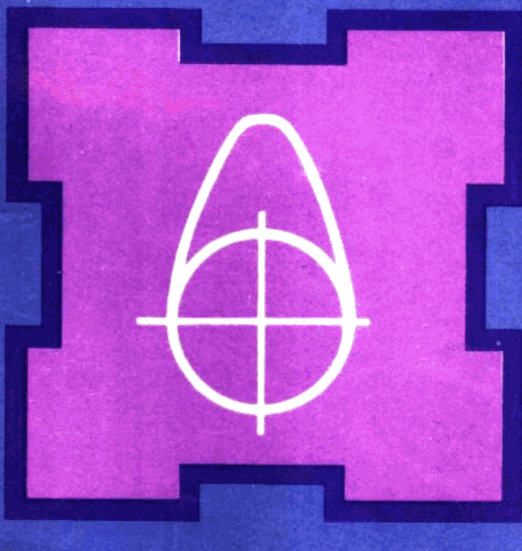
CAD在内燃机工程中的应用

谈荣望 主编　　吴昌华 主审



西南交通大学出版社

CAD在内燃机工程中的应用

谈 荣 望　主编
吴 昌 华　主审

西南交通大学出版社

内 容 简 介

本书介绍了最优化技术、有限单元法、计算机绘图等的基本原理和方法，及其在内燃机工程中的应用实例。这些内容避开了繁杂的数学推导，着眼于使用。各部分内容基本上彼此独立，自成系统，可以全读，也可以只用其中的部分内容。

本书可作为热能动力装置或内燃机专业本科高年级学生选修课教材或教学参考书，也可供相应工厂的技术人员阅读。

CAD 在内燃机工程中的应用
谈荣望　主编
*
西南交通大学出版社出版发行
（成都　二环路北一段　610031）
新华书店经销
成都飞机工业公司印刷厂印刷
*
开本：787×1092　1/16　印张：10.25
字数：244 千字　　印数：1—1000 册
1995 年 9 月第 1 版　1995 年 9 月第 1 次印刷
ISBN 7—81022—721—1/T · 128
定价：10.80 元

前 言

随着社会的不断发展，科学技术日新月异，到本世纪后期，世界的科技面貌发生了巨大的变化，过去神话式的境界已经变为现实，而且有许多已进入到人们的日常生活中。在这些先进的科学技术中，微电子技术可说是其中的佼佼者。

电子计算机是电子技术发展的产物，它以运算速度快、逻辑推理和判断能力强等优点，为科学技术解决了许多过去无法解决的问题。科学技术领域中采用计算机又为计算机的发展开辟了市场，二者相互影响，促进了科学技术和计算机的发展。

计算机在内燃机工程中的应用，国外开始于40年代末。现在在我国的内燃机行业中，也不同程度地采用了计算机，本书旨在介绍在内燃机工程中怎样利用计算机的。

CAD(Computer Aided Design)是在设计过程中，计算机只起辅助作用，而不是全自动的，需要人工干预，需要人们对方案和计算结果进行修改和判断。像内燃机这种动力机械，其零部件的形状、运动和受力都很复杂，它们的设计离开人们的实际经验和技巧，计算机还不能很好地完成任务，至少到目前为止是这样。

内燃机随其用途、功率大小等的不同，设计内容差别很大。既然只介绍CAD，这里只介绍利用计算机优势的内容，即分三篇来分别介绍最优化技术、有限单元法和计算机绘图。这些内容都互相独立，基本上都自成系统。我们知道，这些知识都有很丰富的内容。现在将它们归并在一本书内来介绍，是希望花不多的时间，让读者对CAD的内容有一个全面的了解，并结合这些新技术在内燃机工程中的应用，介绍一些实例。基于本书的这样的编写目的，所以只着重阐述有关的基本原理和方法，涉及的一些纯数学内容，以及相关更深入的内容，都没有作介绍。为了便于感兴趣的读者作深入的了解，在书的后面列有参考文献，可供读者选用。

由于本书的内容是利用计算机来完成的，所以读者在学习本书之前应对计算机的基本知识和算法语言要有一定的了解。要很好掌握本书的有关内容，应该强调上计算机的实践。限于篇幅，本书没有能提供相应的计算程序和习题，但对于在计算中需要判别后才能确定计算走向的内容，为了更清楚起见，书中有些地方，如第一篇，给出了相应的计算框图。如果读者需要相应的计算程序，可以和我们联系，以便提供。

本书可作为热能动力装置和内燃机专业本科高年级选修课教材或教学参考书，也可供有关工厂的技术人员阅读。

本书的第一章至第七章，由西南交通大学谈荣望编写，第八章至第十章由西南交通大学李人宪编写。全书由大连铁道学院吴昌华教授主审。在编写和审稿过程中，还得到上海铁道学院张宗才教授、北方交通大学宋国隆副教授的帮助和指导，在此一并表示深切的谢意。

