第1章

机械优化设计概述

1.1 学习导引

本章主要内容包括机械优化设计概述、机械优化设计的数学模型、优化问题的分类和优化方法简介、优化的几何解释、数值迭代法及终止判定准则等。其中对优化设计数学模型的三要素的定义及性质的掌握是重点,它是机械优化设计的一个基础。

1.2 重难点回顾

1.2.1 机械优化设计的数学模型

用一组设计变量描述优化设计对象的设计内容,即描述优化意图(目标、指标)和有关限制条件的数学表达式,称为优化设计的数学模型。

$$
\begin{aligned}
&\min f(\boldsymbol{X}), \quad X \in \mathbf{R}^n \\
\text{s.t.} \quad &g_j(\boldsymbol{X}) \leqslant 0, \quad j = 1,2,\cdots,l \\
&h_k(\boldsymbol{X}) = 0, \quad k = 1,2,\cdots,m
\end{aligned}
$$

数学模型三要素为:设计变量、目标函数、约束条件。

(1) 设计变量

在优化设计过程中,要优化选择的设计参数为设计变量。设计变量必须是独立变量,即在一个优化设计问题中,任意两个设计变量之间没有函数关系。按照产品设计变量的取值特点,设计变量可分为连续变量(例如轴径、轮廓尺寸等)和离散变量(例如各种标准规格等)。

设计变量的个数为优化问题的维数。

设计变量为列向量:

$$\boldsymbol{X} = [x_1, x_2, \cdots, x_n]^{\mathrm{T}}$$

一般来说,在满足设计要求的前提下,应尽量减少设计变量的个数,使设计的难度降低。

(2) 目标函数

目标函数是用来使设计得以优化的函数,或可以评价设计方案好坏的函数。实际上它是反映优化意向的关于设计变量的数学表达式。

$$\min f(\boldsymbol{X}) = f(x_1, x_2, \cdots, x_n)$$

(3) 约束条件

约束条件是优化设计中设计变量必须满足的条件,这些条件是设计变量的函数。根据约束的性质可分为:

① 边界约束,直接限定设计变量的取值范围的约束条件,即

$$a_i \leqslant x_i \leqslant b_i, \quad i = 1, 2, \cdots, n$$

② 性能约束,即由方案的某种性能或设计要求推导出来的约束条件。

在设计空间中,由所有约束条件所限定的区域叫做设计可行域。

1.2.2 优化的几何解释

等值线(等高线)是由许多具有相同目标函数值的设计点所构成的平面曲线,如图 1-1 所示。

目标函数的等值线的数学表达式为

$$f(\boldsymbol{X}) = C \tag{1-1}$$

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(超曲面)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,一般为目标函数等值面(超曲面)与可行域边界(超曲面)的一个切点。

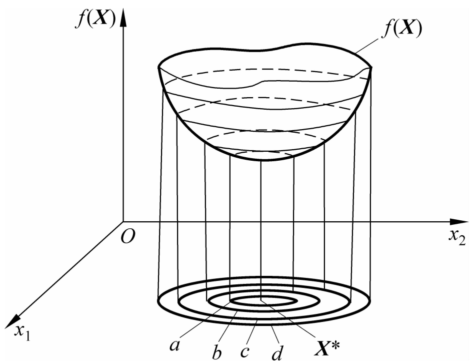

图 1-1 函数的等值线

1.2.3 数值迭代法及终止判定准则

数值迭代法的基本公式为

$$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k + \alpha^k \boldsymbol{S}^k, \quad k = 0, 1, 2, \cdots \tag{1-2}$$

式中,α^k、$\boldsymbol{S}^k$ 分别为第 k 次迭代的步长和搜索方向。

迭代法要解决的问题有:

(1) 选择搜索方向,一般不同算法搜索方向产生的机理不同。

(2) 确定步长因子,在无约束优化设计中一般采用最优步长,而约束优化设计中采用适当步长。

(3) 给定终止判定准则

点距准则:

$$\| \boldsymbol{X}^{k+1} - \boldsymbol{X}^k \| \leqslant \varepsilon$$

函数下降量准则:

$$\| f(\boldsymbol{X}^{k+1}) - f(\boldsymbol{X}^k) \| \leqslant \varepsilon$$

梯度准则:

$$\| \nabla f(\boldsymbol{X}^k) \| \leqslant \varepsilon$$

采用哪种收敛准则,可视具体问题而定。可以取 $\varepsilon=10^{-2}\sim10^{-5}$。

上述准则都在一定程度上反映了逼近最优点的程度,但都有一定的局限性。在实际应用中,可取其中一种或多种同时满足来进行判定。

1.3　典型例题

【例 1-1】 某厂因生产需要,购进五种配件,其个数分别为 x_1、x_2、x_3、x_4、x_5。每种配件的单价分别为 60 元、80 元、85 元、100 元、120 元。要求 x_1 不少于 20 个,x_3 不少于 40 个,其余每种配件不少于 30 个,x_1、x_2 之和不少于 80 个,x_3、x_4 之和不少于 200 个,x_1、x_3、x_4、x_5 之和不少于 400 个。试建立数学模型,问每种配件为多少个,配件总的进价才最低。

解:本例题中影响总进价的因素是每种配件的个数。要求参数 x_1、x_2、x_3、x_4、x_5 在一定的条件下,使所有配件的总进价最低。

设所有配件的总进价为 P,有

$$P=60x_1+80x_2+85x_3+100x_4+120x_5$$

$$\boldsymbol{X}=[x_1,x_2,x_3,x_4,x_5]^{\mathrm{T}}$$

则数学模型为:

目标函数

$$\min f(\boldsymbol{X})=60x_1+80x_2+85x_3+100x_4+120x_5$$

且满足约束条件

$$\begin{aligned}
g_1(\boldsymbol{X})&=x_1+x_2\geqslant 80\\
g_2(\boldsymbol{X})&=x_3+x_4\geqslant 200\\
g_3(\boldsymbol{X})&=x_1+x_3+x_4+x_5\geqslant 400\\
g_4(\boldsymbol{X})&=x_1\geqslant 20\\
g_5(\boldsymbol{X})&=x_2\geqslant 30\\
g_6(\boldsymbol{X})&=x_3\geqslant 40\\
g_7(\boldsymbol{X})&=x_4\geqslant 30\\
g_8(\boldsymbol{X})&=x_5\geqslant 30
\end{aligned}$$

【例 1-2】 试建立数学模型,设计一闭式直齿圆锥齿轮转动。已知:小锥齿轮悬臂支承,大锥齿轮两端支承,轴交角 $\Sigma=90°$,小锥齿轮传递扭矩 $T_1=40\mathrm{N\cdot m}$,转速 $n=960\mathrm{r/min}$,齿数比 $u=3$,精度等级为 7 级,电动机驱动,工作机载电荷稳定,两班制工作,使用期限为 8 年。小锥齿轮选用 40Cr,调质处理,硬度为 241～286HB,大锥齿轮选用 42SiMn,调质处理,硬度为 217～255HB。要求所设计的圆锥齿轮转动体积最小。

解:本例题是要求确定独立设计参数,小齿轮齿数 z、大端模数 m、齿宽系数 Ψ_R 的值。这组值不仅要满足齿轮的强度要求,而且还要使锥齿轮转动体积小(重量轻)。

直齿圆锥齿轮转动的体积可取为小锥齿轮的体积 V_1 和大锥齿轮的体积 V_2 之和,而每个直齿圆锥齿轮的体积又可近似为其大端分度圆与小端分度圆之间的截头圆锥的体积。因此根据截头圆锥的体积计算公式可知,直齿圆锥齿轮转动的体积计算式可写成

$$V=V_1+V_2$$

$$= \frac{\pi}{3} b\cos\delta_1 \left[\left(\frac{mz_1}{2}\right)^2 + \frac{mz_1}{2}\left(\frac{R-b}{R}\cdot\frac{mz_1}{2}\right) + \left(\frac{R-b}{R}\cdot\frac{mz_1}{2}\right)^2 \right]$$
$$+ \frac{\pi}{3} b\cos\delta_2 \left[\left(\frac{mz_2}{2}\right)^2 + \frac{mz_2}{2}\left(\frac{R-b}{R}\cdot\frac{mz_2}{2}\right) + \left(\frac{R-b}{R}\cdot\frac{mz_2}{2}\right)^2 \right]$$

式中，b 为齿宽；δ_1、δ_2 分别为小、大锥齿轮的分锥角；m 为模数；z_1 为小锥齿轮的齿数；R 为锥距；z_2 为大锥齿轮的齿数。

使圆锥齿轮转动体积最小时还应满足如下条件。

(1) 接触强度条件

齿面接触应力必须小于等于许用接触应力

$$\sigma_{\mathrm{H}} = Z_{\mathrm{E}} Z_{\mathrm{H}} Z_{\varepsilon} \sqrt{\frac{4.7KT_1}{\Psi_R(1-0.5\Psi_R)^2 m^3 z_1^3 u}} \leqslant [\sigma_{\mathrm{H}}]$$

式中，Z_{E} 为弹性系数，大、小锥齿轮均为钢制的，故 $Z_{\mathrm{E}}=189.8$；Z_{H} 为节点区域系数，直齿轮 $Z_{\mathrm{H}}=2.5$；Z_{ε} 为重合度系数；K 为载荷系数；$[\sigma_{\mathrm{H}}]$为许用接触应力。

(2) 齿根弯曲强度条件

两齿轮齿根的弯曲应力必须分别小于等于它们各自的许用弯曲应力

$$\sigma_{F1} = \frac{4.7KT_1 Y_{Fa1} Y_{Sa1} Y_{\varepsilon}}{\Psi_R(1-0.5\Psi_R)^2 m^3 z_1^3 \sqrt{u^2+1}} \leqslant [\sigma_{F1}]$$

$$\sigma_{F2} = \sigma_{F1} \frac{Y_{Fa2} Y_{Sa2}}{Y_{Fa1} Y_{Sa1}} \leqslant [\sigma_{F2}]$$

式中，σ_{F1}、σ_{F2}分别为小、大齿轮的齿根弯曲应力；$[\sigma_{F1}]$、$[\sigma_{F2}]$分别为小、大齿轮的许用弯曲应力；Y_{Fa1}、Y_{Sa1}分别为小齿轮的齿形系数和应力修正系数；Y_{Fa2}、Y_{Sa2}分别为大齿轮的齿形系数和应力修正系数；Y_{ε} 为重合度系数。

