

第 17 篇 优 化 设 计

主 编 孙靖民
编写人 孙靖民
米成秋
柯尊忠

第 1 章 优化设计概述

现代化的设计工作已不再是过去那种凭借经验或直观判断来确定结构方案,也不是像过去“安全寿命可行设计”方法那样:在满足所提出的要求的前提下,先确定结构方案,再根据安全寿命等准则,对该方案进行强度、刚度等的分析、校核,然后进行修改,以确定结构尺寸。而是借助电子计算机,应用一些精确度较高的力学的数值分析方法(如有限元法等)进行分析计算,并从大量的可行设计方案中寻找出一种最优的设计方案,从而实现用理论设计代替经验设计,用精确计算代替近似计算,用优化设计代替一般的安全寿命的可行性设计。

优化方法不仅用于产品结构的设计、工艺方案的选择,也用于运输路线的确定、商品流通量的调配、产品配方的配比等等。

机械优化设计包括建立优化设计问题的数学模型和选择恰当的优化方法与程序两方面的内容。由于机械优化设计是应用数学方法寻求机械设计的最优方案,所以首先要根据实际的机械设计问题建立相应的数学模型,即用数学形式来描述实际设计问题。在建立数学模型时需要应用专业知识确定设计的限制条件和所追求的目标,确立各设计变量之间的相互关系等。机械优化设计问题的数学模型可以是解析式、试验数据或经验公式。虽然它们给出的形式不同,但都是反映设计变量之间的数量关系的。

数学模型一旦建立,机械优化设计问题就变成一个数学求解问题。应用数学规划方法的理论,根据数学模型的特点,可以选择适当的优化方法,进而可以选取或自行编制计算机程序,以计算机作为工具求得最佳设计参数。

1 机械优化设计问题示例

在优化设计中,通常是根据分析对象的设计要求,应用有关专业的基础理论和具体技术知识进行推导来建立相应的方程或方程组。对机械类的分析对象来说,主要是根据力学、机械设计基础知识和各专业机械设备的具体知识来推导方程或方程组(动力学问题中多为偏微分或常微分方程组的形式),这些方程反映结构诸参数之间的内在联系,通过它可以研究各参数对设计对象工作性能的影响。

下面通过几个具体例子,说明机械优化设计中建立方程组的方法和步骤(公式的推导尽量简略,以减少

篇幅)。

例 1 平面四连杆机构的优化设计。

平面四连杆机构的设计主要是根据运动学的要求,确定其几何尺寸,以实现给定的运动规律。

图 17.1-1 所示是一个曲柄摇杆机构。图中 x_1 、 x_2 、 x_3 、 x_4 分别是曲柄 AB 、连杆 BC 、摇杆 CD 和机架 AD 的长度。 φ 是曲柄输入角, ψ_0 是摇杆输出的起始位置角。这里,规定 φ_0 为摇杆在右极限位置角 ψ_0 时的曲柄起始位置角,它们可以由 x_1 、 x_2 、 x_3 和 x_4 确定。通常规定曲柄长度 $x_1=1.0$, 而在这里 x_4 是给定的,并设 $x_4=5.0$, 所以只有 x_2 和 x_3 是设计变量。

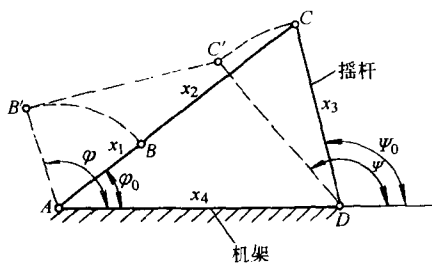


图 17.1-1 曲柄摇杆机构

设计时,可在给定最大和最小传动角的前提下,当曲柄从 φ_0 位置转到 φ_0+90° 时,要求摇杆的输出角最优地实现一个给定的运动规律 $f_0(\varphi)$ 。例如,要求

$$\psi = f_0(\varphi) = \psi_0 + \frac{2}{3\pi}(\varphi - \varphi_0)^2$$

对于这样的设计问题,可以取机构的期望输出角 $\psi = f_0(\varphi)$ 和实际输出角 $\psi_j = f_j(\varphi)$ 的平方误差积分准则做为目标函数,使 $f(x) = \int_{\varphi_0}^{\varphi_0+\frac{\pi}{2}} [\psi - \psi_j]^2 d\varphi$ 最小。

当把输入角 φ 取 s 个点进行数值计算时,它可以化约为 $f(x) = f(x_3, x_4) = \sum_{i=0}^s [\psi_i - \psi_{ji}]^2$ 最小。

相应的约束条件有:

(1) 曲柄与机架共线位置时的传动角

最大传动角 $\gamma_{\max} \leq 135^\circ$

最小传动角 $\gamma_{\min} \geq 45^\circ$

对本题可以计算出

$$\gamma_{\max} = \arccos \left[\frac{x_2^2 + x_3^2 - 36}{2x_2x_3} \right]$$

$$\gamma_{\min} = \arccos \left[\frac{x_2^2 + x_3^2 - 16}{2x_2x_3} \right]$$

所以

$$x_2^2 + x_3^2 - 2x_2x_3\cos 135^\circ - 36 \geq 0$$

$$x_2^2 + x_3^2 - 2x_2x_3\cos 45^\circ + 16 \geq 0$$

(2) 曲柄存在条件

$$x_2 \geq x_1$$

$$x_3 \geq x_1$$

$$x_4 \geq x_1$$

$$x_2 + x_3 \geq x_1 + x_4$$

$$x_4 - x_1 \geq x_2 - x_3$$

(3) 边界约束

当 $x_1 = 1.0$ 时, 若给定 x_4 , 则可求出 x_2 和 x_3 的边界值。例如, 当 $x_4 = 5.0$ 时, 则有曲柄存在条件和边界值限制条件如下

$$x_2 + x_3 - 6 \geq 0$$

$$4 - x_2 + x_3 \geq 0$$

和

$$1 \leq x_2 \leq 7$$

$$1 \leq x_3 \leq 7$$

例2 齿轮减速器的优化设计。

齿轮减速器是一种应用广泛的传动装置。传统的设计方法虽已较完善, 但它们多属校核性质的, 即从给定的条件出发, 根据经验类比和理论计算, 用试凑的方法确定主要参数, 然后进行强度、刚度等方面的校核。如不合格, 则对某些参数进行修改后, 再重复上述过程, 直至满足各项要求为止。显然, 这种方法不能保证得到最优设计方案。

这里通过一个常见的二级圆柱齿轮减速器(其传动简图如图 17.1-2 所示), 说明在对它进行优化设计时, 建立其相应的数学模型的方法。设计时, 通常给定传递的功率 P 、总传动比 i 和输出的转数 n 。要求在满足强度的条件下, 使其体积最小, 以达到使结构紧凑、质量最小的目的。

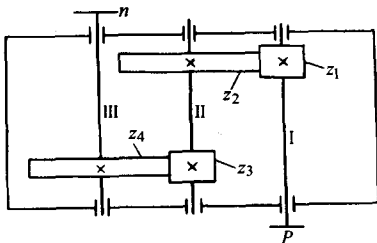


图 17.1-2 减速器传动简图

从图 17.1-2 的减速器传动简图中可以看出, 它由两对圆柱齿轮传动共四个齿轮组成, 它们的齿数分别为 z_1 、 z_2 、 z_3 和 z_4 , 相应的齿数比分别是 $i_1 = \frac{z_2}{z_1}$ 和 $i_2 = \frac{z_4}{z_3}$; 两组传动齿轮的法面模数分别设为 m_{n1} 和 m_{n2} ;

齿轮的螺旋角为 β 。在这里 z_1 、 z_2 、 z_3 、 z_4 、 i_1 、 i_2 、 m_{n1} 、 m_{n2} 和 β 都是设计参数。但由于设计时已给定总传动比 i , 且有 $i = i_1 i_2$, 所以 $i_2 = \frac{i}{i_1}$ 。从而四个齿轮的齿数只要能确定两个即可。例如, 我们可以定两个小齿轮的齿数 z_1 和 z_3 为设计变量。因此, 这个优化设计问题的独立设计变量为: z_1 、 z_3 、 m_{n1} 、 m_{n2} 、 i_1 和 β 六个。可见不是所有的设计参数都是设计变量。

上面提到, 设计时要使该减速器的体积最小, 这就是本优化设计问题追求的目标函数。它可以归结为使减速器的总中心距 A 为最小, 写成

$$A = \frac{1}{2\cos\beta} [m_{n1} z_1 (1 + i_1) + m_{n2} z_3 (1 + i_2)] \rightarrow \min \quad (17.1-1)$$

保证总中心距 A 为最小时应满足的条件是本优化设计问题的约束条件, 它们是: 齿面的接触强度和齿根的弯曲强度以及中间轴 II 上的大齿轮 z_2 不与低速轴 III 发生干涉。

1) 齿面接触强度计算给出

$$\frac{[\sigma_H]^2 \psi m_{n1}^3 z_1^2 i_1}{6.845 \times 10^6 K_1 T_1} - \cos^3 \beta \geq 0 \quad (17.1-2a)$$

和

$$\frac{[\sigma_H]^2 \psi m_{n2}^3 z_3^2 i_2}{6.845 \times 10^6 K_2 T_2} - \cos^3 \beta \geq 0 \quad (17.1-2b)$$

式中 $[\sigma_H]$ ——许用接触应力;

T_1 ——高速轴 I 的扭矩;

T_2 ——中间轴 II 的扭矩;

K_1 、 K_2 ——载荷系数;

ψ ——齿宽系数。

2) 齿根弯曲强度计算给出

高速级大、小齿轮的齿根弯曲强度条件为

$$\frac{[\sigma_w]_1 \psi Y_1}{3K_1 T_1} (1 + i_1) m_{n1}^3 z_1^2 - \cos^2 \beta \geq 0 \quad (17.1-3a)$$

和

$$\frac{[\sigma_w]_2 \psi Y_2}{3K_2 T_2} (1 + i_2) m_{n2}^3 z_3^2 - \cos^2 \beta \geq 0 \quad (17.1-3b)$$