本书涉及的内容较多，编者才疏学浅，疏漏和谬误之处在所难免，尚期读者和专家们不吝指正。

编 者

1993年9月于成都

目　　录

第二篇　有 限 单 元 法

第三篇　计 算 机 绘 图

绪　论

众所周知，设计是将某种设想变为现实的一项工作，它为实现这种设想提供一定的规范、工艺和施工过程。一般的机械设计需要通过调查研究，了解现有同类产品的性能、技术参数、使用情况和市场信息，从而确定新产品的性能、经济技术指标，再通过拟订方案、分析计算、绘图和编制技术文件等阶段后组织试生产。产品投放市场后，还要不断收集产品销售、利润和用户反映等情况，不断改进设计来提高产品质量，扩大产品销路。

评价设计优劣的主要依据是设计质量和设计速度。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是否恰当；设计速度则主要取决于设计方法及运算辅助工具。如果所用设计理论基本一致，则不同的设计方法对设计的影响将更加明显。可见，为了提高设计质量及设计速度，改进设计方法是极其重要的。

从以后几章的内容将会看到，以前曾经有一些理论，学者们公认是先进的，由于手段跟不上，这些理论无法在工程上应用。但计算机出现以后，这些理论便可以用于工程设计。更何况利用了计算机来推动科学技术发展，还会有新的理论出来不断丰富设计理论。另外，计算机的使用，可以替代许多繁重的人工劳动，因而可以大大提高设计速度和设计质量。所以计算机在设计领域中的应用，将大大改变设计面貌，使设计工作进入一个全新的阶段。

根据计算机在设计过程中所起的作用不同，人们把它分为自动设计——AD和计算机辅助设计——CAD。前者是在设计过程中不需要人工干预，由计算机根据编制好的程序，利用存储在计算机内的专家系统，自动地完成设计的各个阶段。目前，这种方法只适用于形状和受力简单的零件，如齿轮和一般轴类。后者则是在设计过程中要求随时检查方案和计算结果，并通过光笔和显示器对设计结果作出必要的修正，这种人、机结合的交互作业过程，构成了CAD过程，它把设计师与计算机组成一个解决问题的整体。这种设计方法是内燃机零部件的主要设计方法，也是一般非标准零部件的设计方法。

这种全新的CAD设计方法，可以用最优化技术来选定最佳设计方案。而在确定最优设计方案时，经常要用强度(应力)作为目标函数来进行优化，这就要利用计算机来进行有限单元法计算，确定强度。对现有产品的改进同样是这样。所以利用计算机进行辅助设计时，最优化技术和有限单元法成为不可缺少的两部分内容。

在新产品设计和现有产品的改进时，绘制零部件的工作图，是不可缺少而又很繁重的工作，由计算机来完成，不仅可以保证质量，而且可以很快出图。英国的里卡特(Ricardo)公司在1982年就已有40%的图纸由计算机来完成。在我国的有些部门已经规定，不是计算机绘制的零部件图，不能下到车间生产，或不能作为新产品验收，以保证生产过程中的产品质量。所以也把计算机绘图作为本书的一个重要内容。

以上是本书的基本内容。其中涉及到纯数学领域的内容，这里不作介绍，而是把着眼点放在使用上。

应用 CAD 最早并处于国际领先地位的是美国，各大计算机公司在 70 年代末不惜工本地装备自己的 CAD 系统；后来在许多小公司中也致力于开发小型的 CAD 系统。日本 CAD 的发展，走的是从美国引进→消化→自己开发的道路。日本《机械设计》杂志 1985 年第五期发表了一篇当时日本使用 CAD/CAM 系统现状的调查报告。接受调查的 247 家公司的行业分布如图 0—1 所示，公司的规模如图 0—2 所示。从图中可以看出，CAD 系统的采用是以大、中型企业为先导而发展起来的，这在很大程度上与公司所具有的技术水平、经济能力有关。

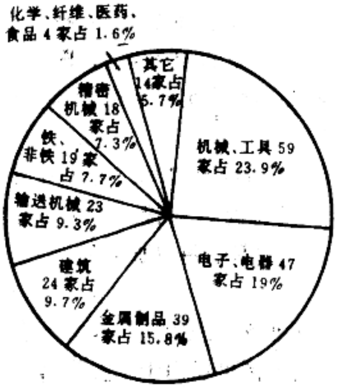


图 0—1 采用 CAD 系统的行业分布

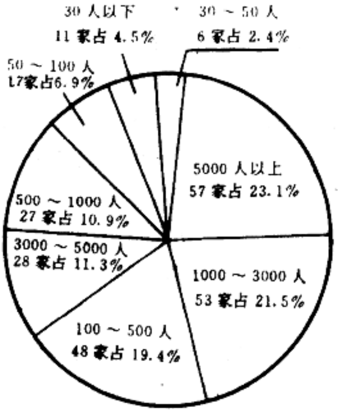


图 0—2 采用 CAD 系统各公司的规模

图 0—3 是被调查公司引进 CAD 系统的时间，由图可以看出，CAD 的高速发展是在后期，可以预见，这种趋势在今后还会保持下去。

应该说我国内燃机的 CAD 工作，从 80 年代开始到现在，做了不少工作，但发展很不平衡。总的来说，CAD 工作在科研单位和高等院校，在中、小型和微型计算机上，都早已开始了，并有许多有关论文发表。而企业部门，这方面的工作则要晚得多，特别是中、小型企业，有的甚至还没有开展这方面的工作。从 CAD 的内容方面来看，以前的工作主要是采用最优化技术和有限单元法计算或单纯的计算机绘图；将三者结合在一起，成为一个完整的 CAD 系统的工作，却进行得很少。本书将这三部分内容汇集在一起，奉献给读者，期望读者花费最少的时间，掌握一个完整的 CAD 系统知识，为深入开展 CAD 工作打下良好基础。

