

内 容 简 介

本书介绍了优化技术的基本理论以及作者近年来应用数学规划理论解决机械设计问题的研究成果。以直接法为主，求导法为辅，附有大量设计实例及FORTRAN通用程序。

全书共分八章，内容为设计方法的改进及优化技术概述；直接法及其应用；优化设计中一些实际问题的处理方法；求导法及其应用；约束条件转化方法及应用；处理多目标函数优化问题的工程方法及应用；求解线性问题的方法；工程设计实例。

此书可作为机械工程，精密机械，光学工程专业高年级本科生教材，也可供有关专业研究生及科技人员参考使用。

优化技术及其应用

何献忠 刘玉桐 金小海 编著

北京工业学院出版社出版

新华书店北京发行所发行 各地新华书店经售

三河县中赵甫印刷厂印刷

787×1092毫米 16开本 10.75印张 260千字

1986年6月第一版 1986年6月第一次印刷

印数：1—10,000册

统一书号：15434·28 定价：1.80元

前 言

由于电子数字计算机的普及,七十年代初期,优化技术(或称最优化技术)在国外已渐普遍应用于各技术领域之中,并取得了显著的成果。

目前,国内正处于引进、学习、使用阶段,并在建筑、航空、光学等方面取得了积极的进展。

1977年底在光学系统自动设计学习、推广的启发下,我们拟将优化技术引入精密机械设计之中。1978年3月,在华北地区弹簧国家标准宣传贯彻会上,我们发表了第一篇〈圆柱形螺旋弹簧优化设计及程序〉学术报告,得到了有关部门的支持,加强了我們继续探索的决心。之后,在兄弟院校及院内兄弟系的支援、帮助下,我们继续探讨,并在不同的学术会议上继续发表了探索性成果论文。它们包括:螺旋弹簧、片弹簧、非标准滚珠轴承、滚珠导轨、齿轮传动系统、片簧式径向调速器等精密机械零、部件的优化设计及约束降维法、图解分析法两种优化方法。故本书亦曾拟定名为〈精密机械优化设计〉。

1980年9月我们写成了〈优化技术及应用〉(初稿)。1981年2月,以此作教学参考书,为光学仪器专业学员开设了选修课。1981年8月,又以此书的修改稿作参考书为机床设计专业开设了选修课。初稿印出后,有关院校及厂所,对它提出了不少宝贵意见及建议。在此基础上,我们又于1981年9月完成了〈优化技术及应用〉铅印版。本书就是根据该版本再次修改后完成的。金小海同志参加了这次的编写工作。

由于基层计算机房的建立及微机普及,书中通用程序作了较大的补充、改进。使用机型以PDP 11/23及IBM PC/XT为主,使用、移植皆极简易。

本书曾经五年教学使用(包括:本科生、研究生、进修教师、外出讲学);有关软件已在机械设计、精密机械设计、光学仪器设计、摄影仪器设计、光学零件膜系设计、激光器件设计、电子光学系统设计等科研实践中,得到了应用,并取得了成果。对求解三十维以下的工程实际问题,比较可靠有效。

我们水平较低,实践不够,书中缺点、错误必然存在,诚恳地期待各方面的批评、指正。

本书及有关学术论文在写作过程中,曾得到清华大学吴宗泽副教授、华东工程学院郑祖炳副教授、哈尔滨工业大学初允纬副教授、张宗兰副教授、华中工学院廖道训副教授以及我院林汉藩教授、吴兆汉教授、梁家玉副教授、韩锡勋副教授的支持和帮助。黄航汉工程师为本书配制了全部附图。梁锁记同志也曾参加过部分研究工作。在此,一并表示深切感谢。

1985年6月

目 录

第一章 设计方法的改进及优化技术概述

- §1-1 为什么要研究设计方法..... 1
- §1-2 设计方法改进概况..... 1
 - 1. 试算法 2. 表格法 3. 图算法 4. 利用一元函数极值理论的设计法 5. 优化设计法
- §1-3 有关优化设计的术语及概念..... 5
 - 1. 实际技术问题及其数学模型抽象 2. 设计变量 3. 目标函数 4. 约束条件 5. 设计空间及设计可行域 6. 优化设计的一般数学模型
- §1-4 优化设计过程形象化表达..... 9
 - 1. 等值线与优化点 2. 择优过程的形象表达

第二章 直接法及其应用

- §2-1 直接法的基本特点及方法.....11
- §2-2 坐标轮换法及其应用.....11
 - 1. 基本设想 2. 逻辑框图 3. 计算示例 4. 最小外径非标准滚珠轴承论证 5. 坐标轮换法的局限性
- §2-3 单纯形法及其应用.....18
 - 1. 基本设想 2. 构成初始单纯形的方法 3. 收敛条件 4. 择优过程与框图 5. 实例运算 6. 最小推力滚珠导轨设计 7. 结论 8. 使用上的局限性
- §2-4 复形法及其应用.....29
 - 1. 基本设想 2. 初始复形的构成 3. 收敛判别条件 4. 择优过程及框图 5. 实例运算 6. 钢丝滚道非标准滚珠轴承优化设计及程序 7. 片簧式径向作用调速器优化设计 8. 结论
- §2-5 约束降维法及其应用.....52
 - 1. 坐标轮换法在求解无约束极小化问题时的局限性 2. 约束降维法的基本设想 3. 有常量约束的最小推力滚珠导轨优化设计及结果对比 4. 最小重量压力弹簧优化设计及程序 5. 片簧式径向作用调速器优化设计对比 6. 约束降维法的优越性
- §2-6 图解分解法.....60
 - 1. 基本设想 2. 半圆弧状片簧优化设计及程序

第三章 优化设计中一些实际问题的处理方法

- §3-1 实用步长确定法.....68
 - 1. 一维搜索、最小函数值步长确定法 2. 简单实用步长确定法
- §3-2 改善局部优化的工程方法.....74
 - 1. 函数凸性判别 2. 局部优化的实用价值 3. 局部优化的相对改善
- §3-3 离散量的处理方法.....77
 - 1. 曲线拟合技术 2. 实用优化参数的处理