低速级大、小齿轮的齿根弯曲强度条件为

$$\frac{[\sigma_w]_3 \psi Y_3}{3K_3 T_3} (1 + i_1) m_{n1}^3 z_3^2 - \cos^2 \beta \geq 0 \quad (17.1-3c)$$

$$\frac{[\sigma_w]_4 \psi Y_4}{3K_4 T_4} (1 + i_2) m_{n2}^3 z_1^2 - \cos^2 \beta \geq 0 \quad (17.1-3d)$$

式中 $[\sigma_w]_1, [\sigma_w]_2, [\sigma_w]_3$ 和 $[\sigma_w]_4$ —— 分别是齿轮 z_1, z_2, z_3 和 z_4 的许用弯曲应力;

Y_1, Y_2, Y_3, Y_4 —— 分别是齿轮 z_1, z_2, z_3 和 z_4 的齿形系数。

3) 根据不干涉条件, 有

$$\frac{m_{n1} z_3 (1 + i_1)}{2 \cos \beta} - \left(m_{n1} + \frac{m_{n1} z_1 i_1}{2 \cos \beta} + s \right) \geq 0$$

式中 s —— 低速轴 III 的轴线和中间轴 II 上大齿轮 z_2 齿顶间的距离。可取 $s = 5 \text{ mm}$ 。则得

$$m_{n1} z_3 (1 + i_1) - 2 \cos \beta (5 + m_{n1}) - m_{n1} z_1 i_1 \geq 0 \quad (17.1-4)$$

4) 另外, 还要考虑传动平稳、轴向力不宜过大、高速级与低速级的大齿轮 z_3 和 z_4 浸油深度大致相同、小齿轮分度圆尺寸不能太小等因素, 来建立一些边界约束条件

$$a_i \leq x_i \leq b_i \quad (17.1-5)$$

式中, $i = 1, 2, \dots, 6$ (6 是设计变量的个数)。这样, 则可写出二级圆柱齿轮减速器优化设计的数学模型如下

$$A = \frac{1}{2 \cos \beta} [m_{n1} z_1 (1 + i_1) + m_{n1} z_3 (1 + i_1)] \rightarrow \min$$

$$\text{s. t. } \frac{[\sigma_H]^2 \psi m_{n1}^3 z_1^3 i_1}{6.845 \times 10^6 K_t T_1} - \cos^3 \beta \geq 0$$

$$\frac{[\sigma_H]^2 \psi m_{n1}^3 z_3^3 i_1}{6.845 \times 10^6 K_t T_1} - \cos^3 \beta \geq 0$$

$$\frac{[\sigma_w]_1 \psi Y_1}{3 K_1 T_1} (1 + i_1) m_{n1}^3 z_1^2 - \cos^2 \beta \geq 0$$

$$\frac{[\sigma_w]_2 \psi Y_2}{3 K_1 T_1} (1 + i_1) m_{n1}^3 z_1^2 - \cos^2 \beta \geq 0$$

$$\frac{[\sigma_w]_3 \psi Y_3}{3 K_1 T_1} (1 + i_1) m_{n1}^3 z_3^2 - \cos^2 \beta \geq 0$$

$$\frac{[\sigma_w]_4 \psi Y_4}{3 K_1 T_1} (1 + i_1) m_{n1}^3 z_3^2 - \cos^2 \beta \geq 0$$

$$m_{n1} z_3 (1 + i_1) - 2 \cos \beta (5 + m_{n1})$$

$$- m_{n1} z_1 i_1 \geq 0$$

$$a_1 \leq z_1 \leq b_1$$

$$a_2 \leq z_3 \leq b_2$$

$$a_3 \leq m_{n1} \leq b_3$$

$$a_4 \leq m_{n1} \leq b_4$$

$$a_5 \leq i_1 \leq b_5$$

$$a_6 \leq \beta \leq b_6$$

或简化写成

$$f(x) = A = \frac{1}{2 \cos \beta} [m_{n1} z_1 (1 + i_1) + m_{n1} z_3 (1 + i_1)] \rightarrow \min$$

$$\text{s. t. } g_j(x) \leq 0 \quad (j = 1, 2, \dots, 7)$$

$$x_{i \min} \leq x_i \leq x_{i \max} \quad (i = 1, 2, \dots, 6)$$

例3 机床传动系统的优化设计。

这里以一个机床主传动系统为例, 说明在优化设计时建立数学模型的方法。

图 17.1-3a、b 所示分别是某车床主传动的传动系统图和相应的传动结构图。图中没有画出摩擦离合器, 因为它的结构尺寸参数是按现有结构选取的, 不需进行优选。由图可知: 它共有四个传动组, 即在 I - II 轴间的由 i_{11} 和 i_{12} 组成的第一传动组, II - III 轴间由 i_{21} 和 i_{22} 组成的第二传动组, III - IV 轴间由 i_{31} 、 i_{32} 和 i_{33} 组成的第三传动组, IV - V 轴间由 i_{41} 和 i_{42} 组成的第四传动组。IV 轴上的齿轮 z_{41} 是公用齿轮。

各传动组的模数依次为 m_1 、 m_2 、 m_3 、 m_4 。

这个传动系统的设计变量有以下三类, 即

1) 各传动组的最低传动比, 分别是 i_{11} 、 i_{21} 、 i_{31} 、 i_{41} 。

2) 各传动组的最小主动轮齿数, 分别是 z_{11} 、 z_{21} 、 z_{31} 、 z_{41} (由于 z_{41} 是公用齿轮, 所以 z_{42} 就不是独立的变量)。

3) 各传动组的模数, 分别是 m_1 、 m_2 、 m_3 、 m_4 。

所以共有 12 个设计变量。

说明一点, 当采用变位齿轮时, 还要考虑变位齿轮传动副和标准齿轮传动副中心距的差值 ΔA_i 。

目标函数取传动路线中各对啮合齿轮中心距之和最小, 它可写成

$$f(x) = \sum_{j=1}^4 \frac{m_j z_{j1}}{2} \left(1 + \frac{1}{i_{j1}} \right) \quad (17.1-6)$$

约束条件包括:

1) 由于结构尺寸引起的齿轮齿数、传动比值、中心距的限制 (可以是上限、下限或上下限)。

2) 齿轮线速度的限制。

3) 齿轮弯曲和接触强度的限制等。

这台车床共有 67 个约束条件, 虽然形式各异, 但都可统一写成不等式约束的形式

$$g_j(x) \leq 0 \quad (j = 1, 2, \dots)$$

$$\text{或} \quad x_{i \min} \leq x_i \leq x_{i \max} \quad (i = 1, 2, \dots)$$

这样, 问题可归结为: 求 $x = [i_{11} \ i_{21} \ i_{31} \ i_{41} \ z_{11} \ z_{21} \ z_{31} \ z_{41} \ m_1 \ m_2 \ m_3 \ m_4]^T$ 的值, 使

$$f(x) = \sum_{j=1}^4 \frac{m_j z_{j1}}{2} \left(1 + \frac{1}{i_{j1}} \right) \rightarrow \min$$

$$\text{s. t. } g_j(x) \leq 0 \quad (j = 1, 2, \dots)$$

⊙ 即受约束于。

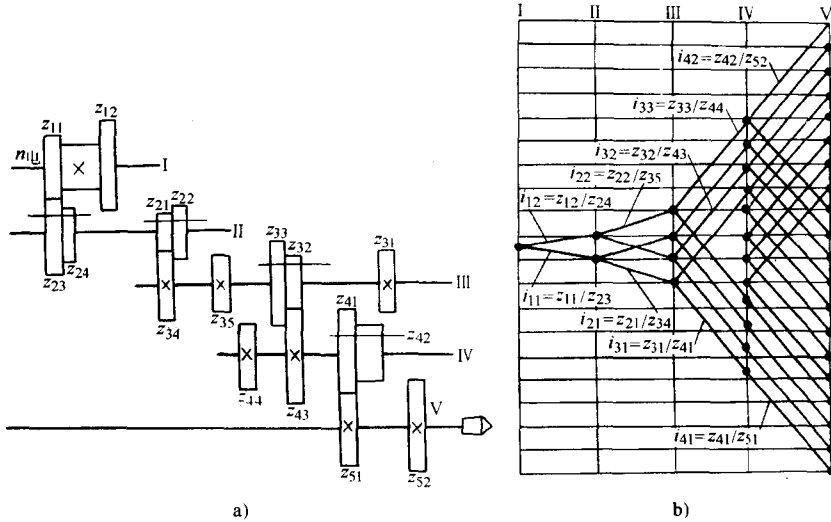


图 17.1-3 车床的主传动系统

a) 传动系统图 b) 传动结构图

$$x_{i\min} \leq x_i \leq x_{i\max} \quad (i=1, 2, \dots)$$

例 4 轴承和轴承系统的优化设计。

对于动压式滑动轴承，当取无量纲形式的表达式时，通过计算可以得出它的

$$\text{承载能力系数} = \frac{F\psi^2}{\eta v L}$$

$$\text{润滑油流量系数} = \frac{q}{\psi v DL}$$

$$\text{轴承的功耗} = \frac{\mu F v}{102}$$

$$\text{轴承的温升} = \frac{\mu F v}{427 c_p \rho q}$$

$$\text{摩擦阻力系数} = \frac{\mu}{\psi}$$

$$\text{圆柱轴承的最小油膜厚度} = \frac{D}{2} \psi \left(1 - \frac{e}{c} \right)$$

轴颈的失稳转速(指开始半速涡动时的轴颈转速)

$$n_w = n_{k1} \sqrt{\frac{\bar{m}}{\gamma^2 k_{eg}}} \text{ 等等。}$$

上述各式中： F 是轴承载荷； D 是轴承直径； L 是轴承长度； v 是轴颈圆周速度； η 是润滑油粘度； c 是半径间隙； e 是轴颈和轴承中心间的偏心距； q 是润滑油流量； μ 是摩擦因数； $\psi = 2c/D$ 是间隙比； c_p 是油的比热容； ρ 是油的密度； $\bar{m} = \omega \psi^3 m / \eta L$ 是转轴分配到轴承上的无量纲质量； m 是转轴分配到轴承上的质量； ω 是转轴的工作角速度； k_{eg} 是当量刚度； γ 是刚度和阻尼的比例系数； n_{k1} 是转轴的第一临界转速。