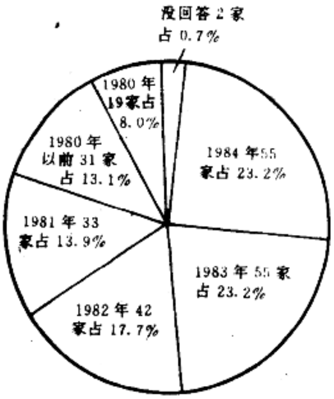


图 0—3 CAD 系统引进的时间

第一篇　优化技术

第一章　优化技术导论

§1—1　优化技术的概念

目前研究和应用较多的优化方法主要有准则优化法和数学规划法两大类，而每一类中又有若干具体的算法。为了叙述方便，现在先对若干主要方法加以分类说明。

一、准则优化法

准则优化法是从力学或物理的观点，规定一个评判“优”的准则，使设计出的结构遵循这个准则。现介绍几个主要的准则优化法。

1. 同时破坏准则

该准则规定，整个结构破坏时每个元件都同时达到了强度极限。由于要求每个元件同时达到强度极限，所以这种准则只能在单独一种载荷情况下的静定结构中实现。

2. 各种破坏同时发生准则(也称同步失效准则)

该准则主要是对结构中的各个元件而言的，特别是受压缩的薄壁元件。本准则的设计法认为，当元件具有多种可能的破坏形式时，如果能使各种破坏同时发生，则元件设计得最合理而且经济。

3. 满应力准则

该准则把同时破坏准则加以推广，使之能考虑多组载荷情况(以后简称多工况)，而且仅规定每个元件内的应力至少在一种工况下达到许用应力，并不要求所有元件同时达到，这就是满应力准则设计法。该法一般要经过多次的迭代计算，最后得出一个设计方案，其中每个元件至少在一种工况下能够达到满应力。

4. 应变能均匀分布准则

该准则认为：当结构的各个部分单位体积内贮藏的应变能都相等，并且等于材料单位体积容许的应变能值时，这样的结构便是最优的结构。

二、数学规划法

在优化设计中，数学规划法(又称最优化算法)是一种以数学规划为理论基础，以计算机为工具的优化方法。它的基本原理是求解极值问题，一般它是将问题转化为一个数学模型，包括确定目标函数和约束条件，接下来是根据数学模型中函数性质，选用合适的优化方法。然后

编制相应的计算程序，在计算机上求得最优值。最后根据计算结果进行分析和正确判断，得出最优设计。

动态规划是数学规划中的一个分支。解题过程是把问题分成若干个阶段，用一种递推关系式一个接一个依次作出最优决策，达到整个过程的最优化。所以它的求解方法是不同于求函数极值的微分法和变分法的。这种方法最适宜于处理桁架、梁和连续梁，因为这类结构可以分成若干串联式的分段，而且相互联接的关系又比较简单。

几何规划是数学规划中较新的一个分支，它是在目标函数的各个部分中，利用 n 个正数的几何平均值不大于它们的算术平均值这一定理，来寻求分配总目标值的最佳方案，而不是直接求变量的最优值。

整数规划是指设计变量的一部分或全部只能取整数或离散值的规划问题。在实际的工程结构优化设计问题中，经常要处理整数或离散设计变量问题，如桁架结构的杆件数只能是整数，而型钢的规格只能是一定型号的离散值等，所以整数规划具有重要的现实意义。

如上所述，准则法与数学规划法相比，准则法物理概念清楚、收敛速度快，算法的迭代次数与设计问题的规模关系不大。但是它们的适用范围较小，而且虽然在本世纪 50 年代就已提出了最早的力学准则（等强度准则），可是迄今对这种准则严格的成立条件一直缺乏科学系统的研究。而数学规划法则相反，它是从 40 年代末开始随着数学和计算机的发展而兴起的一门新兴技术，它严密、精确。这几十年来，进展很快，方兴未艾，本书主要是介绍这种方法。对其中几何规划和整数规划在第四章作简要介绍。

需要指出，数学规划法在进行多工况、多约束、多变量的大型结构优化设计中，即使采用大型计算机，有时也难得到满意结果。这时人们就力图减少结构分析的重复计算，即采用近似分析概念，在结构性态向量空间中直接搜索可行点和最优点，设计基本模式成为“选点、检验、优化、再选点”，这种方法就是将准则法最优性条件和数学规划选优方法相结合的方法。

下面来看一个力学的最优化例子。如图 1—1 所示三杆桁架，设三杆的材料相同，其密度为 ρ，截面积分别为 A_1，A_2 和 A_3。第一种情形的载荷为沿杆 1 的力 P（载荷 1），第二种情形的载荷为沿杆 3 的力 P（载荷 2）。试求在受载荷时桁架的最小重量 G。