§3-4 线性方程组求解	86
1.消元法 2.行列式求解法 3.逆矩阵法	
第四章 求导法及其应用	
§4-1 求导法的基本出发点	90
§4-2 多元函数极值理论	90
1.多元函数的近似表达式和最速下降方向 2.几种特殊类型函数的斜量公式 3.极值点的充要条件	
§4-3 最速下降法及其应用	95
1.基本设想 2.分析及改进的可能 3.最短带长带传动设计	
§4-4 牛顿法及其改进	97
1.基本设想 2.实例运算分析 3.改进的可能性——“阻尼牛顿法”	
§4-5 变尺度法及其应用	100
1.基本设想 2.新矩阵 (H_k) 的构成 3.运算示例 4.框图 5.运算实例及程序	
§4-6 阻尼最小二乘法及其应用	112
1.最小二乘法 2.阻尼最小二乘法 3.阻尼因子的选择 4.计算实例 5.迭代过程及框图	
第五章 约束条件转化方法及应用	
§5-1 等式约束极小化问题的转化	117
1.待定乘数法的基本设想 2.一般应用实例 3.设计应用实例——最低成本偏差设计	
§5-2 一般有约束极小化问题的转化	126
1.基本设想 2.具体方法 3.罚参数的选取及优化点的判别 4.计算实例	
§5-3 不等式约束转化为等式约束	130
第六章 处理多目标函数优化问题的工程方法及应用	
§6-1 什么是多目标函数问题	132
1.定义及基本数学模型 2.多目标函数的等值线图形	
§6-2 一般求解的工程方法	133
1.主目标函数法 2.加权求和法	
§6-3 多技术要求齿轮传动系统分速比优化设计	134
1.建立初始模型 2.用主目标函数法建模并求优化解 3.用加权求和法建模并求优化解 4.两种方法所得实用结果对比 5.专用子程序	
§6-4 考虑到目标函数相互矛盾的求解法	137
1.协调曲线的获得及分析 2.有约束多目标函数极小化的协调曲线及分析 3.利用协调曲线求优化解	
第七章 求解线性问题的方法	
§7-1 另一类优化问题	139
1.基本模型 2.有关定义	
§7-2 线性规划的单纯形法	140
1.基本设想及优化解判别根据 2.实例图解分析 3.具体方法	
§7-3 运算实例	144
第八章 工程实例	
§8-1 光学工程实例	147

1. 自动安平型光学仪器悬挂系统设计	2. 近半球腔激光器最佳参数设计	
§8-2 机械工程实例·····		151
1. 最小应力牙嵌离合器设计	2. 最小重量工程车辆传动轴设计	3. 最小体积二级圆柱齿轮
减速器设计	4. 最低成本切削用量的优选	

第一章 设计方法的改进及优化技术概述

§1-1 为什么要研究设计方法

设计是将某种设想转变为现实的一段过程，是为加工提供根据的必要手段。评价一种设计方法优劣的主要根据，是设计质量及设计速度。设计质量固然根本取决于所用的基本理论是否正确，同时也取决于设计方法恰当与否；设计速度则主要取决于设计方法及运算辅助工具。如果所用基本理论一致，则不同的设计方法对设计的影响将更明显。可见，为提高设计质量及设计速度，改进设计方法是极其重要的。多年来，人们为此所作的努力产生了十分明显的效果。

§1-2 设计方法改进概况

1. 试算法 这种方法始于本世纪二十年代末。试算法以一定的理论公式为根据，利用已知或假定的技术条件，通过多次试算、修改，最终获得适用的设计参数。

例如欲设计一个刚度 P' 一定的圆柱形螺旋压簧，可以根据下列刚度公式进行试算

$$P' = \frac{P}{F} = \frac{G d^4}{8n D_2^3}$$

式中 P —— 弹簧所受的轴向负荷；

F —— 弹簧在负荷 P 作用下所产生的变形量；

D_2 —— 弹簧的平均直径，简称中径；

n —— 弹簧的有效圈数；

d —— 弹簧材料的直径；

G —— 弹簧材料的切变模量。

根据上式，如已知或先预定 D_2 、 n 、 d 、 G 诸参数，通过多次试算、修改，就有可能得到压簧刚度等于或接近于 P' 的设计参数。

刚度公式也可以写成一般的多元函数表达式，即

$$y = f(x_i) \quad i = 1, 2, \dots, n$$

式中 y 代表性能指标 P' ， x_i 是设计参量，分别代表 G 、 d 、 n 、 D_2 ，所以 $n=4$ 。

对于一个多元函数，如要求函数值一定，固然可以通过适当选定诸 x_i 值来满足要求。但在 x_i 既有一定数值范围限制又包括部分离散量的情况下，即使经过多次试算、修改，也难获得理想结果。计算量也会随着试算次数的增多而加大。

2. 表格法 这种方法始于本世纪三十年代。它仍以一定的理论计算公式为根据，参照常用离散数列及规范，预制出系统的表格，供设计者直接查阅。目的在于简化设计过程、减少重复的试算量。如螺旋状拉、压弹簧设计中所用屈度系数表格(见表1-1)。

表 1-1

C	4	5	6	7	8	9	10
K	1.40	1.31	1.25	1.21	1.18	1.16	1.14

制订上表的根据是屈度系数计算式

$$K = \frac{4C-1}{4C-4} + \frac{0.615}{C}$$

其中 $C = D_2/d$ 为弹簧旋绕比。在选定 C 后, 依上表即可查得 K 值。

又如滚动轴承手册中的“轴承结构参数”表格, 只要根据额定负荷及其它条件, 便可查得设计所需标准轴承的类型、型号及尺寸。

表格法比试算法可相对减少计算量, 但由于表格必须分档, 所列数值又皆为离散量; 兼之分档又不能过细, 故表列数值, 难于适应各种需要, 如表列数值不理想, 尚须插值求解。

3. 图算法 这种方法始于本世纪四十年代。它也以一定的理论公式为根据, 建立图尺方程, 确定图尺系数, 作出具有专用图线的算图。这些专用图线, 避免了函数值的离散化, 使用时也无需用插值法求中间值。