优化设计时，可以取滑动轴承的最大承载能力、最小功耗、最小流量、最小温升或振动过程中的油膜稳定性等等之中的一个或几个的组合做为目标函数。其约

束条件可以是最低油膜厚度、轴承温升、轴承功耗、轴承转速、轴承的长径比等等。

对一般的轴承系统，可以从动力学角度考虑它的优化设计。

若把轴承系统看做是由支承和轴承处的轴所组成，则在工作时，由轴和支承的质量、轴承系统刚度和阻尼组成一个振动系统。在外力作用下，它会产生沿垂直和水平两个方向的强迫振动。如果忽略垂直和水平方向上的刚度和阻尼的相互影响，则可以对它的两个方向的振动分别进行研究。若只考虑系统在垂直方向上的振动，则它可以简化成图 17.1-4 所示的力学模型。图中 m_1 、 k_1 和 δ_1 是轴的当量质量、轴承刚度系数和阻尼系数； m_2 、 k_2 和 δ_2 是支承的质量、支承座的刚度系数和阻尼系数。这是一个两自由度的振动系统。

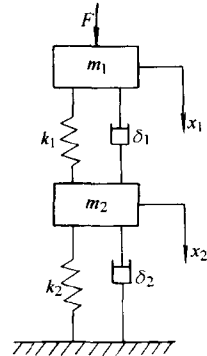


图 17.1-4 轴承系统的力学模型

设计时，可以选择、调整 m_1 、 k_1 、 δ_1 、 m_2 、 k_2 、 δ_2 使系统强迫振动引起的振幅 X_1 和激振力 F 之比 X_1/F ，即动柔度最小(或动刚度最大)。但是必须避免共

振,同时 $m_1, k_1, \delta_1, m_2, k_2, \delta_2$ 等应有一个设计对象所能允许的变化范围。

当忽略阻尼影响时,可以通过系统的两个自由度振动的运动方程

$$M\ddot{x} + Kx = F \quad (17.1-7)$$

解出其动柔度 X_1/F 。式中的 M 和 K 是系统的质量矩阵和刚度矩阵。

避免共振就是要避免激振力频率 ω (例如轴的工作频率 ω) 与系统的固有振动频率 ω_n 重合。工程上按系统固有频率值给出一个频率禁区,使激振力频率不落在频率禁区内。一般要求激振力频率 ω 避开 $(1 \pm 20\%) \omega_n$ 禁区范围。

这样,问题可归结为:确定设计变量 $x = [m_1, m_2, k_1, k_2, \delta_1, \delta_2]^T$ 使目标函数 $\frac{X_1}{F} = f(x)$ 最小,约束条件为

$$\text{若 } \omega_i > \omega, \text{ 则 } \omega_i > 1.2\omega$$

$$\text{若 } \omega_i < \omega, \text{ 则 } \omega_i < 0.8\omega$$

$$x_{i,\min} \leq x_i \leq x_{i,\max}$$

其中 x_i 分别是 $m_1, m_2, k_1, k_2, \delta_1, \delta_2$ 等设计变量。在实际设计中,轴的当量质量一般是给定的。这时,设计变量中不应再包括 m_1 。

例5 机床主轴的优化设计。

图 17.1-5 所示是一个机床主轴的典型结构原理图。对于这类问题,目前也是采用有限元法,利用状态方程来计算轴端变形 y 和固有频率 ω 。

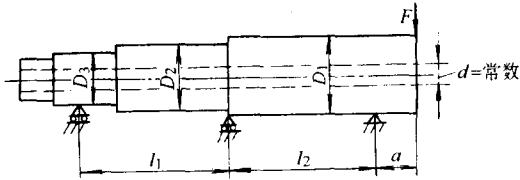


图 17.1-5 机床主轴的典型结构原理图

优化设计的任务是确定 D_i, l_i 和 a , 保证 y 和 ω 在允许限内,使结构的质量最轻。这时,问题归结为:求 D_i, l_i, a 的值,使质量 $f(D_i, l_i) = \rho\pi[\sum(D_i^2 - d^2)l_i + (D_i^2 - d^2)a]$ 为最小,并满足条件

$$y \leq [y]$$

$$\omega^2 \geq \omega_0^2$$

$$D_{i,\min} \leq D_i \leq D_{i,\max} \quad (i=1, 2, \dots, n)$$

$$l_{i,\min} \leq l_i \leq l_{i,\max}$$

$$a_{\min} \leq a \leq a_{\max}$$

$$N_{\min} \leq \frac{l_i}{a} \leq N_{\max}$$

式中 ρ ——材料的密度;

D_i, l_i ——阶梯形主轴的外径和对应的长度;

D_n ——与 a 对应的外径。

在主轴结构动力优化设计时,也可取由振型和质量确定的能耗为目标函数。约束条件可以取激振力频率避开 $(1 \pm 20\%) \omega$ 的禁区范围。

例6 汽车悬挂系统的优化设计。

图 17.1-6 所示是 5 个自由度的汽车悬挂系统。图中的 m_1 是驾驶该车的司机及其座位的质量,它由弹簧 k_1 和阻尼器 δ_1 支持。其它部分,如车体、车轮、车轴等的质量、弹簧和阻尼分别用 m_2, m_4, m_5 和 k_2, k_3 以及 δ_2, δ_3 表示,如图 17.1-6 所示。 k_4, k_5 和 δ_4, δ_5 表示轮胎的刚度和阻尼系数。车体对其质量中心的惯性矩用 I 表示。 L 表示轮距长度。 $f_1(t)$ 和 $f_2(t)$ 表示由于道路表面起伏不平引起的前、后轮的位移函数。 z_i 是坐标。

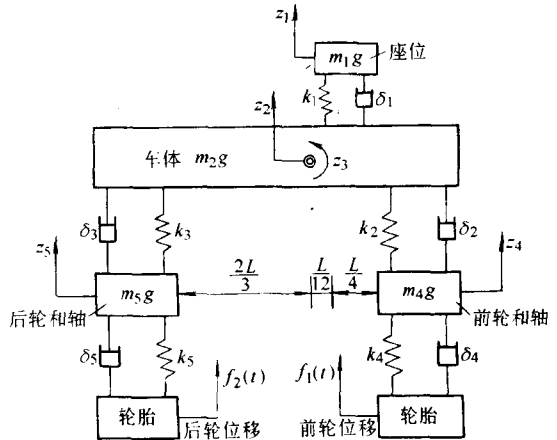


图 17.1-6 汽车悬挂系统

在汽车结构系统设计中,希望汽车能在不同速度和道路条件下,司机座位的最大加速度为最小,同时还须满足一系列的动态响应和设计变量的约束。设计变量是系统的弹簧常数 k_i 和阻尼系数 δ_i 。当然司机座位的最大加速度 d 也可以是设计变量。所以,本优化问题的设计变量取为 $k_1, k_2, k_3, \delta_1, \delta_2, \delta_3$ 和 d , 即

$$x = [k_1, k_2, k_3, \delta_1, \delta_2, \delta_3, d]^T$$

汽车的运动方程可以根据拉格朗日运动方程导出。拉格朗日运动方程的一般形式是

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{z}_i} \right) - \frac{\partial T}{\partial z_i} + \frac{\partial V}{\partial z_i} - F_{Qi} = 0$$

$$(i=1, 2, \dots, 5)$$

系统的动能 T 可表示为 $T = \frac{1}{2} (m \times v^2)$, 即

$$T = \frac{1}{2} m_1 \dot{z}_1^2 + \frac{1}{2} m_2 \dot{z}_2^2 + \frac{1}{2} I \dot{z}_3^2 + \frac{1}{2} m_4 \dot{z}_4^2 + \frac{1}{2} m_5 \dot{z}_5^2$$

$$= \frac{1}{2} \sum_{i=1}^5 m_i \dot{z}_i^2 \quad (m_3 = I)$$

保守力（恢复力）的位置 V 可表示为： $V = \frac{1}{2} (k \times z^2)$ 。

由于车体与司机座位间的相对位移是 $z_2 - z_1 + \frac{L}{12} \times z_3$ ；车体与前、后轮间相对位移分别为 $z_4 - z_2 - \frac{L}{3} z_3$ 和 $z_5 - z_2 + \frac{2L}{3} z_3$ ；前、后轮与路面间相对位移分别为 $z_4 - f_1(t)$ 和 $z_5 - f_2(t)$ 。所以有

$$V = \frac{1}{2} k_1 \left(z_2 - z_1 + \frac{L}{12} z_3 \right)^2 + \frac{1}{2} k_2 \left(z_4 - z_2 - \frac{L}{3} z_3 \right)^2 + \frac{1}{2} k_3 \left(z_5 - z_2 + \frac{2L}{3} z_3 \right)^2 + \frac{1}{2} k_4 (z_4 - f_1(t))^2 + \frac{1}{2} k_5 (z_5 - f_2(t))^2$$

非保守力（阻尼力） F_{Q_i} 可以通过下面方法给出。它的虚功是

$$\sum_{i=1}^5 F_{Q_i} \delta z_i = -c_1 \left(\dot{z}_2 - \dot{z}_1 + \frac{L}{12} \dot{z}_3 \right) \left(\delta z_2 - \delta z_1 + \frac{L}{12} \delta z_3 \right) - c_2 \left(\dot{z}_4 - \dot{z}_2 - \frac{L}{3} \dot{z}_3 \right) \left(\delta z_4 - \delta z_2 - \frac{L}{3} \delta z_3 \right) - c_3 \left(\dot{z}_5 - \dot{z}_2 + \frac{2L}{3} \dot{z}_3 \right) \left(\delta z_5 - \delta z_2 + \frac{2L}{3} \delta z_3 \right) - c_4 (\dot{z}_4 - \dot{f}_1(t)) \delta z_4 - c_5 (\dot{z}_5 - \dot{f}_2(t)) \delta z_5$$

这样，则当 $i=1$ 时，自

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{z}_1} \right) - \frac{\partial T}{\partial z_1} + \frac{\partial V}{\partial z_1} - F_{Q1} = 0$$