(3) 最大圆周速度条件

7 级精度的直齿圆锥齿轮，其平均直径处的圆周速度 V_{m} 应满足

$$V_{\mathrm{m}} < 8\mathrm{m/s}$$

(4) 模数限制条件

传递动力的齿轮，其最小模数不得小于 1.5，因此得

$$m_{\max} \geqslant m \geqslant 1.5$$

式中，$m_{\max}$为 m 的最大值。

(5) 齿宽系数限制条件

一般的圆锥齿轮齿宽系数范围是

$$\Psi_R = 0.25 \sim 0.3$$

即

$$\Psi_R \geqslant 0.25$$
$$\Psi_R \leqslant 0.3$$

(6) 小锥齿轮不根切条件

$$z_{1\max} \geqslant z_1 \geqslant 17\cos\delta_1$$

设

$$\boldsymbol{X}=\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}z_1\\m\\\boldsymbol{\Psi}_R\end{bmatrix}$$

则直齿圆锥齿轮传动的优化数学模型为：

使目标函数

$$f(\boldsymbol{X})=V=\frac{\pi}{3}b\cos\delta_1\left[\left(\frac{mz_1}{2}\right)^2+\frac{mz_1}{2}\left(\frac{R-b}{R}\cdot\frac{mz_1}{2}\right)+\left(\frac{R-b}{R}\cdot\frac{mz_1}{2}\right)^2\right]$$
$$+\frac{\pi}{3}b\cos\delta_2\left[\left(\frac{mz_2}{2}\right)^2+\frac{mz_2}{2}\left(\frac{R-b}{R}\cdot\frac{mz_2}{2}\right)+\left(\frac{R-b}{R}\cdot\frac{mz_2}{2}\right)^2\right]$$
$$\to \min$$

且满足约束条件。

$$g_1(\boldsymbol{X})=Z_{\mathrm{E}}Z_{\mathrm{H}}Z_{\varepsilon}\sqrt{\frac{4.7KT_1}{\boldsymbol{\Psi}_R(1-0.5\boldsymbol{\Psi}_R)^2m^3z_1^3u}}-[\sigma_{\mathrm{H}}]\leqslant 0$$
$$g_2(\boldsymbol{X})=\frac{4.7KT_1Y_{Fa1}Y_{Sa1}Y_{\varepsilon}}{\boldsymbol{\Psi}_R(1-0.5\boldsymbol{\Psi}_R)^2m^3z_1^3\sqrt{u^2+1}}-[\sigma_{F1}]\leqslant 0$$
$$g_3(\boldsymbol{X})=\frac{4.7KT_1Y_{Fa2}Y_{Sa2}Y_{\varepsilon}}{\boldsymbol{\Psi}_R(1-0.5\boldsymbol{\Psi}_R)^2m^3z_1^3\sqrt{u^2+1}}-[\sigma_{F2}]\leqslant 0$$
$$g_4(\boldsymbol{X})=V_{\mathrm{m}}-8<0$$
$$g_5(\boldsymbol{X})=1.5-m\leqslant 0$$
$$g_6(\boldsymbol{X})=m-m_{\max}\leqslant 0$$
$$g_7(\boldsymbol{X})=0.25-\boldsymbol{\Psi}_R\leqslant 0$$
$$g_8(\boldsymbol{X})=\boldsymbol{\Psi}_R-0.3\leqslant 0$$
$$g_9(\boldsymbol{X})=17\cos\delta_1-z_1\leqslant 0$$
$$g_{10}(\boldsymbol{X})=z_1-z_{1\max}\leqslant 0$$

【例 1-3】 试建立数学模型，设计一重量最轻的空心转动轴。空心转动轴的轴截面形状如图 1-2 所示，图中 D、d 分别为轴的外径和内径，轴的长度不得小于 3m，轴的材料为 45 钢，密度 $\rho=7.8\times10^{-6}\,\mathrm{kg/mm^3}$，弹性模量 $E=2\times10^5\,\mathrm{MPa}$，许用切应力$[\tau]=60\mathrm{MPa}$，轴所受扭矩 $M=1.5\times10^6\,\mathrm{N\cdot mm}$。

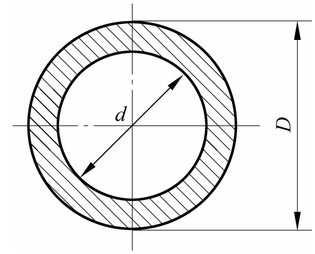

图 1-2　空心转动轴的截面形状

解：空心转动轴的质量 W 的计算式为

$$W=\frac{\pi}{4}(D^2-d^2)l\rho$$

将 ρ 的值代入上式并整理得

$$W=6.12\times10^{-6}\times(D^2-d^2)l$$

所涉及的空心转动轴应满足以下条件：

(1) 扭转强度

(2) 空心转动轴的切应力不得超过许用值，即

$$\tau\leqslant[\tau]$$

空心转动轴的扭转切应力

$$\tau = \frac{16MD}{\pi(D^2 - d^2)}$$

将已知数据代入上式并整理得

$$d^2 - D^2 + 1.27 \times 10^5 D \leqslant 0$$

(3) 扭转稳定性

扭转切应力不得超过扭转稳定的临界切应力 τ'，即

$$\tau \leqslant \tau'$$

空心转动轴的扭转稳定的临界切应力为

$$\tau' = 0.7E\left(\frac{D-d}{2D}\right)^{3/2}$$

将已知数据代入上式并整理得

$$\frac{154.34D}{D^4 - d^4} - \left(\frac{D-d}{D}\right)^{3/2} \leqslant 0$$

(4) 结构尺寸

(5) 空心转动轴的长度不得小于给定值，即

$$l \geqslant l_{\min}$$

式中，$l_{\min}$ 为题目给定的最小长度，为 3m，故

$$3 - l \leqslant 0$$

$$d \geqslant 0$$

$$D - d > 0$$

设

$$\boldsymbol{X} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} D \\ d \\ l \end{bmatrix}$$