例1 屈度系数 K 线图

如图 1-1 所示, 线图由 d 、 D_2 、 K 三条直线组成。它们既可保持弹簧材料直径 d 因标准化所限定的离散性, 而又可使 D_2 适应设计要求, 并可快速而又比较精确地获得屈度系数 K 。

具体使用方法是, 如选 $d = 11\text{mm}$, $D_2 = 41\text{mm}$, 先分别在 d 线及 D_2 线上找到相应的两点,

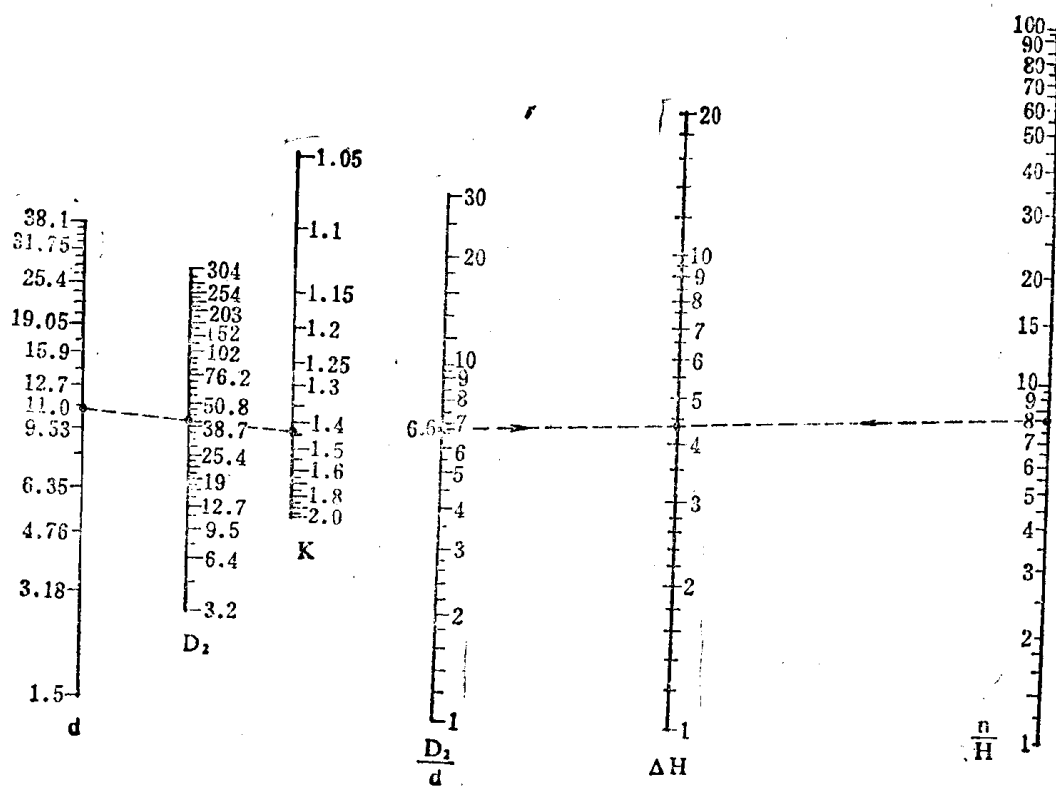


图 1-1

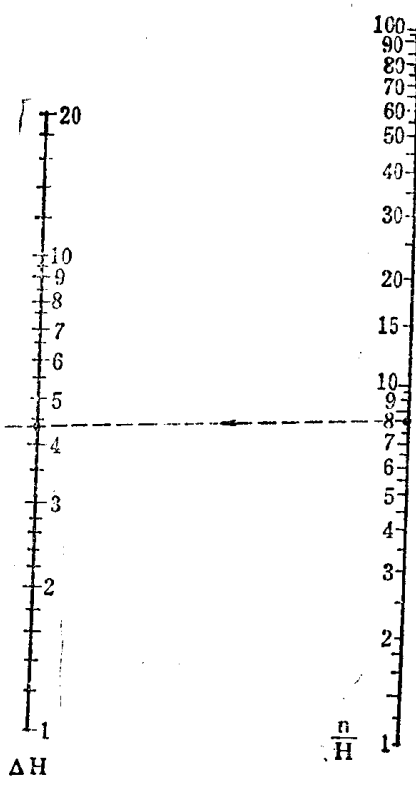


图 1-2

然后联结 d —— D_2 并延长，与 K 线相交，交点即为 K 值，近于1.44。

例2：弹簧允差线图

如已知弹簧中径 D_2 、材料直径 d 、有效圈数 n 、自由高度 H ，可利用弹簧允差图线，直接查出中径偏差 ΔD_2 及自由高度偏差 ΔH 。

如图1-2所示，自由高度偏差图线由 D_2/d 、 ΔH 、 n/H 三条直线组成。如取 $D_2/d=6.6$ 、 $n/H=8$ ，分别在 D_2/d 、 n/H 两线上找到相应的点，并在图上用虚线相联，联线与 ΔH 线相交，交点所对应的值即 $\Delta H=4.4\%$ （它为自由高的百分数）。

如图1-3所示，中径偏差图线由 d 、 D_2 、 ΔD_2 三条直线组成，如取 $d=1.2\text{mm}$ 、 $D_2=8\text{mm}$ ，分别在 d 、 D_2 两线上找到相应的点，并在图上用虚线相联，向右延长与 ΔD_2 线相交，交点所对应的值即 $\Delta D_2=0.12\text{mm}$ 。