每杆在受载荷时的应力应小于许用应力，即 $\sigma_i \leqslant [\sigma]$。在本例中，字母右下方的注角表示杆件号，即 $i = 1,2,3$，而右上方的注角表示哪一种情形的载荷。

设材料的许用应力 $[\sigma]$ 与应力的关系为

$$\sigma_i = \begin{cases} [\sigma] & \text{拉伸应力} \\ -[\sigma]/2 & \text{压缩应力} \end{cases}$$

宏观分析可知，杆长 $l_1 = l_3 = \sqrt{2}\,l$，而且桁架最轻时，杆截面 $A_1 = A_3$，因此实际上只需确定 A_1，A_2。而最小重量为 $G = \rho l(2\sqrt{2}A_1 + A_2)$。

利用力系的静力平衡和位移（或力）法分析，可得

$$\begin{aligned}
&\sigma_1 + \sigma_3 = \sigma_2 \\
&\sigma_1^1 = \sigma_3^2 = (A_1 + A_2/\sqrt{2})P/A_1(A_1 + \sqrt{2}A_2) \\
&\sigma_2^1 = \sigma_2^2 = P/(A_1 + \sqrt{2}A_2) \\
&\sigma_3^1 = \sigma_1^2 = -A_2P/\sqrt{2}/A_1(A_1 + \sqrt{2}A_2)
\end{aligned}$$

以杆的截面 A_2 和 A_1 为纵横坐标绘制图 1—2。在图中，绘出杆件上的应力等于材料许用应力时的曲线 BCD，BF 和 OCE。杆件的截面要大于零，这样就可以确定杆件截面应取在

BCE（或称为边界线 Γ）上或其右上方，这个区域称为可行域。

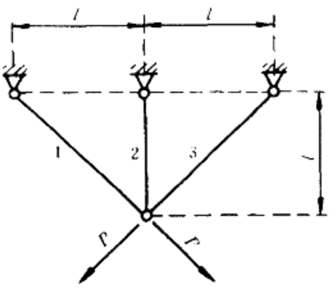


图 1—1　三杆桁架

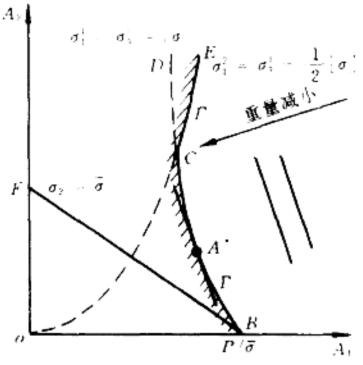


图 1—2　二维问题的图形解

分析重量式 $G = \rho l(2\sqrt{2}A_1 + A_2)$ 可以看出，$A_2 = 1/(\rho l) - 2\sqrt{2}A_1$，当 ρl 一定时，它是一个等截距的直线方程式，不同的重量 G，对应于一定的直线，图上给出了两条不同重量 G 的平行斜线，称为等重量线。重量越小，等重量线越接近坐标轴的顶点。这组等重量线中与边界线 Γ 相切的那点 A^*，就是这个桁架中结构最轻，而杆件中应力小于许用应力的最优方案，它对应于 $A_1^* = (1/2)(P/[\sigma])(1 + 1/\sqrt{3})$，$A_2^* = (P/[\sigma])(1/\sqrt{6})$，最小重量为 $G = Pl(2 + \sqrt{3})/\sqrt{2}$。

从上例可以看出，构成三杆桁架的杆件截面的可能性很多，但是满足设计要求的最优方案只有一个。优化技术就是要在众多的可能方案中选出最优的方案来。更确切地说，优化技术就是要在一定限制（约束）条件下，寻求一组设计参数（变量），使设计对象的某项设计指标（目标函数）达到最优。

三、最优化技术中常用的几个术语

1. 设计变量 x

描述一个设计方案需要用到许多量，例如材料的面积、弹性模量和应力值。如果考虑经济指标，设计变量还应包括单位体积（或重量）的价格，不同材料或不同构件形状与尺寸的加工费用等。在这些描述结构设计方案的量中，一部分是根据结构工作的条件决定的，在调整设计方案的过程中不能改变。另外一部分在优化设计过程中需要反复调整，以使设计出的结构在某种追求的目标方面最理想，这种量称为设计变量。如上述例子中的截面积 A 就是设计变量。在优化设计过程中，取哪些量作为设计变量，这与许多因素有关。首先要根据所追求的目标来分析。例如目标是使结构的重量最轻，就应选择对重量影响最大的参数作为设计变量。至于设计变量的个数如何确定，这既要考虑问题的需要，又要考虑采用的优化算法的特点，同时还要注意到计算机存储量和运算速度等等。一般来说，设计变量越多则越能发挥结构优化的作用，带来的经济效益越大，但计算量也越大，计算费用也越高。