则数学模型为：

极小化目标函数

$$\min f(\boldsymbol{X}) = 6.12(D-d)l \times 10^{-6} = 6.12(x_1^2 - x_2^2)x_3 \times 10^{-6}$$

且满足约束条件

$$g_1(\boldsymbol{X}) = d^4 - D^4 + 1.27D \times 10^5 = x_2^4 - x_1^4 + 1.27 \times 10^5 \leqslant 0$$

$$g_2(\boldsymbol{X}) = \frac{154.34D}{D^4 - d^4} - \left(\frac{D-d}{D}\right)^{3/2} = \frac{154.34x_1}{x_1^4 - x_2^4} - \left(\frac{x_1 - x_2}{x_1}\right)^{3/2} \leqslant 0$$

$$g_3(\boldsymbol{X}) = 3 - l = 3 - x_3 \leqslant 0$$

$$g_4(\boldsymbol{X}) = d = x_2 \geqslant 0$$

$$g_5(\boldsymbol{X}) = D - d = x_1 - x_2 > 0$$

【例 1-4】 已知某约束优化问题的数学模型为

$$\min f(\boldsymbol{X}) = (x_1 - 3)^2 + x_2^2$$

$$D: g_1(\boldsymbol{X}) = x_1 \geqslant 0, \quad g_2(\boldsymbol{X}) = x_2 \geqslant 0, \quad g_3(\boldsymbol{X}) = 6 - x_1 \geqslant 0$$

$$g_4(\boldsymbol{X}) = 4 - x_2 \geqslant 0, \quad g_5(\boldsymbol{X}) = x_2 + 1 \geqslant 0, \quad g_6(\boldsymbol{X}) = x_1 - x_2 + 2 \geqslant 0$$

(1) 按比例画出各约束曲线及 $f(\boldsymbol{X})$ 的等值线、可行域，并指出消极约束。

(2) 图中指出无约束最优点和约束最优点。

(3) 若增加 $h(\boldsymbol{X})=x_2-3=0$ 这一等式约束，求最优点。

解：(1) 各约束曲线及 $f(\boldsymbol{X})$的等值线如图 1-3 所示。消极约束为 $g_5(\boldsymbol{X})$。

(2) 无约束最优点和约束最优点均为 $\boldsymbol{X}^*=[3,0]^{\mathrm{T}}$(见图 1-3)。

(3) 若增加 $h(\boldsymbol{X})=x_2-3=0$ 这一等式约束时，$x_2=3$，结合目标函数 $f(\boldsymbol{X})=(x_1-3)^2+x_2^2$，可知，当 $x_1=3$ 时取得最小值 9，代入约束条件可知$[3,3]^{\mathrm{T}}$ 在可行域内，故此时的最优点为$[3,3]^{\mathrm{T}}$。

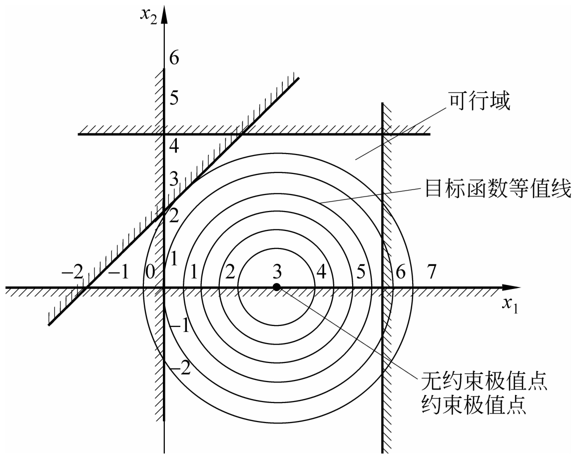

图 1-3　例 1-4 问题的平面图

【例 1-5】　已知目标函数

$$f(\boldsymbol{X})=(x_1-2)^2+x_2^2$$

满足以下约束条件

$$g_1(\boldsymbol{X})=x_1-x_2+2\geqslant 0$$
$$g_2(\boldsymbol{X})=-x_1^2+x_2-1\geqslant 0$$
$$g_3(\boldsymbol{X})=x_1\geqslant 0$$
$$g_4(\boldsymbol{X})=x_2\geqslant 0$$

试画出该数学模型的可行域和目标函数等值线，并在图中指出无约束极值点及约束极值点的位置。

解：在三维空间里，目标函数 $f(\boldsymbol{X})=(x_1-2)^2+x_2^2$ 的几何图形是一个旋转抛物面，如图 1-4 所示。

设

$$f(\boldsymbol{X})=(x_1-2)^2+x_2^2=C_i,\quad i=1,2,3,\cdots$$

且 C_i 为常数，令 $C_1=1/4$，$C_2=1$，$C_3=2$，$C_4=4\cdots$，即可得一系列平面曲线(圆)，它们在 x_1x_2 平面的投影为一系列的同心圆，圆心$[2,0]$即为无约束极值点。约束极值点为约束边界与某一条目标函数等值线的切点，如图 1-5 所示。