还可利用其它图线，查得相应负荷允许偏差值。

可见，利用预制图线既可相对提高设计质量及速度，又可保持各设计参数的数值特性，所得结果只需进行实用数值处理即可。

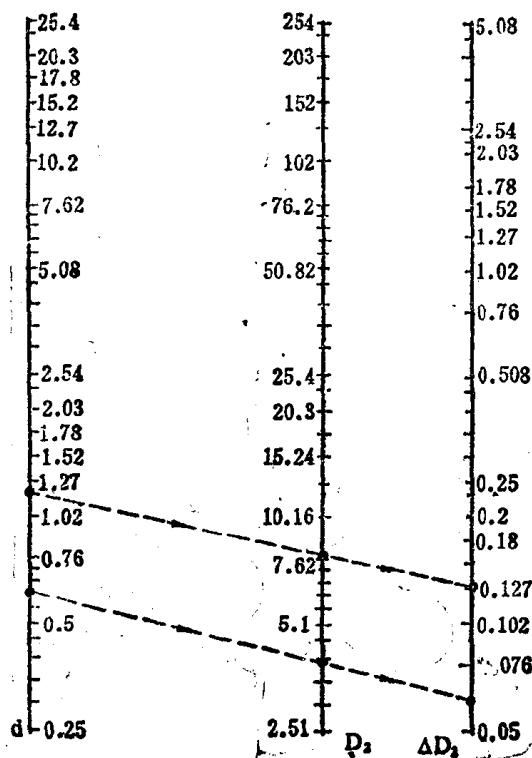


图 1-3

4. 利用一元函数极值理论的设计法 这种方法始于四十年代末，目的在于获得理论上的最优设计性能，是优化设计的萌芽。

如某一设计的性能指标为 y ，诸设计参量为 x_i ，并保持一定函数关系

$$y=f(x_i) \quad i=1, 2, \dots, n$$

y 的极大值或极小值，表征了设计的最优性能。由获得极值的必要条件

$$y'=0$$

所确定的设计参量，即为获得最优性能所应选用的具体值。

实际上，绝大多数设计都非一元问题，为简化运算，都是只保留诸设计参量中的一个变量，而将其余赋值，使多元函数降为一元函数，然后对一元函数求极值得到最优参数。

例1 最多释放圈数蜗线弹簧设计

蜗线弹簧释放圈数 n 与其有关的结构参数关系为

$$n = \frac{R_2 - R_3 - R + \sqrt{R^2 + R_3^2 - R_2^2}}{h}$$

式中 R_2 ——蜗线弹簧全部拧紧时的外半径；

R_3 ——蜗线弹簧所用芯轴半径；

R ——蜗线弹簧盒内半径；

h ——蜗线弹簧厚度。

当选定 R 、 R_3 、 h 的具体尺寸后，则 n 仅与 R_2 保持一元函数关系，如令

$$\frac{dn}{dR_2} = \frac{1}{h} \left(1 - \frac{R_2}{\sqrt{R^2 + R_3^2 - R_2^2}} \right) = 0$$

则可得

$$R_2 = \sqrt{\frac{R^2 + R_3^2}{2}} = f(R, R_3)$$

例2 最小重量或最小体积螺旋压簧设计

最小重量条件式：依

$$\frac{dW}{dC} = 0$$

$$\text{得 } B = \frac{C^3 (C^3 - 0.635C^2 - 0.98C + 0.615)^{\frac{1}{2}} (5C^3 - 7.27C^2 - 1.21C + 1.23)}{4(1.365C^4 - 0.732C^3 - 0.673C^2 + 0.981C - 0.378)}$$

最小体积条件式：依

$$\frac{dV}{dC} = 0$$

$$\text{得 } B = \frac{C^4}{2} \left(\frac{C-1}{C^2 + 0.365C - 0.615} \right)^{\frac{1}{2}} \left(\frac{7C^3 - 5.54C^2 - 8.9C + 3.21}{2C^3 + 0.095C^2 - 3.19C + 1.845} \right)$$

在以上诸式中， W 为压簧重量； V 为压簧体积； C 为压簧旋绕比； B 为系数，它与压簧其它参数，保持以下关系

$$B = \frac{FG}{n\sqrt{8P\pi}[\tau]} = f(F, G, n, P, [\tau])$$

式中， $[\tau]$ 为压簧材料许用切应力，其余符号意义同前。

理论上，只要选定 F 、 G 、 n 、 P 、 $[\tau]$ ，并使旋绕比 C 分别满足以上条件式，即可获得最小重量或最小体积压簧。

例3 齿轮系统最短起动时间传动比设计

最短起动时间，对应于最大起动角加速度。如角加速度为 ε ，依

$$\frac{d\varepsilon}{di} = 0$$

得

$$i = m + \sqrt{m^2 + K}$$

$$m = \frac{M_2}{(M - M_t)\eta}$$

$$K = \frac{J}{J_m}$$

式中 M_2 ——负载力矩;
 M ——电机有效力矩;
 M_f ——系统全部摩擦力矩;
 η ——总效率;
 J ——全部系统转化到负载轴上的转动惯量;
 J_m ——电机的转动惯量。

实际上, 传动比 i 将具有以下一般关系

$$i = f(M_2, M, M_f, \eta, J, J_m)$$

可见, 理论上只要选定上式右端各参数, 并满足该条件式所得的传动比 i , 就可使该齿轮系统获得最短起动时间。

由以上三例可见, 为建立一元函数模型, 需要将一些设计参数取为常数, 而仅保留一个变量。但对于一般情况, 依极值必要条件所得的条件式, 仍然是多元函数, 故具体设计时, 要想合理地满足条件式, 将亦困难。总之, 这种设计方法虽有理论意义, 但少实用价值。

5. 优化设计法 优化设计亦称最优化设计。优化设计始于五十年代末, 而普及应用于七十年代, 它是以数学规划理论为基础, 以电子数字计算机为辅助工具的一种设计方法。

优化的基本点是, 将优化技术应用于设计过程之中, 最终获得较理想的设计参数。由于这种设计一般多在完成初始设计之后进行, 最终获得优化参数及结果, 故称之为优化设计。优化设计大体分为两类:

一为: 直接计算函数值、比较函数值, 并以之作为迭代、收敛根据的方法。

二为: 以多变量函数极值理论为根据、利用函数性态、以之作为迭代、收敛根据的方法。前者泛称为直接法, 后者泛称为求导法, 两种方法的择优、运算过程, 皆按预编程序在电子数字计算机上进行, 故在有的技术领域中, 亦将此过程称之为自动设计。