有

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{z}_1} \right) = m_1 \ddot{z}_1$$

$$\frac{\partial T}{\partial z_1} = 0$$

$$\frac{\partial V}{\partial z_1} = k_1 z_1 - k_1 z_2 - \frac{L}{12} k_1 z_3$$

$$F_{Q1} = c_1 \dot{z}_1 - c_1 \dot{z}_2 - \frac{L}{12} c_1 \dot{z}_3$$

从而给出

$$M_1 \ddot{z}_1 + c_1 \dot{z}_1 - c_1 \dot{z}_2 - \frac{L}{12} c_1 \dot{z}_3 + k_1 z_1 - k_1 z_2 - \frac{L}{12} k_1 z_3 = 0$$

同样可得 $i=2, 3, 4, 5$ 时的运动方程式。

如果把 $z_1, z_2 \dots z_5, \dot{z}_1, \dot{z}_2 \dots \dot{z}_5$ 都写成向量 $z = [z_1 \ z_2 \ \dots \ z_5 \ \dot{z}_1 \ \dot{z}_2 \ \dots \ \dot{z}_5]^T$ ，则五个运动方程式可统一写成如下的状态方程

$$M \ddot{z}(t) + C \dot{z}(t) + K z(t) = f(t)$$

其中 $f(t)$ 是广义力。

通过变换，也可写成下面的形式

$$\dot{z}(t) = M(x)z(t) + F(t) \quad (17.1-8)$$

式中 $M(x)$ —— 由质量、刚度系数、阻尼系数及 L 和 I 组成的矩阵，而不是单纯的质量矩阵；

$F(t)$ —— 由 $m_1, m_5, k_4, k_5, f_1(t), f_2(t)$ 组成的矩阵；

$$z(t) = [z_1 \ z_2 \ z_3 \ z_4 \ z_5 \ \dot{z}_1 \ \dot{z}_2 \ \dot{z}_3 \ \dot{z}_4 \ \dot{z}_5]^T$$

前、后轮的垂直位移和路面有关，设它们分别为按正弦规律变化的函数 $f_1(t)$ 和 $f_2(t)$ ，其值可表述成

$$f_1(t) = \begin{cases} v(t) & 0 \leq t \leq t_1 \\ 0 & \text{非上述情况} \end{cases} \quad (17.1-9)$$

$$f_2(t) = f_1(t - t_0) \text{ (即比前轮滞后 } t_0 \text{)}$$

式中 t_1 —— 路面不平的停止时间。

根据运动方程和位移函数 $f_1(t), f_2(t)$ ，可以建立数学模型。

设计要求是在路面条件下和车速在一定范围内尽量使司机舒适些。因此，设计的目标是通过调整汽车悬挂特征 k 和 c 等 (m 不便于调整，所以不作为设计变量)，使司机座位的最大绝对加速度 $\max |\ddot{z}_i(t)|$ 达到最小，即 $f = \max |\ddot{z}_i(t)| \rightarrow \min (i = 1, 2, \dots, p)$ ，其中的 $\ddot{z}_i(t)$ 是对第 i 种道路条件 $f_1(t)$ 和 $f_2(t)$ 下的司机座位加速度。当规定最大加速度的上限值为 d (可由设计者选取) 时，则 $|\ddot{z}_i(t)| \leq d_0$ 极端情况下 $|\ddot{z}_i(t)| \leq \theta_0$ (θ_0 最大允许加速度)。

此外，还应考虑到汽车的运动要受到一定的约束，因而对车体和司机座位（也要考虑其它乘车人员的座位）之间的相对位移；车体与前、后轮间的相对位移以及路面和前、后轮间的相对位移等，即汽车的各组成部件之间的相对位移要规定一个允许值。例如，车体与司机座位间的相对位移规定为

$$\left| z_2 - z_1 + \frac{L}{12} z_3 \right| \leq 0$$

等等

若设函数 $\eta(t)$ 是连续的，则上述约束条件 $\eta(t) \leq 0 (0 \leq t \leq \tau)$ 相当于积分约束条件 $\int_0^\tau [\eta(t) + |\eta(t)|] dt = 0$ 。所以，对上述的连续函数 $[z, (t) \dots]$ 形式的约束条件，可以统一写成积分形式

$$\psi_j = \int_0^\tau L_j[t, z(t), x] dt = 0 (j = 1, 2, \dots, p)$$

其中 L_j 是拉格朗日函数。

设计变量的变化范围

$$x_{j\min} \leq x_j \leq x_{j\max}$$

可以写成

$$g_s(x) \leq 0$$

这样，该优化问题的数学模型是

目标函数

$$f = \max |z_1(t)| \rightarrow \min$$

或写成

$$f = d \text{ (} d \text{ 是 } |z_1(t)| \leq d \text{ 中的最小者)}$$

约束条件

$$\dot{z}(t) = M(x)z(t) + F(t) \text{ (状态方程的形式)}$$

$$\phi_j = \int_0^T L_j(t, z(t), x) dt = 0 \text{ (函数约束的形式)}$$

$$g_i(x) \leq 0$$

例 7 单工序加工时, 单件生产率的优化。

在机械加工时, 工艺人员常把单件生产率最大, 或单件加工的工时最短做为一个追求的目标。现在说明此优化问题数学模型的建立方法。

设 t_p 是生产准备时间; t_m 是加工时间; t_c 是刀具更换时间或嵌入一片不重磨刀片所需的时间。若用 T 表示刀具寿命, 则每个工件占用的刀具更换时间为 $t_c \frac{t_m}{T}$ ($\frac{t_m}{T}$ 表示刀具切削刃在其寿命期间内平均可以加工的工件数)。这样, 则单件生产时间是 (min/件)

$$t = t_p + t_m + t_c = t_p + t_m + t_c \frac{t_m}{T}$$

因而单位时间内生产的工件数, 即生产率为

$$q = \frac{1}{t} = \frac{1}{t_p + t_m + t_c \frac{t_m}{T}}$$

刀具寿命 T 和切削速度 v 存在 $vT^n = C$ 的关系, 加工时间和切削速度成反比, 即有 $t_m = \frac{\lambda}{v}$ (λ 是切削加工常数), 则有

$$t = t_p + \frac{\lambda}{v} + \frac{t_c \lambda}{C^{\frac{1}{n}} v^{\frac{1}{n}-1}} \quad (17.1-10)$$

式 (17.1-10) 就是本优化问题的目标函数。

在实际加工中, 典型的约束条件有

进给速度约束条件: $s_{\min} \leq s \leq s_{\max}$

切削速度约束条件: $v_{\min} \leq v \leq v_{\max}$

表面粗糙度约束条件: $\frac{s^2}{8R} \leq R_{a\max}$ (其中的 R 是刀尖半径, $R_{a\max}$ 是允许的表面粗糙度) 或写成: $s \leq \sqrt{8RR_{a\max}} = s_a$ (s_a 是一个常数值)。把它和进给速度约束结合起来, 则有约束

$$s_{\min} \leq s \leq \min(s_{\max}, s_a)$$

功率约束条件: $\frac{F_t h^a s^b v}{4500} \leq P$ (其中的 h 是切削深度, F_t 是切削阻力, P 是电动机功率)。

考虑到约束条件中的变量是 s 和 v , 所以宜把目标函数式 (17.1-10) 中的变量也用 s 和 v 表述。这可以通过用 $t_m = \frac{\lambda_0}{sv}$, $\lambda_0 = \lambda s$ (λ_0 是切削加工常数), $T s^{\frac{1}{m_0}} v^{\frac{1}{n_0}} = C_0$ (其中的 m_0 , n_0 和 C_0 均是常数) 来处理。则得单

件的生产时间为

$$t = t_p + \frac{\lambda_0}{sv} + t_c \frac{\lambda_0}{C_0} s^{\frac{1}{m_0}-1} v^{\frac{1}{n_0}-1}$$

$$= t_p + \frac{\lambda_0}{sv} + t_c \frac{\lambda_0}{C_0} s^m v^n$$

或取下述形式

$$t = t_p + \frac{\lambda_0}{sv} + \lambda_0 a s^m v^n \quad \left(\text{其中的 } a = \frac{t_c}{C_0} \right)$$

可以把它改写成

$$\frac{t}{\lambda_0} = \frac{t_p}{\lambda_0} + \frac{1}{sv} + a s^m v^n$$

由于 $\frac{t_p}{\lambda_0}$ 是常值项, 可以从目标函数中略去, 则本问题的数学模型可以表述为求 s 和 v , 使目标函数 (单件加工时间——每一个工件的加工时间的分钟数值)

$$f(s, v) = \frac{1}{sv} + a s^m v^n \rightarrow \min$$

s. t.

$$v_{\min} \leq v \leq v_{\max}$$

$$s_{\min} \leq s \leq \min(s_{\max}, s_a)$$

$$\frac{F_t h^a s^b v}{4500} \leq P$$

例 8 生产计划的优化示例。

某车间生产甲、乙两种产品。生产甲种产品每件需用材料 9kg、3 个工时, 4kW 电, 可获利 60 元。生产乙种产品每件需用材料 4kg、10 个工时, 5kW 电, 可获利 120 元。若每天能供应材料 360kg, 有 300 个工时, 能供 200kW 电, 问每天生产甲、乙两种产品各多少件, 才能够获得最大的利润。

设每天生产的甲、乙两种产品分别为 x_1 、 x_2 件, 则此问题的数学模型如下

$$f(x_1, x_2) = 60x_1 + 120x_2 \rightarrow \max$$

$$9x_1 + 4x_2 \leq 360 \quad \text{(材料约束)}$$

$$3x_1 + 10x_2 \leq 300 \quad \text{(工时约束)}$$

$$4x_1 + 5x_2 \leq 200 \quad \text{(电力约束)}$$

$$x_1 \geq 0 \quad x_2 \geq 0$$

当然还可以举出一些其它行业的例子。但不不管是哪个专业范围内的问题, 都可以按照如下的方法和步骤来建立相应的优化设计问题的数学模型:

1) 根据设计要求, 应用专业范围内的现行理论和经验等, 对优化对象进行分析。必要时, 需要对传统设计中的公式进行改进, 并尽可能反映该专业范围内的现代技术进步的成果。

