — В кн.: Применение вычислительной техники для автоматизации. — М.: Машгиз, 1961, с. 140—164.
27. Каган Б. М. Научные и технические проблемы кибернетических систем автоматизации проектирования. — В кн.: Материалы семинара «Кибернетические системы автоматизации проектирования». — М.: МДНТП, 1973, с. 3—14.
28. Каган Б. М., Тер-Микаэлян Т. М. Решение инженерных задач на цифровых вычислительных машинах. — М.: Энергия, 1964. — 592 с.
29. Каган Б. М., Даниленко С. Е. Вопросы архитектуры систем оптимального проектирования. — В кн.: Материалы семинара «Кибернетические системы автоматизации проектирования». — М.: МДНТП, 1973, с. 132—136.
30. Каган Б. М., Даниленко С. Е. Применение метода случайного поиска с обучением при оптимальном проектировании асинхронных двигателей. — В кн.: Автоматика и вычислительная техника, — Рига: Зинатне, 1966, вып. 13, с. 169—172.
31. Основные проблемы автоматизации проектирования серий электрических машин/ Каган Б. М., Бердичевский А. А., Даниленко С. Е., Розенкоп В. Д. — В кн.: Применение вычислительной техники в электротехнической промышленности. — М.: 1971, с. 313—319.
32. Каплинский А. И., Красненкер А. С. О формировании диалоговых алгоритмов векторной оптимизации. — Автоматика и вычислительная техника, 1977, № 5, с. 32—37.
33. Каплинский А. И., Красненкер А. С., Назин А. В. Обучение принципу свертывания в задаче векторной оптимизации. — Автоматика и вычислительная техника, 1978, № 4, с. 43—49.
34. Моцкус И. Б. Многоэкстремальные задачи в проектировании. — М.: Наука, 1967. — 215 с.

35. Мозгова Э. А. Об одном методе поиска минимума функций при наличии ограничений. — Автоматика и телемеханика, 1962, № 12, с. 51—58.

36. Налимов В. В. Теория эксперимента. — М.: Наука, 1971. — 207 с.

37. Пашас Н., Моранди Р. Усовершенствованный алгоритм прямого поиска для задач математического программирования. — Конструирование и технология машиностроения, 1975, № 4, с. 208.

38. Подвальный Л. Д., Пугачева Л. И. Метод сопряженных случайных направлений в задачах овражного типа. — Кибернетика, 1976, № 4, с. 22—29.

39. Подшивковский В. В., Газрилов В. М. Оптимизация по последовательно применяемым критериям. — М.: Советское радио, 1975. — 192 с.

40. Подшивковский В. В. Многокритериальные задачи с упорядоченными по важности критериями. — Автоматика и телемеханика, 1976, № 11, с. 118—127.

41. Алгоритмы оптимизации проектных решений/ Половинкин А. И., Грудачев В. Т. и др. — М.: Энергия, 1976. — 265 с.

42. Полак Э. Численные методы оптимизации. Единый подход. — М.: Мир, 1974. — 376 с.

43. Прагер В. Основы теории оптимального проектирования конструкций. — М.: Мир, 1977. — 109 с.

44. Шеневичкий Б. И., Данилин Ю. М. Численные методы в экстремальных задачах. — М.: Наука, 1975. — 319 с.

45. Гаврилин Л. А., Рина К. К., Тарасович Г. С. Адаптация случайного поиска. — Рига: Знание, 1978. — 242 с.

46. Растрингин Л. А. Системы экстремального управления. — М.: Наука, 1974. — 630 с.

47. Оболь М. И. Численные методы Монте-Карло. — М.: Наука, 1973. — 318 с.

48. Сорокер Г. Г., Каган Б. М. О применении ЭВМ для работ серий электрических машин. — Вестник электронпромышленности, 1968, № 9, с. 96—118.

49. Оптимальное проектирование серий асинхронных двигателей на ЦВМ/ Сорокер Г. Г., Воскресенский А. П., Даниленко С. Е., Мёрдвинюв Ю. В. — Труды ВНИИЭМ, 1976, т. 47, с. 5—12.

50. Старосельский Л. А., Шелудько Г. А., Кантор Б. Я. Об одной реализации метода оврагов с адаптацией величины овражного шага по экспоненциальному закону. — Вычислительная математика и математическая физика, т. 8, 1968, № 5, с. 1161—1167.

51. Строингин Р. Г. Численные методы в многоэкстремальных задачах. — М.: Наука, 1978. — 240 с.

52. Тарасенко Г. С. Сходимость адаптивного алгоритма случайного поиска. — Кибернетика, 1977, № 5, с. 88—90.

53. Фианко А., Мак-Корриик Г. Нелинейное программирование. Методы последовательной безусловной оптимизации. — М.: Мир, 1970. — 240 с.