近十年来, 优化技术在理论和应用上都得到了很大发展。实践证明, 优化设计可明显地提高设计质量及速度, 为进一步改进设计方法展示了广阔的前景。

§1-3 有关优化设计的术语及概念

1. 实际技术问题及其数学模型抽象 为具体说明与优化设计有关的基本概念及模型, 现举例于下。

例1 设计如图 1-4 示梯形槽使其截面、容积最大。已知板料尺寸为, 宽度 $b=24\text{mm}$ 、长度 $L=50\text{mm}$ 。

解 板长即槽长。槽长一定时, 槽截面积最大, 容积亦最大。

依梯形截面公式, 得槽截面积 A 为

$$A = \frac{1}{2}[(24-2x) + (24-2x+2x \cos \alpha)]x \sin \alpha = f(x, \alpha)$$

如上式有极大值存在, 则按二元函数求极值的方法, 就可求得最大截面积 A 所对应的 x 及

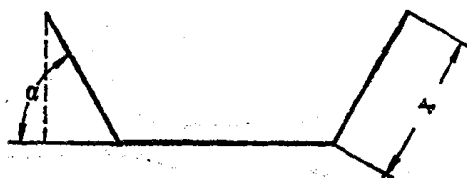


图 1-4

α 值。如实际设计按此取值,就可获得最大容积的槽。

以上问题也可简写为

$$\max f(x, \alpha) \quad (1-1)$$

它表示了求解截面积A极大化的模型, max表示极大化。由于式(1-1)为非线性函数,故称非线性问题。

例2 最小重量螺旋压簧设计

设计要求:最大工作负荷 $P=30\text{N}$,最大工作变形量 $F=10\text{mm}$,压并高小于 50mm ,压簧内径大于 16mm ,有效圈数在 $3\sim 10$ 圈之内。

解 因压簧重量 W 为

$$W = \pi D_2 n_1 \frac{\pi d^2}{4} \gamma = f(D_2, n_1, d, \gamma)$$

式中 n_1 ——压簧总圈数; γ ——弹簧材料重度;其余符号意义同前。

设计时应满足的条件(技术要求)为,在最大工作负荷作用下不破坏,即

$$\frac{8PD_2}{\pi d^3} K \leq [\tau]$$

小于要求的压并高,即

$$H_b = n_1 d \leq 50$$

大于要求的内径,即

$$D_1 = D_2 - d \geq 16$$

最大变形量等于 10mm ,即

$$F = \frac{8PD_2^3 n}{Gd^4} = 10$$

有效圈数范围,即

$$10 \geq n \geq 3$$

根据以上条件,因负荷 P 为常量,若再令重度 γ 亦为常量,取 $n_1 = n$,则整个问题的模型可以归纳并写为

$$\min f(D_2, n, d) \quad (1-2)$$

$$\text{sub } g_1(D_2, d) - [\tau] \leq 0 \quad (1-3)$$

$$g_2(n_1, d) - 50 \leq 0 \quad (1-4)$$

$$g_3(D_2, d) - 16 \geq 0 \quad (1-5)$$

$$h_1(D_2, n, d) = 10 \quad (1-6)$$

$$10 \geq n \geq 3 \quad (1-7)$$

式(1-2)至式(1-7)概括表明,在满足各项技术要求时,求解压簧极小化重量的数学模型。min表示极小化、sub表示应满足的要求, g_i 表示不等式的函数形式, h_1 表示等式的函数形式。

由于式(1-2)至式(1-6)多为非线性函数,故这个问题是非线性的。

例3 选定最低总价格轴承

欲选定四种滚珠轴承,其个数分别为 x_1, x_2, x_3, x_4 ,轴承单价分别为 y_1, y_2, y_3, y_4 。要求 x_1, x_2 之和不少于24个, x_3, x_4 之和不少于32个, x_1, x_2, x_3 之和不少于36个,且总价格最低。

解 根据以上条件及要求, 即可写出其模型。

$$\min \quad Y = x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4 \quad (1-8)$$

$$\text{sub} \quad x_1 + x_2 \geq 24 \quad (1-9)$$

$$x_3 + x_4 \geq 32 \quad (1-10)$$

$$x_1 + x_2 + x_3 \geq 36 \quad (1-11)$$

式(1-8)至式(1-11)概括表明, 在满足各项要求时, 求解四种滚珠轴承总价格最低的数学模型。由于它们都是线性函数, 故这个问题是线性的。当然, 以上模型也可写成一般形式, 即

$$\left. \begin{array}{l} \min \quad f(x_1, x_2, x_3, x_4) \\ \text{sub} \quad g_1(x_1, x_2) - 24 \geq 0 \\ \quad \quad g_2(x_3, x_4) - 32 \geq 0 \\ \quad \quad g_3(x_1, x_2, x_3) - 36 \geq 0 \end{array} \right\}$$

由以上三例可以看出, 有关设计问题的数学模型建立和抽象的根据是设计任务、技术要求及其有关技术知识。模型概括地表达了有关设计问题的全部要求, 它是选择优化方法及择优计算的主要根据。

2. 设计变量 设计变量是设计模型的基本成分, 是设计最后所需确定的参数, 如例1中的 x, a ; 例2中的 D_1, n, d ; 例3中的 x_1, x_2, x_3, x_4 。设计变量的个数, 即是所求解问题的维数, 故例1为二维问题, 例2为三维问题, 例3为四维问题。有的工程技术问题可有几十维, 几百维, 甚至有更高维数的。

3. 目标函数 以所选定的设计变量为自变量, 以所要求的性能指标为因变量, 并按一定关系(如几何关系、物理关系、传动关系、……)所建立起来的函数式, 即为目标函数。它反映了设计性能要求与设计参数之间的关系, 如式(1-1)、式(1-2)、式(1-8)。由于目标函数的函数值大小, 可以评价设计质量的优劣, 故在有的著作中, 亦称之为评价函数。

设计变量的个数, 确定了目标函数的维数。设计变量的幂及函数的性态, 确定了目标函数的性质, 如式(1-1)为二维非线性函数, 式(1-2)为三维非线性函数, 式(1-8)为四维线性函数。