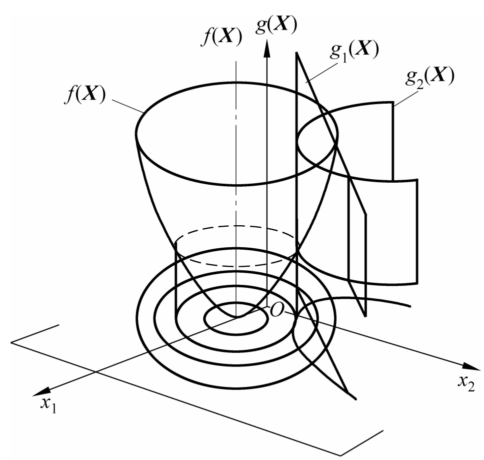

图 1-4　目标函数和约束函数的立体图

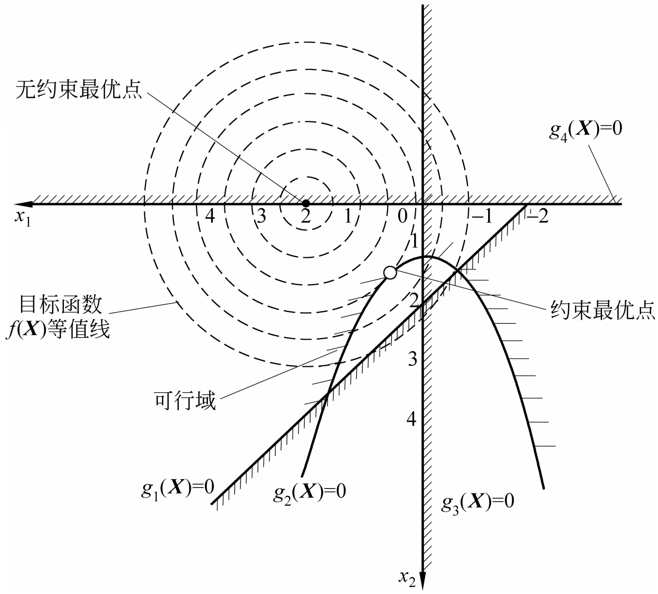

图 1-5　例 1-5 的平面图

【例 1-6】　已知目标函数

$$\min\ f(\boldsymbol{X}) = -x_1 - 4x_2$$

且满足不等式约束条件

$$g_1(\boldsymbol{X}) = x_1 + x_2 - 6 \leqslant 0$$

$$g_2(\boldsymbol{X}) = -x_1 + 2x_2 - 6 \leqslant 0$$

$$g_3(\boldsymbol{X}) = -x_1 \leqslant 0$$

$$g_4(\boldsymbol{X}) = -x_2 \leqslant 0$$

用图解法求其最优点 $\boldsymbol{X}^*$ 和最优值 $f(\boldsymbol{X}^*)$。

解：作出目标函数等值线，并给定目标函数一系列数值。

作出最优点 $\boldsymbol{X}^* = [2,4]^{\mathrm{T}}$，作出最优值 $f(\boldsymbol{X}^*) = -18$，如图 1-6 所示。

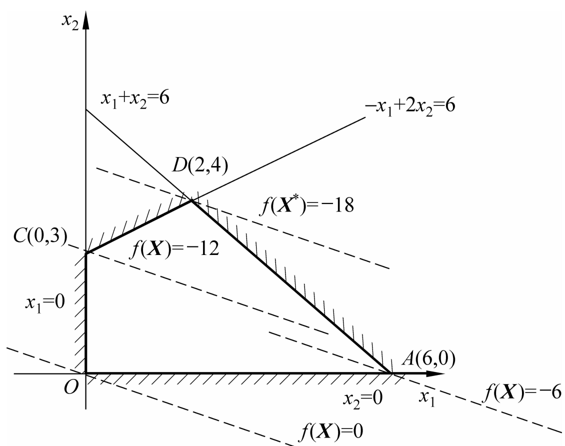

图 1-6　例 1-6 的平面图

1.4 习题

1. 某车间生产甲、乙两种产品。生产甲种产品每件需使用 9kg 材料、3 个工时、4kW·h 电能，可获利润 60 元，生产乙种产品每件需使用 4kg 材料，10 个工时、5kW·h 电能，可获利润 120 元，若使每天能供应 360kg 材料、300 个工时、200kW·h 电能，确定两种产品每天的产量，使每天可能获得的利润最大。试建立该问题的数学模型。

2. 要用薄钢板制造一体积为 5m^3 的货箱，由于运输装卸要求其长度不小于 4m，要求钢板用料最省，试写出这个优化问题的数学模型。

3. 将一批每根 10m 长的钢材截成 3m 和 4m 长一段的毛坯，要求这两种尺寸的毛坯不少于 100 段，试建立用钢材根数最少的数学模型。

4. 将如下优化问题

$$\min f(\boldsymbol{X}) = (x_1 - 3)^2 + (x_2 - 4)^2$$

$$\text{s.t.}\begin{cases} g_1(\boldsymbol{X}) = 5 - x_1 - x_2 \geqslant 0 \\ g_2(\boldsymbol{X}) = x_1 - x_2 - 2.5 \geqslant 0 \\ g_3(\boldsymbol{X}) = x_1 \geqslant 0 \\ g_4(\boldsymbol{X}) = x_2 \geqslant 0 \end{cases}$$