每一个设计变量 x 都可以看成向量 X 的一个元素，即 $X = \{x_i\}(i = 1,2,\cdots,n)$。

2. 目标函数 $f(X)$

目标函数又称评价函数。它是用来评价目标优劣的数学关系式，自变量是设计变量。上述

例子中的 $G = \rho l(2\sqrt{2}A_1 + A_2)$ 就是目标函数。目标函数的求优分两种情况：求极小化（成本、功耗、尺寸、重量等）$f(X) \to \min$ 或求极大化（如承载能力、效率、可靠性等）$f(X) \to \max$。因为 $f(X)$ 的极大化等价于 $1/f(X)$ 或 $-f(X)$ 的极小化，所以在优化设计中通常均用目标函数极小化这种形式。可供选择的目标有：强度、刚度、耐磨性、效率、可靠性、寿命极大，或成本、重量、功耗、温度、动态响应、轨迹误差极小等。在设计过程中，可能只追求某一单项指标最优（单目标优化），也可能追求若干项的综合指标最优（多目标优化）。前者理论较成熟，应用较广。

3. 约束条件（$g_j(X) \leqslant 0$ 或 $h_e(X) = 0$）

约束条件就是在最优化设计过程中，对设计变量所加的各种限制。约束条件可表示成不等式约束

$$g_j(X) = g_j(x_1, x_2, \cdots, x_n) \geqslant 0$$

或

$$g_j(X) = g_j(x_1, x_2, \cdots, x_n) \leqslant 0 \qquad (j = 1, 2, \cdots, m)$$

和等式约束

$$h_e(X) = h_e(x_1, x_2, \cdots, x_n) = 0 \qquad (e = 1, 2, \cdots, p < n)$$

式中　m —— 施加于该项设计的不等式约束件数；

　　　p —— 施加于该项设计的等式约束件数。

一般来说，约束还可分为几何约束和性态约束。属于几何约束的如：对第 i 个设计变量进行的单边界限约束可表达为

$$x_i \geqslant 0 \qquad \text{（非负约束）}$$

$$x_i - x_{i\min} \leqslant 0 \qquad \text{（上界约束）}$$

$$x_{i\min} - x_i \leqslant 0 \qquad \text{（下限约束）}$$

而对其进行的双边界限约束则表达为

$$x_{i\min} \leqslant x_i \leqslant x_{i\max} \qquad \text{（上、下界约束）}$$

式中　$x_{i\min}$ —— 第 i 个设计变量的下限；

　　　$x_{i\max}$ —— 第 i 个设计变量的上限。

性态约束指结构在外因（如载荷、温度、振动等）作用下各种响应值与设计规范中的许用值的关系，如

强度约束为：$\sigma - [\sigma] \leqslant 0$

刚度（或位移）约束为：$\delta - [\delta] \leqslant 0$

稳定性约束为：$p - [p_{临界}] \leqslant 0$

固有频率约束为：$[\omega_{\min}] - \omega \leqslant 0$

在最优化设计中，不等式约束经常遇到。一个不等式约束（如 $g_j(X) \leqslant 0$）将设计空间划分成两部分，一部分满足约束条件（如 $g_i(X) < 0$），另一部分不满足约束条件（如 $g_i(X) > 0$）。这两部分的分界面称为约束面（如 $g_j(X) = 0$）。若某个问题有 m 个不等式约束条件，则由 m 个约束在设计空间中围成两个区。凡满足不等式约束方程组（如 $g_j(X) \leqslant 0$）的设计变量选择区，称为可行域。而不满足不等式约束方程组中任一约束条件的设计变量选择区，称为非可行域。图 1—2 中 Γ 线右侧为可行域，左侧为非可行域。

设计问题各式各样，优化技术随设计问题的不同，可以采用不同的方法。对于简单的问题，设计人员可以凭掌握的知识、经验和逻辑推理来直觉判断。对于影响设计指标参数的主次尚不清楚的问题，可以选用几个参数范围来做试验，通过试验结果的分析来进行优化，这就是探索

试验优化。对于复杂问题，采用探索试验优化，不仅花费时间和金钱，而且也难得到精确结果。为此，人们力争寻找科学的方法。

对于机械设计领域，优化技术能比较好地把各种现代设计理论和经过长期实践验证的设计内容结合在一起。

机械优化技术与常规设计有着密不可分的内在联系。常规设计中所涉及到的各种设计准则、设计规范和设计数据等，是建立优化技术数学模型的依据。常规设计中的主要内容，仍然是优化技术中的重要组成部分。优化技术显著区别于常规设计之处，主要是前者利用了先进的数学方法（如数学规划论）和计算机，使得分析过程能快速地、自动地进行多种方案计算、比较和寻优。在优选方案时，考虑问题更全面，计算速度更快，分析结果更加准确。而后者则主要依赖设计者个人的才智和经验，在有限的设计方案中，经过分析比较、判断，确定出相对满意的设计结果。

应当指出，优化技术虽然更多地依赖于技术手段 —— 计算机的硬、软件，但设计人员的经验仍然极为重要，必不可少。无论在建立优化的数学模型、选用合适的优化方法、编制计算机软件，以及对计算结果进行分析等过程的哪个环节中，都离不开设计人员决定性作用。因此，一个经验丰富的机械设计工作者，只要掌握了优化技术的基本技能，就不难成为一名机械优化设计者。