2) 对结构诸参数进行分析, 以确定设计的原始参数、设计常数和设计变量 (说明见本章第 3 节)。

3) 根据设计要求, 确定并构造目标函数和相应的约束条件, 有时要构造多目标函数。

4) 必要时对数学模型进行规范化, 以消除诸组成

项间由于量纲不同等原因导致的数量悬殊的影响。

有时不了解结构(或系统)的内部特性,则可建立黑箱(Black box)模型。

2 优化设计问题的数学模型

(1) 设计变量

一个设计方案可以用一组基本参数的数值来表示。这些基本参数可以是构件长度、截面尺寸、某些点的坐标值等几何量,也可以是重量、惯性矩、力或力矩等物理量,还可以是应力、变形、固有频率、效率等代表工作性能的导出量。但是,对某个具体的优化设计问题,并不是要求对所有的的基本参数都用优化方法进行修改调整。例如,对某个机械结构进行优化设计,一些工艺、结构布置等方面的参数,或者某些工作性能的参数,可以根据已有的经验预先取为定值。这样,对这个设计方案来说,它们就成为设计常数。而除此之外的基本参数,则需要在优化设计过程中不断进行修改、调整,一直处于变化的状态,这些基本参数称作设计变量,又叫做优化参数。

设计变量的全体实际上是一组变量,可用一个列向量表示

$$\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$$

称作设计变量向量。向量中分量的次序完全是任意的,可以根据使用的方便任意选取。例如,在例3中的 i_j, z_j, m_j 相当于 $x_1, x_2, x_3, \dots, x_{12}$ 为向量 \mathbf{x} 中的12个分量;例4中的 $m_1, m_2, k_1, \dots, \delta_2$ 相当于 x_1, x_2, \dots, x_6 六个变量;例6中的 $k_1, k_2, k_3, \delta_1, \delta_2, \delta_3$ 和 d 相当于向量 \mathbf{x} 的七个分量等等。这些设计变量可以是一些结构尺寸参数,也可以是一些化学成分的含量或电路参数等。一旦规定了这样一种向量的组成,则其中任意一个特定的向量都可以说是一个“设计”。由 n 个设计变量为坐标所组成的实空间称作设计空间。一个“设计”,可用设计空间中的一点表示,此点可看成是设计变量向量的端点(始点取在坐标原点),称作设计点。

(2) 约束条件

设计空间是所有设计方案的集合,但这些设计方案有些是工程上所不能接受的(例如面积取负值等)。如果一个设计满足所有对它提出的要求,就称为可行(或可接受)设计,反之则称为不可行(或不可接受)设计。

一个可行设计必须满足某些设计限制条件,这些限制条件称作约束条件,简称约束。在工程问题中,根据约束的性质可以把它们区分成性能约束和侧面约束两大类。针对性能要求而提出的限制条件称作性能约束。例如,选择某些结构必须满足受力的强度、刚度或稳定性等要求,桁架某点变形不超过给定值。不是针对

性能要求,只是对设计变量的取值范围加以限制的约束称作侧面约束。例如,允许选择的尺寸范围,桁架的高在其上下限范围之间的要求就属于侧面约束。侧面约束也称作边界约束。

约束又可按其数学表达形式分成等式约束和不等式约束两种类型。等式约束

$$h(\mathbf{x}) = 0$$

要求设计点在 n 维设计空间的约束曲面上。不等式约束

$$g(\mathbf{x}) \leq 0$$

要求设计点在设计空间中约束曲面 $g(\mathbf{x}) = 0$ 的一侧(包括曲面本身)。所以,约束是对设计点在设计空间中的活动范围所加的限制。凡满足所有约束条件的设计点,它在设计空间中的活动范围称作可行域。如满足不等式约束

$$g_j(\mathbf{x}) \leq 0 \quad (j = 1, 2, \dots, m)$$

的设计点活动范围,它是由 m 个约束曲面

$$g_j(\mathbf{x}) = 0 \quad (j = 1, 2, \dots, m)$$

所形成的 n 维子空间(包括边界)。满足两个或更多个 $g_j(\mathbf{x}) = 0$ 点的集合称作交集。在三维空间中两个约束的交集是一条空间曲线,三个约束的交集是一个点。在 n 维空间中 r 个不同约束的交集的维数是 $n-r$ 的子空间。等式约束 $h(\mathbf{x}) = 0$ 可看成是同时满足 $h(\mathbf{x}) \leq 0$ 和 $h(\mathbf{x}) \geq 0$ 两个不等式约束,代表 $h(\mathbf{x}) = 0$ 曲面。

约束函数有的可以表示成显式形式,即反映设计变量之间明显的函数关系,如例1.4中的约束条件,这类约束称作显式约束。有的只能表示成隐式形式,如例5、6中的复杂结构的性能约束函数(变形、应力、频率等),需要通过有限元法或动力学计算求得,机构的运动误差要用数值积分来计算,这类约束称作隐式约束。

(3) 目标函数

在所有的可行设计中,有些设计比另一些要“好些”,如果确实是这样,则“较好”的设计比“较差”的设计必定具备某些更好的性质。倘若这种性质可以表示成设计变量的一个可计算函数,则我们就可以考虑优化这个函数,以得到“更好”的设计。这个用来使设计得以优化的函数称作目标函数。用它可以评价设计方案的好坏,所以它又被称作评价函数,记作 $f(\mathbf{x})$,用以强调它对设计变量的依赖性。目标函数可以是结构重量、体积、功耗、产量、成本或其它性能指标(如变形、应力等)和经济指标等。

建立目标函数是整个优化设计过程中比较重要的问题。当对某一设计性能有特定的要求,而这个要求又很难满足时,则若针对这一性能进行优化将会取得满意的效果。但在某些设计问题中,可能存在两个或两个

以上需要优化的指标, 这将是多目标函数的问题。例如, 设计一台机器, 期望得到最低的造价和最少的维修费用。

目标函数是 n 维变量的函数, 它的函数图像只能在 $n+1$ 维空间中描述出来。为了在 n 维设计空间中反映目标函数的变化情况, 常采用目标函数等值面的方法。目标函数的等值面, 其数学表达式为

$$f(x) = c \quad (17.1-11)$$

(c 为一系列常数), 代表一族 n 维超曲面。如在二维设计空间中 $f(x_1, x_2) = c$, 代表 x_1-x_2 设计平面上的一族曲线。

(4) 优化问题的数学模型

优化问题的数学模型是实际优化设计问题的数学抽象。在明确设计变量、约束条件、目标函数之后, 优化设计问题就可以表示成一般数学形式。

求设计变量向量 $x = [x_1 \ x_2 \ \dots \ x_n]^T$ 使

$$f(x) \rightarrow \min$$

且满足约束条件

$$h_k(x) = 0 \quad (k=1, 2, \dots, l) \quad (17.1-12)$$

$$g_j(x) \leq 0 \quad (j=1, 2, \dots, m)$$

利用可行域概念, 可将数学模型的表达进一步简练。设同时满足 $g_j(x) \leq 0 (j=1, 2, \dots, m)$ 和 $h_k(x) = 0 (k=1, 2, \dots, l)$ 的设计点集合为 R , 即 R 为优化问题的可行域, 则优化问题的数学模型可简练地写成

求 x 使

$$\min_{x \in R} f(x) \quad (17.1-13)$$

符号“ \in ”表示“从属于”。

在实际优化问题中, 对目标函数一般有两种要求形式: 目标函数极小化 $f(x) \rightarrow \min$ 或目标函数极大化 $f(x) \rightarrow \max$ 。由于求 $f(x)$ 的极大化与求 $-f(x)$ 的极小化等价, 所以今后优化问题的数学表达一律采用目标函数极小化形式。

优化问题可以从不同的角度进行分类。例如, 按其有无约束条件分成无约束优化问题和约束优化问题。也可以按约束函数和目标函数是否同时为线性函数, 分成线性规划问题和非线性规划问题。如例 8 的目标函数和约束条件都是线性的, 属于线性规划问题。例 1 和例 3 的目标函数和约束条件都是非线性的, 而例 4 的目标函数是非线性的, 约束条件则是线性的, 这属于非线性规划问题。还可以按问题规模的大小进行分类, 例如, 设计变量和约束条件的个数都在 50 以上的属大型, 10 个以下的属小型, 10~50 属中型。随着电子计算机容量的增大和运算速度的提高, 划分界限将会有所变动。

(5) 优化问题的几何解释

无约束优化问题就是在没有限制的条件下, 对设计变量求目标函数的极小点。在设计空间内, 目标函数是以等值面的形式反映出来的, 则无约束优化问题的极小点即为等值面的中心。

约束优化问题是在可行域内对设计变量求目标函数的极小点, 此极小点在可行域内或在可行域边界上。用图 17.1-7 可以说明有约束的二维优化问题极值点

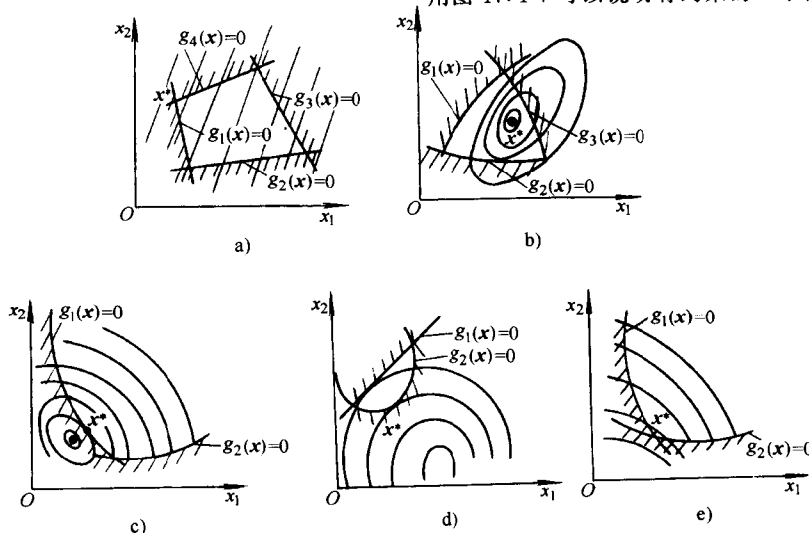


图 17.1-7 极值点所处位置不同的情况

- a) 极值点处于多边形的某一顶点上
- b) 极值点处于等值线的中心
- c) 极值点处于约束曲线与等值线的切点上
- d) 极值点处于约束曲线与等值线的切点上
- e) 极值点处于两个约束曲线的交点上