如果所选的设计变量与所要求的性能指标之间无精确的函数关系, 亦可采用曲线拟合、多元回归或其它近似计算方法, 获得近似的函数式作为目标函数。

如果在同一设计中, 需要满足一个以上的性能指标, 则可分别建立一个以上的目标函数表达式, 并以之作为初始模型。

4. 约束条件 约束条件包括常量约束与约束方程两类。常量约束亦称边界约束, 它表明设计变量的允许取值范围, 如式(1-7)。约束方程亦称性能约束, 它是所选定的设计变量为自变量, 以要求加以限制的设计参数为因变量, 按一定关系(如性能关系、几何关系、设计规范、……)建立起来的函数式, 它常用来限制某些设计性能。式(1-3)、式(1-4)、式(1-5)、式(1-6)等都属于约束方程。约束方程又分不等式约束和等式约束。

可以利用一定方法, 将约束形式相互转变, 如 $g(x_i) \leq 0$, 可转变为 $-g(x_i) \geq 0$, 亦可转变为 $g(x_i) + x_i = 0$ 。在一般表达式中, 只写一种而代其余。

在所求解的问题中, 有时并无约束, 有时则有约束。如例1为非线性无约束问题, 例2为非线性有约束问题, 例3为线性有约束问题。

5. 设计空间及设计可行域 为便于分析、研究,应用矩阵及向量的知识,可将设计模型转化为设计空间,并在此空间内,讨论择优过程。

如将设计方案抽象为一个空间向量 X ,并将与之有关的诸设计变量抽象为诸分向量 $(x_1, x_2, \dots, x_n)$,且各分向量线性独立,则以诸分向量为轴所构成的空间,即为设计空间。如有 n 个独立的设计变量,就可相应地构成 n 维空间,称为 n 维欧氏空间,记为 E^n 或 R^n 。例如

$n=2$, 则有

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = (x_1, x_2)^T$$

$n=3$, 则有

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = (x_1, x_2, x_3)^T$$

.....

$n=n$, 则有

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = (x_1, x_2, \dots, x_n)^T$$

显然,在设计空间中的每一个点,都唯一地确定了一个空间向量,它代表了一组分向量及其数值,实际上也就代表了一种具体的设计方案。

如在设计空间中存在着使目标函数达到极值的点,则该点就代表着一种优化设计方案。如在该空间中仅有一极值点,则它就是全空间中的最优设计方案。该点常用 X^* 或 X_{opt} 表示。

在设计空间中,被约束条件所限定的区域,即为设计可行域(如图 1-5 示),它是设计空间中的一个局部。优化设计的寻优过程,一般只应在此区域内进行。最后确定的优化点,也只能在此可行域内,或在可行域的边界上。否则,所得的设计参数将因超出约束而失去实用价值。

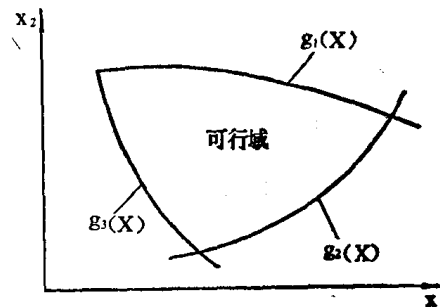


图 1-5

6. 优化设计的一般数学模型 优化设计一般常按两大类情况建模:无约束极小化模型和有约束极小化模型。

无约束极小化模型的一般型式为

$$\min f_k(X), X = (x_1, x_2, \dots, x_n)^T, X \in E^n \quad (1-12)$$

式(1-12)为多目标函数、无约束极小化模型。如 $k=1$, 则为

$$\min f(X), X = (x_1, x_2, \dots, x_n)^T, X \in E^n \quad (1-13)$$

式(1-13)为单目标函数、无约束极小化模型,它是最基本的模型。根据不同的实际要求, k 可为1或大于1的有限数。

有约束极小化模型的一般型式为

$$\begin{array}{ll}
 \min & f_k(X), X=(x_1, x_2, \dots, x_n)^T, X \in E_n \\
 & k=1, 2, \dots, K \\
 \text{sub} & \left. \begin{array}{ll} g_j(X) \geq 0, & j=1, 2, \dots, J \\ h_l(X)=0, & l=1, 2, \dots, L \\ a_i \leq x_i \leq b_i, & i=1, 2, \dots, n \end{array} \right\} \quad (1-14)
 \end{array}$$

它一般地概括了设计中要求最复杂的情况。如 $k=1$ ，则为单目标函数、有约束极小化模型。如 $k \geq 2$ ，则为多目标函数、有约束极小化模型。如仅有 $g_j(X) \geq 0$ （或包括 $g_j(X) \leq 0$ ），则称具有不等式约束。如仅有 $h_l(X)=0$ ，则称具有等式约束。如仅有 $a_i \leq x_i \leq b_i$ ，则称具有常量约束。

由于精密机械设计所涉及的目标函数及约束条件多为非线性函数，且 $\max f(X) = \min [-f(X)]$ ，故将研究、分析的重点放在非线性极小化方面。

又由于可以利用一定的数学方法，将多目标函数转变为单目标函数，有约束的模型转化为无约束的模型，故对单目标函数、无约束极小化问题的探讨，应成为探讨优化技术的基点及起点。

§1-4 优化设计过程的形象化表达

所谓优化设计，就是根据设计模型及初始设计参数，利用一定的优化方法编出程序，通过电子数字计算机，求出优化参数及优化性能指标。为便于形象化表达择优过程，现以二元函数为例进行分析。

1. 等值线与优化点 等值线亦称等高线，它是表达二元函数函数值大小及其变化规律的一种直观图形。

二元函数的几何图形通常是一个空间曲面。设

$$f(x_1, x_2) = C_1$$

且 C_i 为常数。如依次令 $i=1, 2, \dots, n$ 则可得一系列平面曲线

$$f(x_1, x_2) = C_1$$

$$f(x_1, x_2) = C_2$$

.....