(1) 试按一定的比例画出等目标函数值分别等于 1、4、9、16 时的四条等值线，并在图中画出可行域。

(2) 从图中确定无约束最优解 $\boldsymbol{X}_1$ 和约束最优解 $\boldsymbol{X}_2$。

(3) 该问题属于线性规划问题还是非线性规划问题？

(4) 若在该问题中再加上等式约束 $h(\boldsymbol{X})=x_1-x_2=0$ 后，试确定该问题有无最优解。

1.5 习题参考答案

1. 解：该生产计划问题的优化数学模型可建立为

$$\boldsymbol{X} = [x_1, x_2]^{\mathrm{T}}$$

$$\min f(\boldsymbol{X}) = -60x_1 - 120x_2$$

$$\text{s.t.}\begin{cases} g_1(\boldsymbol{X}) = 9x_1 + 4x_2 - 360 \leqslant 0 \\ g_2(\boldsymbol{X}) = 3x_1 + 10x_2 - 300 \leqslant 0 \\ g_3(\boldsymbol{X}) = 4x_1 + 5x_2 - 200 \leqslant 0 \\ g_4(\boldsymbol{X}) = -x_1 \leqslant 0 \\ g_5(\boldsymbol{X}) = -x_2 \leqslant 0 \end{cases}$$

2. 解：设 x_1、x_2 和 x_3 分别为箱体的长、宽、高，其数学模型为

$$\min f(\boldsymbol{X}) = \min[x_1x_2 + 2(x_2x_3 + x_1x_3)]$$

$$\begin{aligned} \text{s.t.}\quad & 4.0 - x_1 \leqslant 0 \\ & x_1 \geqslant 0 \\ & x_2 \geqslant 0 \end{aligned}$$

$$x_1x_2x_3 = 5.0$$

3. 解：设 x_1、x_2 和 x_3 分别为截法Ⅰ、Ⅱ、Ⅲ所用钢材的根数，其数学模型为

$$\min f(\boldsymbol{X}) = x_1 + x_2 + x_3$$

$$\begin{aligned} \text{s.t.} \quad & 3x_1 + 2x_2 \geqslant 100 \\ & 2x_2 + x_3 \geqslant 100 \\ & x_1 \geqslant 0 \\ & x_2 \geqslant 0 \\ & x_3 \geqslant 0 \end{aligned}$$

4. 解：(1) 目标函数等值线和可行域如图 1-7 所示。

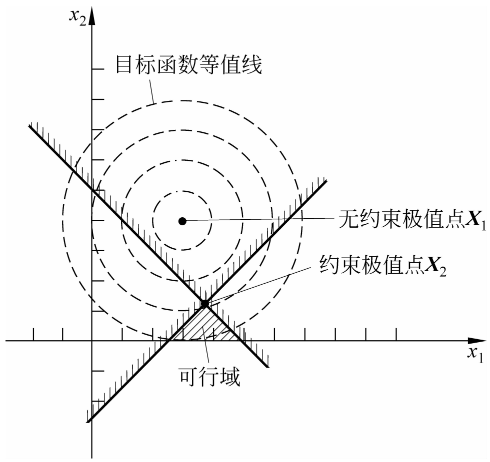

图 1-7　习题 4 的平面图

(2) 无约束最优解：

$$\boldsymbol{X}_1 = [3,4]^{\mathrm{T}}, \quad f_1 = 0$$

约束最优解：

$$\boldsymbol{X}_2 = [3.75,1.25]^{\mathrm{T}}, \quad f_2 = 8.125$$

(3) 该问题属于非线性规划问题。

(4) 加入该约束后，无解。

第2章

优化设计的数学基础

2.1 学习导引

本章主要内容包括方向导数、梯度、海塞矩阵、二次型函数、多元函数的极值、约束函数的极值、凸集、凸函数、凸规划、库恩-塔克条件等。其中库恩-塔克条件的几何意义及判断某点是否为约束极值点是本章的重点。

2.2 重难点回顾

2.2.1 函数的方向导数与梯度

1. 方向导数

偏导数是函数 $f(\boldsymbol{X})$沿平行于坐标轴的各个特殊方向的变化率，对于函数沿任意给定方向的变化率，则需采用方向导数的概念。

多元函数的方向导数可写成

$$\frac{\partial f(\boldsymbol{X}^0)}{\partial \boldsymbol{S}}=\sum_{i=1}^{n}\frac{\partial f(\boldsymbol{X}^0)}{\partial x_i}\cos\alpha_i \tag{2-1}$$

$\alpha_i(i=1,2,\cdots,n)$为某方向 $\boldsymbol{S}$ 的 n 个方向角。

2. 梯度

梯度是函数 $f(\boldsymbol{X})$对各个设计变量的偏导数所组成的列矢量，并以$\nabla f(\boldsymbol{X})$或 $\mathrm{grad} f(\boldsymbol{X})$表示，即梯度是一个矢量，它是函数变化率最大方向上的矢量。

2.2.2 无约束目标函数的极值条件

1. 海塞(Hessian)矩阵

将二元函数的泰勒展开式推广到多元函数时，则 $f(\boldsymbol{X})$在点 $\boldsymbol{X}^k$ 处泰勒展开的矩阵表示为

$$f(\boldsymbol{X})\approx f(\boldsymbol{X}^k)+\nabla^{\mathrm{T}} f(\boldsymbol{X}-\boldsymbol{X}^k)+\frac{1}{2}(\boldsymbol{X}-\boldsymbol{X}^k)^{\mathrm{T}}\nabla^2 f(\boldsymbol{X}-\boldsymbol{X}^k)$$