54. Химмельблау Д. Анализ процессов статистическими методами. — М.: Мир, 1973. — 957 с.

55. Химмельблау Д. Прикладное нелинейное программирование. — М.: Мир, 1975. — 535 с.

56. Цыпкин Я. З. Алгоритмы динамической адаптации. — Автоматика и телемеханика, 1972, № 1, с. 72—81.

57. Численные методы условной оптимизации/ Под ред. Гилл Ф. и Мюррей У. — М.: Мир, 1977. — 290 с.
58. Шор Н. З. Методы отсечения с растяжением пространства для решения задач выпуклого программирования. — Кибернетика, 1977, № 1, с. 94—95.
59. Asaadi J. A computational comparison of some nonlinear programs. — Math. Program., 1973, vol. 4, № 2, p. 144—154.
60. Annocost R., Fiacco A. Sensitivity analysis for parametric nonlinear programming, using penalty methods. — US Dep. Commer. Nat. Bur. Stand. Spec. Publ., 1978, № 502, p. 261—269.
61. Avriel M. Nonlinear programming. Analysis and methods. — Englewood Cliffs, Prentice Hall, 1976. — 322 p.
62. Bandler J. W., Chu W. I. Nonlinear programming using least pth optimization with extrapolation, Internat. — Systems Sci., 1976, vol. 7, № 11, p. 1239—1248.
63. Baier H. Optimization of elastic structures by mathematical programming methods. — Optimization and operations research (Proc. Conf. Oberwolfach, 1975), Berlin, 1976, p. 9—22.
64. Beckman M., Künzi H. P. (eds). Multiple Criteria Problem Solving. — N. Y.: 1978. — 567 p.
65. Biggs M. C., Laughton M. A. Optimal electric power scheduling a large nonlinear programming test problem solved by recursive quadratic programming. — Math. Program., 1977, vol. 13, № 2, p. 167—182.
66. Bightler C. S., Phillips D. T. Applied geometric programming. — N. Y.: Wiley, 1976. — 274 p.
67. Broyden C. G. Quasi-Newton methods and their application to a function minimization. — Math. of Comp., 1970, vol. 24, № 11, p. 647—656.
68. Cesario F. I. An interesting test problem for nonlinear programming algorithms optimizations. — Operations Research, 1975, vol. 23, № 3, p. 599—601.
69. Dimitry V., Moga S. Testing of efficiency of the sequential optimization technique. — Econ. Comput. and Econ. Cybern. Studies a. Research, 1973, vol. 4, p. 55—62.
70. Fletcher R. (ed.). Optimization. — N. Y.: Acad. Press, 1969. — 412 p.
71. Garcia P., Mangasarjan P. Superlinearly convergent quasi-Newton algorithms for nonlinearly constrained optimization problems. Math. — Program., 1976, vol. 11, № 1, p. 1—13.
72. Greenstadt I. A quasi-Newton method with no derivatives. — Math. of Comp., 1972, vol. 26, № 117, p. 146—166.
73. Grossman Ch., Kaplan A. A. Penalty method in nonlinear programming (survey). — Math. Operationsforsch. Statist. ser. — Optimization, 1977, vol. 8, № 2, p. 281—298.
74. Goldstein A. A. Optimization of Lipschitz continuous functions. — Math. Program., 1977, vol. 3, № 1, p. 14—22.
75. Han S. P. A globally convergent method for nonlinear programming. — Optimiz. Theory and Appl., 1977, vol. 22, № 3, p. 297—309.
76. Holenbalken B. Simplicial decomposition in nonlinear programming algorithms. — Math. Program., 1977, vol. 13, № 1, p. 49—68.

77. Lev O. E. On the solution of a nonlinear programming problem by decomposition.—*Optimiz. Theory Appl.*, 1977, vol. 22, № 1, p. 31—34.

78. Lin I. G. Multiple-objective optimization: proper equality constraints (PEC) and maximization of index vectors.—*Multicr. Decis. and Econ.*—London: 1976, p. 68—79.

79. Manna Z. Theory of computation.—N. Y.: 1974.—346 p.

80. Mockus I. On the Bayes methods for seeking the extremal point.—*Proc. IFAC 6th World Congr.*—Boston: 1975, pt. 1, p. 297—301.

81. Mayne D., Polak E. Feasible directions algorithms for optimization problems with equality and inequality constraints.—*Math. Program.*, 1976, vol. 11, p. 67—80.

82. Monarchi D., Wieber I. E. An interactive multiple objective decision-making aid using nonlinear goal programming. *Papers 22nd Internat. Meeting Inst. Management Sci.*, 1976, p. 235—253.