$$f(x_1, x_2) = C_n$$

若将这些曲线投影于 x_1, x_2 轴所构成的平面内，则得一组等值线，如图1-6所示。它们亦相当于过 C 轴上 C_i 诸点所作、平行于 x_1x_2 平面的诸平面与空间曲面的交线在 x_1x_2 平面上的投影。它们形象地表达了函数值的大小及其变化规律， $C_1 > C_2 > C_3 > \dots > C^*$ ， X^* 点即函数值的极小点， C^* 表示该点的函数值。（ C 轴垂直于纸面）

可以证明，对于二元函数，如有极值点存在，则在该点附近的等值线为一族共心椭圆。设 $f(x_1^*, x_2^*)$ 为函数 $f(x_1, x_2)$ 的一个极值点，则

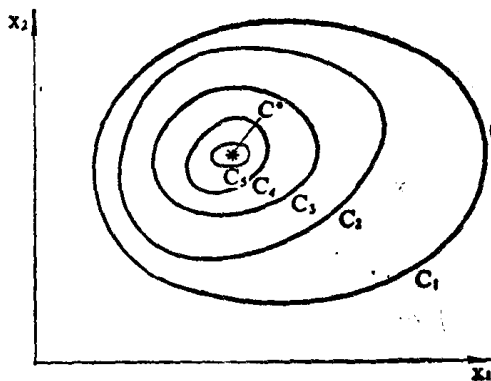


图 1-6

$$f'_1(x_1^*, x_2^*) = f'_2(x_1^*, x_2^*) = 0$$

将函数 $f(x_1, x_2)$ 在极值点 (x_1^*, x_2^*) 附近展成泰勒级数, 且仅取到二阶导数项, 则得近似式为

$$\begin{aligned} f(x_1, x_2) \approx & f(x_1^*, x_2^*) + \frac{1}{2} \left[f''_{x_1 x_1}(x_1^*, x_2^*) (x_1 - x_1^*)^2 \right. \\ & + 2f''_{x_1 x_2}(x_1^*, x_2^*) (x_1 - x_1^*) (x_2 - x_2^*) \\ & \left. + f''_{x_2 x_2}(x_1^*, x_2^*) (x_2 - x_2^*)^2 \right] \end{aligned}$$

如令
$$\begin{aligned} a_{11} &= f''_{x_1 x_1}(x_1^*, x_2^*), & a_{12} &= f''_{x_1 x_2}(x_1^*, x_2^*), \\ a_{22} &= f''_{x_2 x_2}(x_1^*, x_2^*) \end{aligned}$$

则上式可写为

$$\begin{aligned} f(x_1, x_2) \approx & f(x_1^*, x_2^*) + \frac{1}{2} [a_{11}(x_1 - x_1^*)^2 + 2a_{12}(x_1 - x_1^*)(x_2 - x_2^*) \\ & + a_{22}(x_2 - x_2^*)^2] = C \end{aligned}$$

上式亦即等值线方程。如 (x_1^*, x_2^*) 为极小点, 则其充分条件为

$$a_{11}a_{22} - a_{12}^2 < 0$$

如以上条件成立, 则亦表明等值线方程为椭圆方程, 故在极值点附近, 对于某些常数 C , 的等值线, 都是以 (x_1^*, x_2^*) 为中心的椭圆, 该中心即优化点 X^* 。

可见, 在图示 x_1, x_2 座标所构成的设计场内, 椭圆族的中心 X^* 即优化设计方案所在。

2. 择优过程的形象表达 仍以二元函数为例, 已知该函数的极小值点 X^* 及其附近的等值线族(如图1-7示)。图中的 X_0 点代表任选的一组初始设计参数。极小化的择优过程, 就是从 X_0 点开始, 按照一定的方向, 以一定步长, 一步步地接近 X^* 点, 直到满足要求的条件时止。

具体的一种过程为: 如从 X_0 点始, 先沿 x_1 轴方向一步步前进, 由于 $C_1 > C_2 > C_3 > \dots > C^*$, 故每走一步后, 函数值皆有所改善。当达到 X_{13} 点时, 函数值反而增大, 这时应退回到 X_{12} 点。然后沿 x_2 轴方向一步步前进, 直到 X_{24} 点, 函数值反而增大时再退回到 X_{23} 点。然后再沿 x_1 轴方向搜索, 当发现函数值增大时止。后撤一步, 再次换轴。如此反复搜索、反复迭代, 直到接近 X^* 时止。这就是择优、搜索、迭代的一种过程, 由于它直观易懂, 故以此为例。

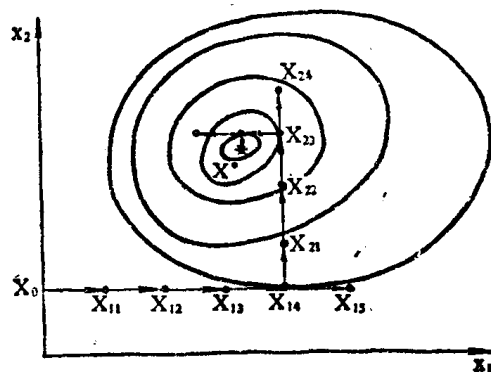


图 1-7

依上例可见, 择优过程也就是按照一定方向, 一步步地接近优化点 X^* 的过程。它的根本问题可归纳为, 如何确定搜索方向? 如何确定每步步长? 如何制定收敛判别条件等三方面。尽管优化技术多种多样, 但其差别亦仅在于此。

不难想象, 迭代过程的计算量, 将依设计变量的多少、目标函数及约束条件的繁简、以及所选优化方法而定。

第二章 直接法及其应用

§2-1 直接法的基本特点及方法

优化技术中的直接法，主要是在迭代过程中直接比较函数值的大小，再根据一定的收敛判别条件，获得优化解。它的基本思想及迭代过程，直观易懂，易于为工程技术人员接受。但是，由于它并未利用目标函数的性态及其解析性质，故收敛较慢。这类方法适用于处理低维问题。