其中

$$\nabla^{\mathrm{T}} f(\boldsymbol{X})=\left[\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\cdots,\frac{\partial f}{\partial x_n}\right]$$

函数在 $\boldsymbol{X}^k$ 处的二阶偏导数矩阵，称为海塞矩阵，常用 $\boldsymbol{H}(\boldsymbol{X})$表示

$$\boldsymbol{H}(\boldsymbol{X})=\nabla^2 f(\boldsymbol{X})=\begin{bmatrix}\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\partial x_n}\\ \frac{\partial^2 f}{\partial x_2\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\partial x_n}\\ \vdots & \vdots & & \vdots\\ \frac{\partial^2 f}{\partial x_n\partial x_1} & \frac{\partial^2 f}{\partial x_n\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2}\end{bmatrix} \tag{2-2}$$

要判别函数的驻点是极大点还是极小点，可以通过 $\boldsymbol{H}(\boldsymbol{X})$是正定还是负定来决定。若 $\boldsymbol{H}(\boldsymbol{X})$是正定矩阵，对应的是极小值点，反之为极大值点。

2．二次型函数

在线性函数中，将关于变量 $x_1,x_2,\cdots,x_n$ 的二次齐次函数（函数中只含有变量的二次项）称为二次型函数。

$$f(\boldsymbol{X})=[x_1,x_2,\cdots,x_n]\boldsymbol{A}\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X} \tag{2-3}$$

其中

$$\boldsymbol{A}=\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix}$$

$\boldsymbol{A}$ 称为二次型矩阵，它是对称矩阵。对于所有的非零矢量 $\boldsymbol{X}$，若 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}>0$，则称二次型 $f(\boldsymbol{X})=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}$ 是正定的，若 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}<0$，则称二次型 $f(\boldsymbol{X})=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}$ 是负定的。

3．多元函数的极值

对于多元函数，其极值点的必要条件是梯度$\nabla f(\boldsymbol{X})=\boldsymbol{0}$。根据线性代数中有关二次型的讨论可知，极值问题可以由海塞矩阵判定。

当 $\boldsymbol{H}(\boldsymbol{X}^*)$正定，点 $\boldsymbol{X}^*$ 为极小值点；

当 $\boldsymbol{H}(\boldsymbol{X}^*)$负定，点 $\boldsymbol{X}^*$ 为极大值点；

当 $\boldsymbol{H}(\boldsymbol{X}^*)$不定，点 $\boldsymbol{X}^*$ 为鞍点。

判定矩阵 $\boldsymbol{A}$ 是正定还是负定的方法：

如果 $\boldsymbol{A}$ 的各阶顺序主子式都大于零，则矩阵 $\boldsymbol{A}$ 正定；若 $\boldsymbol{A}$ 的各阶顺序主子式正负相间，即所有的偶数阶主子式都为正，而所有的奇数阶主子式都为负，则矩阵 $\boldsymbol{A}$ 负定。

2.2.3　凸集、凸函数和凸规划

1. 凸集

设 D 为 n 维欧氏空间中的一个集合，若其中任意两点 $\boldsymbol{X}^1$、$\boldsymbol{X}^2$ 之间的连接直线都属于 D，则称这种集合 D 为 n 维欧氏空间的一个凸集。图 2-1(a)是二维空间的一个凸集，而图 2-1(b)不是凸集。

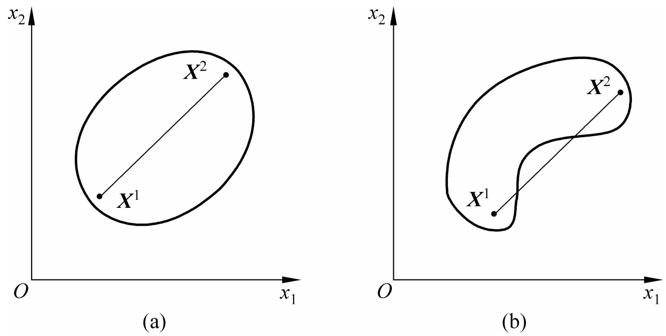

图 2-1　二维空间的凸集与非凸集

(a) 凸集；(b) 非凸集

2. 凸函数

设 $f(\boldsymbol{X})$为定义在 n 维欧氏空间中的一个凸集 D 上的函数，如果对任何实数 $\alpha(0\leqslant\alpha\leqslant 1)$以及对 D 中任意两点 $\boldsymbol{X}^1$、$\boldsymbol{X}^2$ 恒有

$$f(\alpha\boldsymbol{X}^1+(1-\alpha)\boldsymbol{X}^2)\leqslant\alpha f(\boldsymbol{X}^1)+(1-\alpha)f(\boldsymbol{X}^2) \tag{2-4}$$

则称 $f(\boldsymbol{X})$为 D 上的凸函数。

凸函数的几何意义：在凸函数曲线上取任意两点连成一直线线段，则该直线段始终位于函数曲线的上方。

3. 凸规划

对于约束优化问题

$$\begin{aligned}&\min\ f(\boldsymbol{X})\\&\text{s.t.}\quad g_j(\boldsymbol{X})\leqslant 0,\quad j=1,2,\cdots,m\end{aligned}$$