83. Price W. L. A controlled random search procedure for global optimization.—*Comput. Journ.*, 1977, vol. 20, № 4, p. 367—370.

84. Rosen J. B. The gradient projection method for nonlinear programming. Linear constraints.—*SIAM Journal*, 1961, № 9, p. 180—217.

85. Russel S. S. Nonlinear programming—gradient projection via orthogonalization.—*Internat Numer Methods Engrg.*, 1976, vol. 10, № 4, p. 950—953.

86. Schips B. Ein Betrag zum Thema: Methodenbanken.—*Angew. Inform.*, 1977, № 11, S. 465—470.

87. Shanno D. F. Conditioning of quasi-Newton methods for function minimization.—*Math. of Comp.*, 1970, vol. 24, № 111, p. 647—650.

88. Varda S. Problems in linear and nonlinear programming.—London: 1975.—234 p.

89. Varda J. Praktische Optimierung: Verfahren und Anwendungen der linearen und nichtlinearen Optimierung.—Munich—Vienna: 1974.—365 S.

90. Zoutendijk G. Mathematical programming methods.—N. Y.: 1976.—500 p.

91. Zoutendijk G. Some recent developments in nonlinear programming Fifth Conference of Optimization Techniques, pt. 1.—Berlin: 1973, p. 407—417.

92. Zwart P. B. Nonlinear programming: a quadratic analysis of ridge paralysis.—*Journ. of Optimiz. Theory a. Appl.*, 1970, vol. 6, № 4, p. 331—339.

ОГЛАВЛЕНИЕ

Предисловие	3
Глава первая. Проблема оптимального проектирования	5
1-1. Общая постановка задачи оптимального проектирования	5
1-2. Классификация методов решения задач оптимального проектирования	14
1-3. Понятие о системе оптимального проектирования. Особенности организации программного комплекса	23
1-4. Анализ эффективности алгоритмов поиска оптимума	28
1-5. Тестовые задачи	34
1-6. Пример. Система автоматизации оптимального проектирования асинхронных двигателей	37
Глава вторая. Поиск глобального оптимума	39
2-1. Проблема поиска глобального оптимума	39
2-2. Метод последовательности независимых испытаний	42
2-3. Использование априорной информации о функции критерия оптимальности	47
2-4. Адаптация при поиске глобального оптимума	53
Глава третья. Однопараметрические методы оптимизации	61
3-1. Поиск по признаку возрастания (убывания) критерия оптимальности	63
3-2. Метод квадратичной аппроксимации	66
3-3. Пример. Проектирование системы охлаждения криогенного электродвигателя	68
Глава четвертая. Детерминированные методы определения направления движения к оптимуму	70
4-1. Прямые методы оценки направлений	71
4-2. Выбор направления поиска с помощью градиента	77
4-3. Методы переменной метрики	84
Глава пятая. Стохастические методы определения направления движения к оптимуму	90
5-1. Стохастический подход в оптимальном проектировании	90
5-2. Прямые стохастические методы поиска	91
5-3. «Овражные» методы поиска	96
5-4. Методы стохастической оценки градиента	100
5-5. Применение методов планирования эксперимента	102
Глава шестая. Методы возможных направлений	106
6-1. Формирование направлений поиска в допустимой области	107
6-2. Проекционный градиентный метод и его модификации	112
6-3. Метод переменной метрики в ограниченной области	117
6-4. Вхождение в допустимую область. Выбор длины шага	119
6-5. Метод линейного локального моделирования	122
6-6. Пример. Оптимальное проектирование диска турбины	127
Глава седьмая. Методы штрафных функций	134
7-1. Формирование штрафных функций	134
7-2. Особенности поиска с использованием штрафных функций	143
7-3. Модификация штрафных функций для дискретной области поиска	148
Список литературы	154
	159

**ВЛАДИМИР ИЛЬИЧ ГЕМИНТЕРН,
БОРИС МОИСЕВИЧ КАГАН**

Методы оптимального проектирования

Редактор Д. А. Аветисян
Редактор издательства А. Н. Гусьякая
Обложка художника И. Е. Сайко
Технический редактор Н. Н. Хотулева
Корректор Г. А. Полонская
ИБ № 1694

Сдано в набор 24.01.80. Подписано в печать 27.05.80 Т-11203
Формат 84×108^{1/2} Бумага типографская № 3 Гарн. шрифта литературная
Печать высокая Усл. пел. л. 8,4 Уч.-изд. л. 8,93 Тираж 9500 экз.
Знак 513 Цена 45 к.

Издательство «Энергия», 113114, Москва, М-114, Шлюзовая наб., 10
Московская типография № 10 Союзполиграфпрома при Государствен-
ном комитете СССР по делам издательств, полиграфии и книжной
торговли. 113114, Москва, М-114, Шлюзовая наб., 10.