对某一项设计任务来说，可能对其中主要部分进行优选，而对另一部分则不要求优选，这时就要进行优选和常规两部分设计内容。

优化设计的主要特点表现在求优过程完全由计算机按规定内容和顺序自动进行，经过反复迭代直至满足预定的计算精度，输出最优解为止。

目前数学规划法已经有很多有效的算法，根据目标函数和约束特性的不同可以分为：

(1) 线性规划　当 $f(X)$、$g_j(X)$ 和 $h_e(X)$ 均为线性函数时，这种规划称为线性规划。

(2) 二次规划　当 $f(X)$ 为二次函数，$g_j(X)$ 和 $h_e(X)$ 均为线性函数时，称为二次规划。

(3) 非线性规划　只要 $f(X)$、$g_j(X)$ 和 $h_e(X)$ 中有一个是非线性的，则称为非线性的。

如果按有无约束条件来分类，可以分为无约束最优化问题（即无约束规划）和有约束最优化问题。结构的优化设计问题，大多为有约束的。

§1—2　最优化技术的数学和几何描述

从数学角度来说，最优化设计就是要求出一组设计变量 $x_1, x_2, \cdots, x_n$，使目标函数最小（或最大），即

$$f(X) \to \min \ (\text{或} \ \max)$$

且这组设计变量必须满足约束条件

$$g_j(X) \leqslant 0 \qquad (j = 1, 2, \cdots, m)$$

和

$$h_e(X) = 0 \qquad (e = 1, 2, \cdots, p)$$

满足上述要求的一组设计变量值 $X^* = [x_1^* \quad x_2^* \quad \cdots x_n^*]^T$ 称为最优化点，而这时的目标函数 $f(X^*)$ 的值，称为最优化值。最优化点和最优化值，构成一组最优化解。

最优化设计问题，也可以用几何图形来描述，图 1—2 即为一例。图 1—3 也是一个两设计变量 x_1 和 x_2 最优化设计的几何描述。图中两条约束线 $g_1(X) \geqslant 0$（即 AKB 线）和 $g_2(X) \geqslant$

0(即 CKD 线)，形成约束边界线 AKD，这条边界线的右上角成为可行域，这条线左下角(即阴影线部分)成为非可行域。其中 X 点为可行设计点，而 $\overline{X}$ 点为非可行设计点。图中斜直线为线性的目标函数 $f(X)$ 的等值线，且 $f_1 > f_2 > \cdots$ 约束边界线与目标函数等值线的切点(K 点)即为最优化设计点 X^*，而这条等值线即为最优化值 $f(X^*)$。

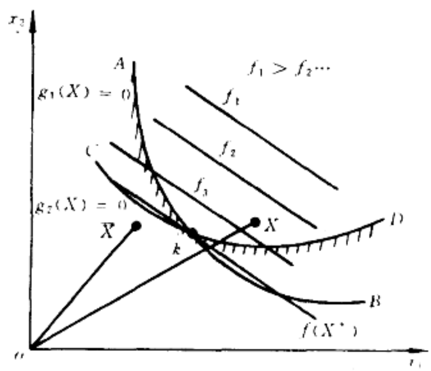


图 1—3　两个设计变量的最优化设计图形

图 1—4(a) 和 (b) 分别表示了目标函数 $f(X)=(x_1-2)^2+x_2^2$ 和约束数 $g_1(X)$，$g_2(X)$ 和 $g_3(X)$ 的立体图和最优化问题的平面图。当目标函数 $f(x_1,x_2)=1/4,1,\cdots$ 时，即在(x_1，

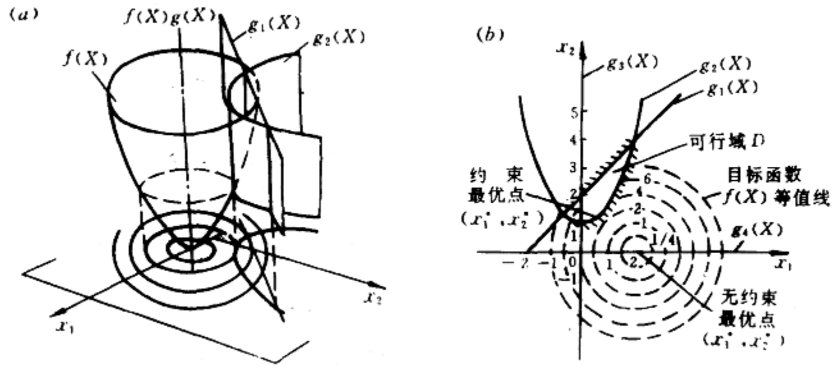


图 1—4　二维非线性最优化问题图形

x_2) 坐标平面内得到一系列平面曲线(同心圆)的等值线，它表示了目标函数值的变化情况，最里层的代表目标函数的最小值。由约束面：直线 $g_1(X)=0$、$g_3(X)=0$、$g_4(X)=0$ 和抛物线 $g_2(X)=0$ 组成的阴影线里侧的区域即为可行域 D。约束边界与等值线的切点 X^* 即为约束最优化点，$x_1^*=0.58$，$x_2^*=1.34$，$f(x_1^*,x_2^*)=3.80$ 就是约束最优解。无约束最优解就是目标函数等值线的中心。

n 个设计变量的最优化问题，可以组成一个 n 维的设计空间，该空间中的任意一个点 X 就代表一个设计方案。目标函数 $f(X)$ 为某一定值时，则在 n 维空间中形成一个等值超曲面。同样道理，每一个约束条件在 n 维空间中形成一个超曲面，这些超曲面组合成约束边界面，它们将设计空间分为可行域和非可行域。可行域中的点对应于可行的设计方案，但目标函数不一定是最优的。非可行域中的点对应于非可行的设计，这种设计至少有某一个约束条件不能满足。目标函数的等值超曲面与约束边界的切点，便是最优化问题的解。