所处位置的不同情况。图 17.1-7a 是约束函数和目标函数均为线性函数的情况，等值线为直线，可行域为 n 条直线围成的多边形，则极值点处于多边形的某一顶点上。图 17.1-7b 是约束函数和目标函数均为非线性函数的情况，极值点位于可行域内等值线的中心处，约束对极值点的选取无影响，这时的约束为不起作用约束，约束极值点和无约束极值点相同。图 17.1-7c、d 均为约束优化问题极值点处于可行域边界的情况，约束对极值点的位置影响很大。图 17.1-7c 中的约束 $g_1(x)=0$ 在极值点处是起作用约束，图 17.1-7d 中的约束 $g_2(x)=0$ 在极值点处是起作用约束，而图 17.1-7e 中的约束 $g_1(x)=0$ 和 $g_2(x)=0$ 同时在极值点处为起作用约束。多维问题最优解的几何解释可借助于二维问题进行想像。

3 优化问题的极值条件

在“机械优化设计问题示例”中，我们可以看到，机械优化设计问题一般是非线性规划问题，实质上是多元非线性函数的极小化问题。由此可见，机械优化设计是建立在多元函数的极值理论基础上的。无约束优化问题就是数学上的无条件极值问题，而约束优化问题则是数学上的条件极值问题。微分学中所研究的极值问题仅限于等式条件极值，很少涉及优化设计中经常出现的不等式条件极值。为了便于学习以后各章所列举的优化方法，有必要先对极值理论作概略地介绍，重点讨论等式约束优化问题的极值条件和不等式约束优化问题的极值条件。

(1) 多元函数的方向导数与梯度

1) 方向导数 二元函数 $f(x_1, x_2)$ 在 $x_0(x_{10}, x_{20})$ 点处沿某一方向 d 的变化率如图 17.1-8 所示，其定义应为

$$\frac{\partial f}{\partial d} \Big|_{x_0} = \lim_{\Delta d \rightarrow 0} \frac{f(x_{10} + \Delta x_1, x_{20} + \Delta x_2) - f(x_{10}, x_{20})}{\Delta d}$$

称它为该函数沿此方向的方向导数。据此，偏导数

$$\frac{\partial f}{\partial x_1} \Big|_{x_0}, \frac{\partial f}{\partial x_2} \Big|_{x_0}$$

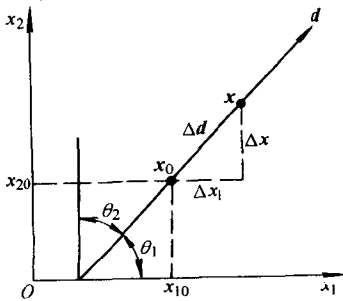


图 17.1-8 二维空间中的方向

坐标轴方向的方向导数。所以方向导数是偏导数概念的推广，偏导数是方向导数的特例。

方向导数与偏导数之间的数量关系是

$$\frac{\partial f}{\partial d} \Big|_{x_0} = \frac{\partial f}{\partial x_1} \Big|_{x_0} \cos\theta_1 + \frac{\partial f}{\partial x_2} \Big|_{x_0} \cos\theta_2$$

依此类推，即可得到 n 元函数 $f(x_1, x_2, \dots, x_n)$ 在 x_0 点处沿 d 方向的方向导数

$$\begin{aligned} \frac{\partial f}{\partial d} \Big|_{x_0} &= \frac{\partial f}{\partial x_1} \Big|_{x_0} \cos\theta_1 + \frac{\partial f}{\partial x_2} \Big|_{x_0} \cos\theta_2 + \dots + \frac{\partial f}{\partial x_n} \Big|_{x_0} \cos\theta_n \\ &= \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{x_0} \cos\theta_i \end{aligned} \quad (17.1-14)$$

其中的 $\cos\theta_i$ 为 d 方向和坐标轴 x_i 方向之间夹角的余弦。

2) 二元函数的梯度 考虑到二元函数具有鲜明的几何解释，并且可以象征性地把这种解释推广到多元函数中去，所以梯度概念的引入也先从二元函数入手。二元函数 $f(x_1, x_2)$ 在 x_0 点处的方向导数 $\frac{\partial f}{\partial d} \Big|_{x_0}$ 的

表达式可改写成下面的形式

$$\begin{aligned} \frac{\partial f}{\partial d} \Big|_{x_0} &= \frac{\partial f}{\partial x_1} \Big|_{x_0} \cos\theta_1 + \frac{\partial f}{\partial x_2} \Big|_{x_0} \cos\theta_2 \\ &= \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix}_{x_0} \begin{bmatrix} \cos\theta_1 \\ \cos\theta_2 \end{bmatrix} \end{aligned}$$

$$\text{令 } \nabla f(x_0) \equiv \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}_{x_0} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix}_{x_0}^T$$

并称它为函数 $f(x_1, x_2)$ 在 x_0 点处的梯度。

$$\text{设 } d \equiv \begin{bmatrix} \cos\theta_1 \\ \cos\theta_2 \end{bmatrix}$$

为 d 方向单位向量，则有

$$\frac{\partial f}{\partial d} \Big|_{x_0} = \nabla f(x_0)^T d = \|\nabla f(x_0)\| \cos(\nabla f, d) \quad (17.1-15)$$

在 x_0 点处函数沿各方向的方向导数是不同的，它随 $\cos(\nabla f, d)$ 变化，即随所取方向的不同而变化。其最大值发生在 $\cos(\nabla f, d)$ 取值为 1 时，也就是当梯度方向和 d 方向重合时其值最大。可见梯度方向是函数值变化最快的方向，而梯度的模就是函数变化率的最大值。

当在 x_1-x_2 平面内画出 $f(x_1, x_2)$ 的等值线

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

(c 为一系列常数) 时，从图 17.1-9 可以看出，在 x_0 处等值线的切线方向 d 是函数变化率为零的方向，即有

$$\frac{\partial f}{\partial d} \Big|_{x_0} = \|\nabla f(x_0)\| \cos(\nabla f, d) = 0$$

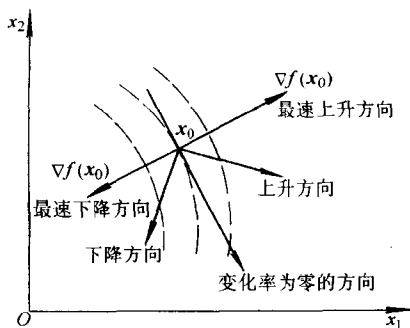


图 17.1-9 梯度方向与等值线的关系

所以

$$\cos(\nabla f, d) = 0$$

可知梯度 $\nabla f(x_0)$ 和切线方向 d 垂直, 从而推得梯度方向为等值面的法线方向。梯度 $\nabla f(x_0)$ 方向为函数变化率最大方向, 也就是最速上升方向。负梯度 $-\nabla f(x_0)$ 方向为函数变化率取最小值方向, 即最速下降方向。与梯度成锐角的方向为函数上升方向, 与负梯度成锐角的方向为函数下降方向。

3) 多元函数的梯度 将二元函数推广到多元函数, 则对于函数 $f(x_1, x_2, \dots, x_n)$ 在 $x_0(x_{10}, x_{20}, \dots, x_{n0})$ 处的梯度 $\nabla f(x_0)$, 可定义为

$$\nabla f(x_0) \equiv \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}_{x_0} = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \dots \quad \frac{\partial f}{\partial x_n} \right]_{x_0}^T$$

对于 $f(x_1, x_2, \dots, x_n)$ 在 x_0 处沿 d 的方向导数可表示为

$$\begin{aligned} \frac{\partial f}{\partial d} \Big|_{x_0} &= \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{x_0} \cos \theta_i = \nabla f(x_0)^T d \\ &= \|\nabla f(x_0)\| \cos(\nabla f, d) \end{aligned} \quad (17.1-16)$$

函数的梯度方向与函数等值面 $f(x) = c$ 相垂直, 也就是和等值面上过 x_0 的一切曲线相垂直, 如图 17.1-10 所示。

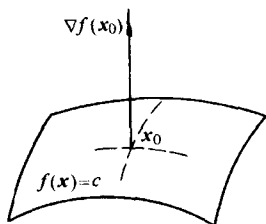


图 17.1-10 梯度方向与等值面的关系

(2) 多元函数的泰勒展开

多元函数的泰勒(Taylor)展开在优化方法中十分重要, 许多方法及其收敛性证明都是从它出发的。

二元函数 $f(x_1, x_2)$ 在 $x_0(x_{10}, x_{20})$ 点处的泰勒展开式为

$$\begin{aligned} f(x_1, x_2) &= f(x_{10}, x_{20}) + \frac{\partial f}{\partial x_1} \Big|_{x_0} \Delta x_1 + \frac{\partial f}{\partial x_2} \Big|_{x_0} \Delta x_2 + \\ &\quad \frac{1}{2} \left(\frac{\partial^2 f}{\partial x_1^2} \Big|_{x_0} \Delta x_1^2 + 2 \frac{\partial^2 f}{\partial x_1 \partial x_2} \Big|_{x_0} \Delta x_1 \Delta x_2 + \right. \\ &\quad \left. \frac{\partial^2 f}{\partial x_2^2} \Big|_{x_0} \Delta x_2^2 \right) + \dots \end{aligned}$$

其中 $\Delta x_1 \equiv x_1 - x_{10}, \Delta x_2 \equiv x_2 - x_{20}$ 。

将上述展开式写成矩阵形式, 有

$$\begin{aligned} f(x) &= f(x_0) + \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix}_{x_0} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} + \\ &\quad \frac{1}{2} [\Delta x_1 \quad \Delta x_2] \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}_{x_0} \times \\ &\quad \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} + \dots \\ f(x) &= f(x_0) + \nabla f(x_0)^T \Delta x + \\ &\quad \frac{1}{2} \Delta x^T G(x_0) \Delta x + \dots \end{aligned} \quad (17.1-17)$$