直接法所包括的具体方法很多，如坐标轮换法，单纯形法，复形法，约束降维法，可变容差法，序贯加权因子法，CFPM……等。现仅研究、讨论其中的几种。

§2-2 坐标轮换法及其应用

1. 基本设想 坐标轮换法亦称降维法。现以二维问题为例，说明它的基本思想及迭代过程。

如图 2-1 所示的设计空间，以箭头所示各轴方向为正，相反方向为负。如设计初始点为 $X_0 = (x_{10}, x_{20})^T$ ，从该点出发，先沿 x_1 轴方向搜索，第一步得点 X_{11} ，其函数值为 $f(X_{11})$ 。将其与初始点函数值 $f(X_0)$ 比较，如 $f(X_{11}) < f(X_0)$ ，表明搜索有效，可以继续前进，直到 $f(X_{14}) > f(X_{13})$ 时止。这时应返回至 X_{13} 点。再以该点为起点，沿 x_2 轴正向搜索得 X_{21} ，如 $f(X_{21}) > f(X_{13})$ ，则表明搜索方向不对，应反向搜索得 X_{22} ，如 $f(X_{22}) < f(X_{13})$ ，表明搜索有效，可继续前进得 X_{23} 、 X_{24} 诸点。如 $f(X_{24}) > f(X_{23})$ ，则应返回 X_{23} 点。依次再沿 x_1 轴、 x_2 轴、按上述原则进行搜索，直到接近（或达到）理论优化点 X^* 时止。

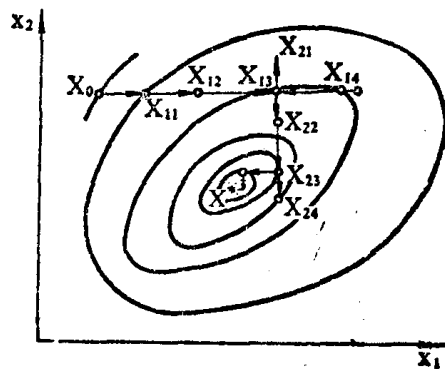


图 2-1

从工程设计的实际情况出发，不难想象，实际设计只能在一定程度上接近理论优化点，而难于达到该优化点。所以，必须提出一定的收敛判别条件，作为停算的根据。如以迭代过程中，相邻两次函数值之差值作为收敛条件，即 $|f(X_{i,k}) - f(X_{i,k-1})| \leq \varepsilon$ (ε 为收敛系数)，则在每次迭代过程中都应以该条件进行验算，一旦满足该不等式，即停止搜索，所得即为满足设计要求的相对优化解。当然也可采用其它收敛条件验算。

再以四维问题为例，如图 2-2 所示的设计空间，仍以箭头所示方向为正，相反为负。如设计初始点 $X_0 = (x_{10}, x_{20}, x_{30}, x_{40})^T$ ，先沿 x_1 轴正向搜索，前进五步，因 $f(X_{15}) > f(X_{14})$ ，故止于 X_{14} 。然后以此作起点，再沿 x_2 轴正向搜索，因 $f(X_{21}) > f(X_{14})$ ，故反向前进，至第四步时，因 $f(X_{24}) > f(X_{23})$ ，故应止于 X_{23} ，并换轴。再依以上原则迭代搜索，换轴，换向，直到

满足收敛条件时止。

由此可见，坐标轮换法迭代过程的基本逻辑为：不论从哪一个坐标轴、沿哪个方向出发，如第一步不满足函数值优化趋势，则应改变搜索方向。多步不满足函数值优化趋势，则应后撤一步，并以该点为起点，变换坐标轴，继续搜索，如此反复轮换，直到满足收敛条件时止。

根据以上逻辑，可以总结出坐标轮换法的基本设想是：将一个 n 维无约束优化问题，依次转变为一系列的一维问题求解。即先将 $(n-1)$ 个设计变量固定不动，只沿第一个变量的方向进行搜索，直到 $(K+1)$ 次迭代所得目标函数值，大于第 K 次迭代所得的目标函数值时止。第 K 次迭代所得的点，即新的 $X_{i,k}$ 点。然后将该点固定在 x_i 轴上，并保持其它变量不变，沿第二个变量的方向进行搜索。如此轮换，可得一系列新点 $X_{i,k}$ 。所得结果如已满足极小化精度要求，则停止计算。否则，再从 x_i 轴开始，直到取得满意的结果时止。如首次迭代发现目标函数并不下降，则应反向搜索，继续迭代。

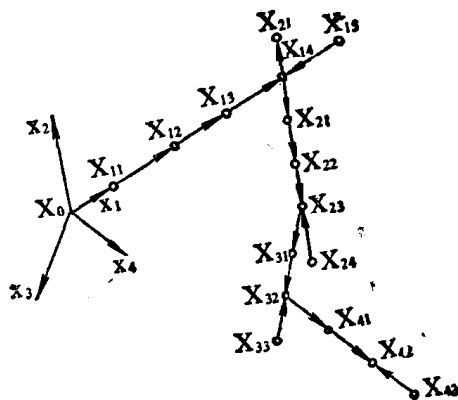


图 2-2

不难想象，坐标轮换法的基本迭代关系式，对于 i 轴， k 步时应为

$$\begin{aligned} X_{i,k} &= X_{i,(k-1)} + a_{i,k} S_i \\ i &= 1, 2, \dots, n \\ k &= 1, 2, \dots, K \end{aligned}$$

式中， $a_{i,k}$ 为步长因子， S_i 为迭代方向。轮换所取 n 维坐标的单位向量，即可代表该迭代方向，对于 i 轴即

$$S_i = e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

它表明，其中只有第 i 个单位坐标方向为 1，其余为零。迭代过程只在 i 轴方向进行，而其它设计参数不变。

步长因子 $a_{i,k}$ ，可以按随机选定法、加速步长法、或一维搜索法获得。如果事先已经了解设计变量的量纲及其最低计量单位，并有一定设计经验，用随机选定法，将十分方便。使用加速步长法时，将依迭代过程加大步长因子，它可提高收敛速度。如在迭代过程中减小步长因子，则可提高收敛精度。一维搜索法所得为理论最小步长，收敛精度最高，但收敛较慢。

收敛条件有二：

1) $|f(X) - f(X_{i,k})| \leq \varepsilon_1$ ，即以相对优化为出发点，只要相对于要求的目标函数降低的数值，满足收敛系数 ε_1 后，即停算。