曲线极大值和极小值的概念早已熟悉了，这里要介绍函数凸性的概念，有了这个概念后，才能讨论最优解。

设 D 为 R^n 中的一个集合，若对任意两点 $X^{(1)} \in D$，$X^{(2)} \in D$，并且连接它们的线段仍在 D 中。对任意实数 $0 \leqslant a \leqslant 1$，使连接线 $aX^{(1)}+(1-a)X^{(2)} \in D$，则称 D 为凸集，如图 1—5(a) 所示；否则 D 为非凸集，如图 1—5(b) 所示。

设 D 为 R^n 中的一个凸集，$f(X)$ 为定义在 D 上的函数，若对任意实数 $0 \leqslant a \leqslant 1$，以及 D 中任意两点 $x^{(1)}$ 和 $x^{(2)}$ 恒有

$$f(ax^{(1)} + (1-a)x^{(2)}) \leqslant af(x^{(1)}) + (1-a)f(x^{(2)})$$

则称 $f(X)$ 为定义在凸集 D 上的一个凸函数，也称为单峰函数。它的几何意义是：对于单变量 $f(x)$ 的情况，如图 1—6 所示。若连接函数曲线上任意两点的直线段，某一点 x 的函数值恒低于此直线段上相应的纵坐标值。

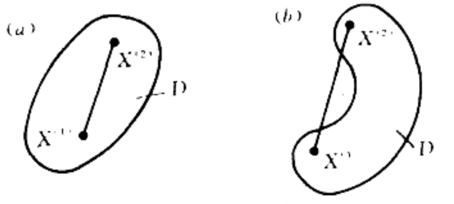


图 1—5　凸集的概念

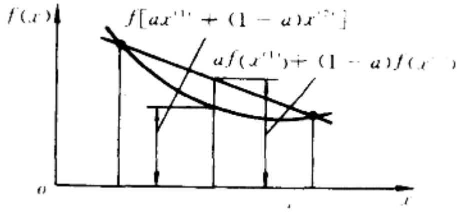


图 1—6　单变量的单峰函数

对于 $\min f(x)(x \in R^n)$，满足约束条件 $g_j(x) < 0(j = 1,2,\cdots,m)$ 的数学规划（最优化）问题，若 $f(x)$ 是凸函数，g_i 是凸函数，这样的数学规划问题，称为凸规划。

现在再回过来介绍最优化技术：

在无约束情况下的最优化问题，实际上是求一个多元函数 $f(X)$ 的极小值问题。即

$$\min f(X) = f(X^*)$$

使目标函数 $f(X)$ 到达极值的向量端点 $X^* = [X_1^*, X_2^*, \cdots, X_n^*]^T$ 就是最优点，而 $f(X^*)$ 为最优值。

在求目标函数极小值时，对于单峰函数，最优点就是极小点。但对于具有多峰的函数来说，由于它具有两个以上的极值点，每个极值点都只对它附近的点来说是极值，而在全域内的多个极值点中，究竟哪一个极值点是全局的最优点，需要从多个极值点中来确定。图 1—7 表示了两个极值点的情况。其中图 (a)，X_1 和 X_2 都是极值点，但 $f(X_1) < f(X_2)$，所以 X_2 是局部最优点，而 X_1 为全局最优点。其中图 (b) 表示了具有鞍形点目标函数的局部最优点和全局最优点。

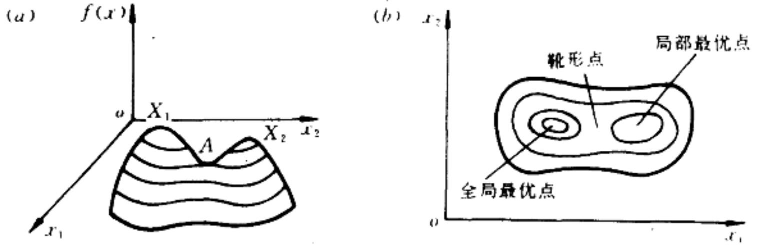


图 1—7　无约束最优化问题的局部最优和全局最优

在约束优化问题中，情况就复杂得多，它不仅与目标函数的性质有关，还与约束条件的函数性质有关。如图 1—8 所示，目标函数 $f(X)$ 有两个不等式约束条件 $g_1(X) \geqslant 0$ 和 $g_2(X) \geqslant 0$，构成两个可行域 D_1 和 D_2。在 D_1 内的 X_1^* 是目标函数的最小值点，在 D_2 内有 X_2^* 和 X_3^* 是目标函数的最小值点。但是 $f(X_1^*) > f(X_2^*) > f(X_3^*)$，所以 X_3^* 是全局最优解点，而 X_1^* 和 X_2^* 是局部最优解点。