其中 $G(x_0) \equiv \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}_{x_0}, \Delta x \equiv \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix}$

$G(x_0)$ 称作函数 $f(x_1, x_2)$ 在 x_0 点处的海赛(Hessian)矩阵, 它是由函数 $f(x_1, x_2)$ 在 x_0 处的二阶偏导数所组成的对称方阵。

将二元函数的泰勒展开式推广到多元函数时, 则 $f(x_1, x_2, \dots, x_n)$ 在 x_0 点处泰勒展开式的矩阵形式为

$$\begin{aligned} f(x) &= f(x_0) + \nabla f(x_0)^T \Delta x + \\ &\quad \frac{1}{2} \Delta x^T G(x_0) \Delta x + \dots \end{aligned}$$

其中

$$\nabla f(x_0) = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \dots \quad \frac{\partial f}{\partial x_n} \right]_{x_0}^T$$

为函数 $f(x)$ 在 x_0 点处的梯度

$$G(x_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}_{x_0} \quad (17.1-18)$$

为函数 $f(x)$ 在 x_0 点处的海赛矩阵。

在优化计算中,当某点附近的函数值采用泰勒展开式作近似表达时,研究该点邻域的极值问题需要分析二次型函数是否正定。当对任何非零向量 x 使

$$f(x) = x^T G x > 0$$

即二次型函数正定, G 为正定矩阵。

(3) 无约束优化问题的极值条件

无约束优化问题是使目标函数取得极小值,所谓极值条件就是指目标函数取得极小值时极值点所应满足的条件。

对于二元函数 $f(x_1, x_2)$,若在 $x_0(x_{10}, x_{20})$ 点处取得极值,其必要条件是

$$\left. \frac{\partial f}{\partial x_1} \right|_{x_0} = \left. \frac{\partial f}{\partial x_2} \right|_{x_0} = 0$$

即 $\nabla f(x_0) = \mathbf{0}$ (黑体字“0”代表零向量)

为了判断从上述必要条件求得的 x_0 是否是极值点,需要建立极值的充分条件。根据二元函数 $f(x_1, x_2)$ 在 x_0 点处的泰勒展开式,考虑上述极值必要条件,

$$\text{设 } A = \left. \frac{\partial^2 f}{\partial x_1^2} \right|_{x_0}, B = \left. \frac{\partial^2 f}{\partial x_1 \partial x_2} \right|_{x_0}, C = \left. \frac{\partial^2 f}{\partial x_2^2} \right|_{x_0}$$

则

$$\begin{aligned} f(x_1, x_2) &= f(x_{10}, x_{20}) + \frac{1}{2}(A\Delta x_1^2 + \\ &\quad 2B\Delta x_1\Delta x_2 + C\Delta x_2^2) + \dots \\ &= f(x_{10}, x_{20}) + \frac{1}{2A}[(A\Delta x_1 + B\Delta x_2)^2 + \\ &\quad (AC - B^2)\Delta x_2^2] + \dots \end{aligned}$$

若 $f(x_1, x_2)$ 在 x_0 点处取得极小值,则要求在 x_0 点附近的一切点 x 均须满足

$$f(x_1, x_2) - f(x_{10}, x_{20}) > 0$$

即要求

$$\frac{1}{2A}[(A\Delta x_1 + B\Delta x_2)^2 + (AC - B^2)\Delta x_2^2] > 0$$

或要求 $A > 0, AC - B^2 > 0$

即 $\left. \frac{\partial^2 f}{\partial x_1^2} \right|_{x_0} > 0$

$$\left[\frac{\partial^2 f}{\partial x_1^2} \frac{\partial^2 f}{\partial x_2^2} - \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \right)^2 \right]_{x_0} > 0$$

此条件反映了 $f(x_1, x_2)$ 在 x_0 点处的海赛矩阵 $G(x_0)$ 的各阶主子式均大于零,即

$$|G(x_0)| = \begin{vmatrix} \left. \frac{\partial^2 f}{\partial x_1^2} \right|_{x_0} & & & \\ & \left. \frac{\partial^2 f}{\partial x_1^2} \right|_{x_0} & & \\ & & \left. \frac{\partial^2 f}{\partial x_1 \partial x_2} \right|_{x_0} & \\ & & & \left. \frac{\partial^2 f}{\partial x_2^2} \right|_{x_0} \end{vmatrix} > 0$$

所以,二元函数在某点处取得极值的充分条件是要求

在该点处的海赛矩阵为正定。

对于多元函数 $f(x_1, x_2, \dots, x_n)$,若在 x^* 点处取得极值,则极值的必要条件为

$$\nabla f(x^*) = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \dots \quad \frac{\partial f}{\partial x_n} \right]_{x^*}^T = \mathbf{0} \quad (17.1-19)$$

极值的充分条件为

$$G(x^*) = \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{vmatrix}_{x^*} \quad \text{正定} \quad (17.1-20)$$

即要求 $G(x^*)$ 的下列各阶主子式均大于零

一般说来,多元函数的极值条件在优化方法中仅具有理论意义。因为对于复杂的目标函数,海赛矩阵不易求得,它的正定性就更难判定了。

(4) 等式约束优化问题的极值条件

求解等式约束优化问题:

$$\begin{aligned} \min f(x) \\ \text{s. t. } h_k(x) = 0 \quad (k=1, 2, \dots, m) \end{aligned}$$

需要导出极值存在的条件,这是求解等式约束优化问题的理论基础。对这一问题在数学上有两种处理方法:消元法(降维法)和拉格朗日乘子法(升维法),现分别予以介绍。

1) 消元法 对于 n 维情况

$$\begin{aligned} \min f(x_1, x_2, \dots, x_n) \\ \text{s. t. } h_k(x_1, x_2, \dots, x_n) = 0 \quad (k=1, 2, \dots, l) \end{aligned}$$

由 l 个约束方程将 n 个变量中的前 l 个变量用其余 $n-l$ 个变量表示,即有

$$\begin{aligned} x_1 &= \varphi_1(x_{l+1}, x_{l+2}, \dots, x_n) \\ x_2 &= \varphi_2(x_{l+1}, x_{l+2}, \dots, x_n) \\ &\vdots \\ x_l &= \varphi_l(x_{l+1}, x_{l+2}, \dots, x_n) \end{aligned}$$

将这些函数关系代入到目标函数中,从而得到只含 $x_{l+1}, x_{l+2}, \dots, x_n$ 共 $n-l$ 个变量的函数 $F(x_{l+1}, x_{l+2}, \dots, x_n)$,这样就可以利用无约束优化问题的极值条件求解。

消元法虽然看起来很简单,但实际求解困难却很大。因为将 l 个约束方程联立往往求不出解来。即便能求出解,当把它们代入目标函数之后,也会因函数十分复杂而难于处理。所以这种方法作为一种分析方法实用意义不大,而对某些数值迭代方法来说,却有很大的启发意义。

2) 拉格朗日乘子法 拉格朗日乘子法是求解等式约束优化问题的另一种经典方法,它是通过增加变量将等式约束优化问题变成无约束优化问题。所以又称作升维法。

对于具有 l 个等式约束的 N 维优化问题

$$\min f(x)$$

$$\text{s. t. } h_k(x) = 0 \quad (k=1, 2, \dots, l)$$

在极值点 x^* 处有

$$df(x^*) = \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i = \nabla f(x^*)^T dx = 0$$

$$dh_k(x^*) = \sum_{i=1}^n \frac{\partial h_k}{\partial x_i} dx_i = \nabla h_k(x^*)^T dx = 0$$

$$(k=1, 2, \dots, l)$$

$$\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} + \lambda_1 \frac{\partial h_1}{\partial x_i} + \lambda_2 \frac{\partial h_2}{\partial x_i} + \dots + \lambda_l \frac{\partial h_l}{\partial x_i} \right) dx_i = 0$$

$$(17.1-21)$$

可以通过其中的 l 个方程

$$\frac{\partial f}{\partial x_i} + \lambda_1 \frac{\partial h_1}{\partial x_i} + \lambda_2 \frac{\partial h_2}{\partial x_i} + \dots + \lambda_l \frac{\partial h_l}{\partial x_i} = 0$$

$$(17.1-22)$$

来求解 l 个 $\lambda_1, \lambda_2, \dots, \lambda_l$, 使得 l 个变量的微分 dx_1, dx_2, \dots, dx_l 的系数全为零。这样,式(17.1-21)的等号左边就只剩下 $n-l$ 个变量的微分 $dx_{l+1}, dx_{l+2}, \dots, dx_n$ 的项,即它变成

$$\sum_{j=l+1}^n \left(\frac{\partial f}{\partial x_j} + \lambda_1 \frac{\partial h_1}{\partial x_j} + \lambda_2 \frac{\partial h_2}{\partial x_j} + \dots + \lambda_l \frac{\partial h_l}{\partial x_j} \right) dx_j = 0$$

$$(17.1-23)$$

但 $dx_{l+1}, dx_{l+2}, \dots, dx_n$ 应是任意的量,则应有

$$\frac{\partial f}{\partial x_j} + \lambda_1 \frac{\partial h_1}{\partial x_j} + \lambda_2 \frac{\partial h_2}{\partial x_j} + \dots + \lambda_l \frac{\partial h_l}{\partial x_j} = 0$$

$$(17.1-24)$$

$$(j = l+1, l+2, \dots, n)$$

式(17.1-22)和式(17.1-24)及等式约束 $h_k(x) = 0 (k=1, 2, \dots, l)$ 就是点 x 达到约束极值的必要条件。

式(17.1-22)和式(17.1-24)可以合并写成

$$\frac{\partial f}{\partial x_i} + \lambda_1 \frac{\partial h_1}{\partial x_i} + \lambda_2 \frac{\partial h_2}{\partial x_i} + \dots + \lambda_l \frac{\partial h_l}{\partial x_i} = 0$$

$$(i=1, 2, \dots, n) \quad (17.1-25)$$

把原来的目标函数 $f(x)$ 改造成为如下形式的新目标函数:

$$F(x, \lambda) = f(x) + \sum_{k=1}^l \lambda_k h_k(x) \quad (17.1-26)$$

式中的 $h_k(x)$ 就是原目标函数 $f(x)$ 的等式约束条件,而待定系数 λ_k 称为拉格朗日乘子, $F(x, \lambda)$ 称为拉格朗日函数。这样,拉格朗日乘子法可以叙述如下:

把 $F(x, \lambda)$ 作为一个新的无约束条件的目标函数来求解它的极值点,所得结果就是在满足约束条件

$h_k(x) = 0 (k=1, 2, \dots, l)$ 的原目标函数 $f(x)$ 的极值点。

自 $F(x, \lambda)$ 具有极值的必要条件

$$\frac{\partial F}{\partial x_i} = 0 \quad (i=1, 2, \dots, n)$$

$$\frac{\partial F}{\partial \lambda_k} = h_k(x) = 0 \quad (k=1, 2, \dots, l)$$

可得 $l+n$ 个方程,从而解得 $x = [x_1 \ x_2 \ \dots \ x_n]^T$ 和 $\lambda_k (k=1, 2, \dots, l)$ 共 $l+n$ 个未知变量的值。由上述方程组求得的 $x^* = [x_1^* \ x_2^* \ \dots \ x_n^*]^T$ 是函数 $f(x)$ 极值点的坐标值。

按照式(17.1-25)给出的条件,拉格朗日乘子法也可以用另一种方式表示如下:

$$\nabla F = \nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$$

$$(17.1-27)$$

式中 $\lambda^T = [\lambda_1 \lambda_2 \dots \lambda_l]$

$$\nabla h(x^*)^T = [\nabla h_1(x^*) \ \nabla h_2(x^*) \ \dots \ \nabla h_l(x^*)]$$

例 用拉格朗日乘子法计算在约束条件 $h(x_1, x_2) = 2x_1 + 3x_2 - 6 = 0$ 的情况下,目标函数 $f(x_1, x_2) = 4x_1^2 + 5x_2^2$ 的极值点坐标。

解 改造的目标函数是 $F(x, \lambda) = 4x_1^2 + 5x_2^2 + \lambda(2x_1 + 3x_2 - 6)$, 则由 $\frac{\partial F}{\partial x_1}$ 和 $\frac{\partial F}{\partial x_2}$ 等于零两式解得极值点坐标是

$$x_1 = -\frac{1}{4}\lambda \quad x_2 = -\frac{3}{10}\lambda$$

把它们代入 $\frac{\partial F}{\partial \lambda} = 0$ (即约束条件 $2x_1 + 3x_2 - 6 = 0$) 中去,得 $\lambda = -\frac{30}{7}$ 即极值点 x^* 坐标是 $x_1^* = 1.071, x_2^* = 1.286$ 。

(5) 不等式约束优化问题的极值条件

在工程上大多数优化问题都可表示为具有不等式约束条件的优化问题。因此研究不等式约束极值条件是很有意义的。受到不等式约束的多元函数极值的必要条件是著名的库恩-塔克(Kuhn-Tucker)条件,它是非线性优化问题的重要理论。

1) 库恩-塔克条件 对于多元函数不等式约束优化问题

$$\min f(x)$$

$$\text{s. t. } g_j(x) \leq 0 \quad (j=1, 2, \dots, m)$$

(其中设计变量向量 $x = [x_1 \ x_2 \ \dots \ x_i \ \dots \ x_n]^T$ 为 n 维向量,它受有 m 个不等式约束的限制),同样可以应用拉格朗日乘子法推导出相应的极值条件。为此,需要引入 m 个松弛变量 $\bar{x} = [x_{n+1} \ x_{n+2} \ \dots \ x_{n+m}]^T$, 使不等式约束 $g_j(x) \leq 0 (j=1, 2, \dots, m)$ 变成等式约束 $g_j(x) + x_{n+j}^2 = 0 (j=1, 2, \dots, m)$, 从而组成相应的拉格朗日函数。

$$F(x, \bar{x}, \mu) = f(x) + \sum_{j=1}^m \mu_j (g_j(x) + x_{n+j}^2) \quad (17.1-28)$$

其中 μ 是对应于不等式约束的拉格朗日乘子向量 $\mu = [\mu_1 \mu_2 \dots \mu_j \dots \mu_m]^T$, 并有非负的要求, 即 $\mu \geq 0$ 。

根据无约束极值条件, 可以得到具有不等式约束多元函数极值条件

$$\begin{cases} \frac{\partial f(x^*)}{\partial x_i} + \sum_{j=1}^m \mu_j \frac{\partial g_j(x^*)}{\partial x_i} = 0 & (i = 1, 2, \dots, n) \\ \mu_j g_j(x^*) = 0 & (j = 1, 2, \dots, m) \\ \mu_j \geq 0 & (j = 1, 2, \dots, m) \end{cases} \quad (17.1-29)$$

这就是著名的库恩-塔克(Kuhn-Tucker)条件。

若引入起作用约束的下标集合

$$J(x^*) = \{j | g_j(x^*) = 0, j = 1, 2, \dots, m\}$$

库恩-塔克条件又可写成如下形式

$$\begin{cases} \frac{\partial f(x^*)}{\partial x_i} + \sum_{j \in J} \mu_j \frac{\partial g_j(x^*)}{\partial x_i} = 0 & (i = 1, 2, \dots, n) \\ g_j(x^*) = 0 & (j \in J) \\ \mu_j \geq 0 & (j \in J) \end{cases} \quad (17.1-30)$$

将上式偏微分形式表示为梯度形式, 得

$$\nabla f(x^*) + \sum_{j \in J} \mu_j \nabla g_j(x^*) = 0 \quad (17.1-31)$$

或
$$-\nabla f(x^*) = \sum_{j \in J} \mu_j \nabla g_j(x^*)$$

它表明库恩-塔克条件的几何意义是, 在约束极小值点 x^* 处, 函数 $f(x)$ 的负梯度一定能表示成所有起作用约束在该点梯度(法向量)的非负线性组合。

下面以二维问题为例, 说明其几何意义。

图 17.1-11 是考虑 $g_1(x)$ 和 $g_2(x)$ 两个约束都起作用的情况, 并考虑在点 x^k 处目标函数的负梯度 $-\nabla f(x^k)$ 时的图形。约束函数的梯度 $\nabla g_1(x^k)$ 和 $\nabla g_2(x^k)$, 它们分别垂直于 $g_1(x) = 0$ 和 $g_2(x) = 0$ 二曲

面并形成一维形夹角区域。此时可能出现两种情况。

第一, $-\nabla f(x^k)$ 落在 $\nabla g_1(x^k)$ 和 $\nabla g_2(x^k)$ 所张成的锥角区外的一侧, 如图 17.1-11b 所示。这时, 当过点 x^k 作出与 $-\nabla f(x^k)$ 垂直的切平面, 并从 x^k 出发向此切平面的 $-\nabla f(x^k)$ 所在一侧移动时, 目标函数值可以减小。由于这一侧有一部分区域是可行域(在图中, 这样的区域是由 $f(x) = C$ 和 $g_2(x) = 0$ 形成的), 结果是既可减小目标函数值, 又不破坏约束条件。这说明 x^k 仍可沿约束曲面移动而不致破坏约束条件, 且目标函数值还能够得到改变(减小)。所以点 x^k 不是稳定的最优点, 即它不是约束最优点或局部极值点。

第二, $-\nabla f$ 落在 ∇g_1 和 ∇g_2 张成的锥角之内, 如图 17.1-11a 所示。此时, 作出和 $-\nabla f$ 垂直的过 x^k 的目标函数等值面的切平面, 把空间分成两个区域。当从 x^k 出发向包含 $-\nabla f$ 的一侧移动时, 将可使目标函数值减小。但这一侧的任何一点都不落在可行区域内。显然, 此时的点 x^k 就是约束最优点或局部极值点 x^* 。沿此点再作任何移动都将破坏约束条件, 故它是稳定点。

由于 $-\nabla f(x^*)$ 和 $\nabla g_1(x^*)$, $\nabla g_2(x^*)$ 在一个平面内, 则前者可看成是后两者的线性组合。又因 $-\nabla f(x^*)$ 处于 $\nabla g_1(x^*)$ 和 $\nabla g_2(x^*)$ 的夹角之间, 所以线性组合的系数为正, 即有

$$-\nabla f(x^*) = \mu_1 \nabla g_1(x^*) + \mu_2 \nabla g_2(x^*)$$

其中 $\mu_1 > 0, \mu_2 > 0$ 。

这就是目标函数在两个起作用的约束条件下, 使 x^* 成为条件极值点的必要条件。

当约束条件有三个且同时起作用时, 则要求 $-\nabla f(x^*)$ 处于 $\nabla g_1(x^*)$ 、 $\nabla g_2(x^*)$ 和 $\nabla g_3(x^*)$ 形成的角锥之内。

对于同时具有等式和不等式约束的优化问题:

$$\begin{aligned} & \min f(x) \\ \text{s. t. } & g_j(x) \leq 0 \quad (j = 1, 2, \dots, m) \\ & h_k(x) = 0 \quad (k = 1, 2, \dots, l) \end{aligned}$$

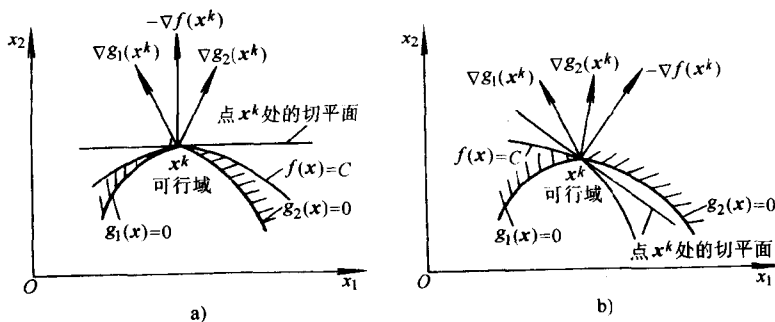


图 17.1-11 库恩-塔克条件的几何意义
a) 负梯度位于锥角区之内 b) 负梯度位于锥角区之外