目标函数面是凸性或凹性，对于判定最优化解的唯一性具有重要意义。对于无约束的最优化问题，如果目标函数是单峰的，它只有一个最优解，即最优解是唯一的。对于约束的最优化问题，如果目标函数是单峰的，而约束面是凸的，局部最优解和全局最优解将重合。凸规划的局部最小值解一定是全局最小值解。除此之外的最优化解，一般可以取不同的初始点，看最后是否收敛到同一个点上。如果是多峰函数，出现了多个最优点，则要如上所述，由这些最优点的目标函数值作比较来确定全局的最优解。一般来说，要确定函数的凸性也是很费事的。

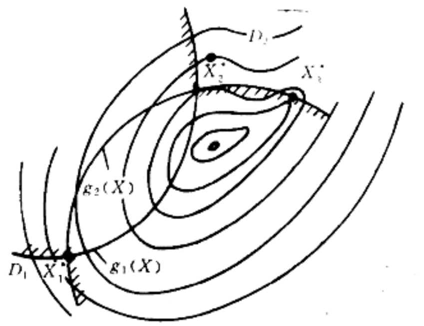


图 1—8　约束问题的局部最优和全局最优

§1—3　最优化设计的迭代过程

对于无约束的优化问题

$$\min f(X) \qquad (X \in R^n)$$

为求它的极小点，必须先求其梯度并等于零，即

$$\nabla f(X) = 0$$

求出其稳定点。然后利用充分条件进行判别。当函数 $f(X)$ 较为复杂时，这个求解和判别过程也较困难甚至无法实现。因此，对于这类问题，一般采用从直接分析目标函数 $f(X)$ 的特征出发，对一些点的函数值来进行搜索。这种数值计算寻优方法的基本思路有消去法和爬山法两类。黄金分割法可作为消去法的例子。这种方法在处理单变量函数的极值问题时很有效，它是将极值点所在的区间不断按比例缩小，直至当区间缩小到与极值点足够靠近时，即认为近似求得了最优解。但对于多变量函数，由于消去的不是一个线段，而是消去平面、立体甚至多维空间的一部分，这就使得求解问题变得很复杂。在这种情况下，爬山法就比较好。爬山法就是将求函数极值的过程比喻为向山的顶峰攀登的过程。当然，“山顶”可以是函数的极大值或极小值。前者称为上升算法，后者称为下降算法，它们的共同点是每前进一步应该使目标函数的值有所改善，同时还要为下步移动方向提供有用的信息。

上述搜索过程是通过所谓迭代法来实现的。它是利用计算机工作特点的一种数值计算方法，一般由以下四部分组成：

(1) 选择初始近似点 $X^{(0)}$，如图 1—9 所示。

(2) 选择一个方向 S_1 为搜索方向，使目标函数值沿方向 S_1 下降。

(3) 在方向 S_1 上选定步长因子 α，使 $X^{(1)} = X^{(0)} + \alpha S$(如 X_{11})，看计算函数值是否有所改善。如有改善，继续进行。当进行到函数值反而增大时(如 X_{15})，则退回到前一点(如 X_{14})，然后改变方向(如 S_2 方向)，再一步步前进，直到下一个点，函数值又增加(如 X_{24})，再退回到前一点(如 X_{23})，再改变方向搜索。

(4) 这样不断搜索，直至第 $k+1$ 个搜索点 X_{k+1} 与第 k 个搜索点 X_k，比较其矢量模满足不等式

$$\| X_{k+1} - X_k \| < \varepsilon$$

或者目标函数绝对值 $|f(X_{k+1}) - f(X_k)| < \varepsilon$ 为止。其中 ε 为标志计算精度要求的足够小的正数。

当然，这个极值点是局部最优点还是全局最优点，这可用 §1—2 中所述方法来判别。

可见，搜索算法是一种迭代过程，搜索方向和步长因子构成了每一次迭代的修正量。

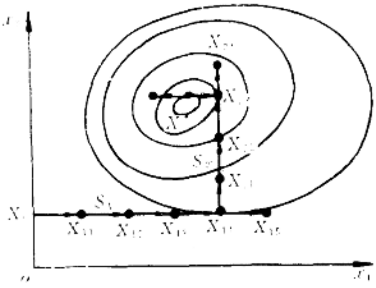


图 1—9　搜索线路示意图

目前已有的具体的优化计算方法很多，其区别在于确定方向和步长的方法不同。这些方法大致分为两大类：

(1) 直接法　这种方法只进行目标函数值的计算与比较来确定优化方向和步长。例如一维搜索中的黄金分割法、二次插值法等。又如多维问题中的随机法、共轭方向(Powell) 法、复合形法等。

(2) 间接法　这种方法是不直接利用目标函数而是用它的一阶、二阶偏导数或其他形式来确定最优化方向和步长。例如梯度法、共轭梯度法和惩罚函数法等。

第二章和第三章将根据有无约束条件来介绍，从中可以具体了解直接法和间接法是如何求解最优化问题的。