若 $f(\boldsymbol{X}),g_j(\boldsymbol{X}),j=1,2,\cdots,m$ 都为凸函数，则称此问题为凸规划。

凸规划的任何局部最优解就是全局最优解。

2.2.4　有约束目标函数的极值条件

约束函数的极值存在的条件可用库恩-塔克(Kuhn-Tucker)条件去判别。库恩-塔克条件表述为：一个约束极值点存在的必要条件是，目标函数梯度$\nabla f(\boldsymbol{X})$可以表示成诸约束面(不等式和等式约束)梯度的线性组合的负值，即

$$\nabla f(\boldsymbol{X})=-\left\{\sum_{i=1}^{m}\lambda_i\nabla g_i(\boldsymbol{X})+\sum_{j=1}^{n}\lambda_j\nabla h_j(\boldsymbol{X})\right\} \tag{2-5}$$

其中

$$\lambda_i \geqslant 0, \quad \lambda_j \geqslant 0$$

库恩-塔克条件的几何意义在于：起作用约束的梯度矢量，在设计空间的某点 $\boldsymbol{X}^k$ 处构成一个锥体，目标函数的负梯度矢量若包含在此锥体范围以内，则 $\boldsymbol{X}^k$ 为约束极值点，否则不是约束极值点。

需要指出的是库恩-塔克条件是约束极值问题的必要条件，而不是充分条件。只有当目标函数为凸函数，约束函数也全部为凸函数即凸规划的问题时，其局部最优点就是全局最优点。

2.3 典型例题

【例 2-1】 求函数 $f(\boldsymbol{X})=3x_1^2+2x_1x_2+x_2^2-5x_3^3+x_2x_3-1$ 在点 $\boldsymbol{X}=[1,1,1]^{\mathrm{T}}$ 的梯度和梯度的模。

解：因为

$$\frac{\partial f}{\partial x_1}=6x_1+2x_2$$

$$\frac{\partial f}{\partial x_2}=2x_1+2x_2+x_3$$

$$\frac{\partial f}{\partial x_3}=-15x_3^2+x_2$$

所以

$$\nabla f(\boldsymbol{X})=[8,5,-14]^{\mathrm{T}}$$

$$\|\nabla f(\boldsymbol{X})\|=\sqrt{8^2+5^2+(-14)^2}=16.882$$

【例 2-2】 求下列函数在点 $\boldsymbol{X}^1=[1,1]^{\mathrm{T}}$，$\boldsymbol{X}^2=[1,2]^{\mathrm{T}}$，$\boldsymbol{X}^3=[4,1]^{\mathrm{T}}$ 的梯度及其模长。

(1) $f(\boldsymbol{X})=x_1^2+x_2^2-6x_1$

解：由题知，函数的梯度为

$$\nabla f(\boldsymbol{X})=\begin{bmatrix}2x_1-6\\2x_2\end{bmatrix}$$

所以，函数在 3 个点的梯度分别为

$$\nabla f(\boldsymbol{X}^1)=\begin{bmatrix}-4\\2\end{bmatrix}$$

$$\nabla f(\boldsymbol{X}^2)=\begin{bmatrix}-4\\4\end{bmatrix}$$

$$\nabla f(\boldsymbol{X}^3)=\begin{bmatrix}2\\2\end{bmatrix}$$

梯度的模分别为

$$\|\nabla f(\boldsymbol{X}^1)\|=2\sqrt{5}$$

$$\|\nabla f(\boldsymbol{X}^2)\|=4\sqrt{2}$$

$$\|\nabla f(\boldsymbol{X}^3)\|=2\sqrt{2}$$

（2）$f(\boldsymbol{X})=x_1^2+x_2^2-4x_1-3x_2$

解：由题知，函数的梯度为

$$\nabla f(\boldsymbol{X})=\begin{bmatrix}2x_1-4\\2x_2-3\end{bmatrix}$$

所以，函数在 3 个点的梯度分别为

$$\nabla f(\boldsymbol{X}^1)=\begin{bmatrix}-2\\-1\end{bmatrix}$$

$$\nabla f(\boldsymbol{X}^2)=\begin{bmatrix}-2\\1\end{bmatrix}$$

$$\nabla f(\boldsymbol{X}^3)=\begin{bmatrix}4\\-1\end{bmatrix}$$

梯度的模分别为

$$\|\nabla f(\boldsymbol{X}^1)\|=\sqrt{5}$$

$$\|\nabla f(\boldsymbol{X}^2)\|=\sqrt{5}$$

$$\|\nabla f(\boldsymbol{X}^3)\|=\sqrt{17}$$

【例 2-3】 求函数 $f(\boldsymbol{X})=(x_1-x_2+x_3)^2+(-x_1+x_2+x_3)^2+(x_1+x_2+x_3)^2$ 的海塞矩阵，并判断它是否为正定矩阵。

解：（1）$\dfrac{\partial f}{\partial x_1}=2(x_1-x_2+x_3)-2(-x_1+x_2+x_3)+2(x_1+x_2+x_3)=6x_1-2x_2+2x_3$

$$\frac{\partial f}{\partial x_2}=-2(x_1-x_2+x_3)+2(-x_1+x_2+x_3)+2(x_1+x_2+x_3)=-2x_1+6x_2+2x_3$$

$$\frac{\partial f}{\partial x_3}=2(x_1-x_2+x_3)+2(-x_1+x_2+x_3)+2(x_1+x_2+x_3)=2x_1+2x_2+6x_3$$

$$\frac{\partial^2 f}{\partial x_1^2}=\frac{\partial^2 f}{\partial x_2^2}=\frac{\partial^2 f}{\partial x_3^2}=6$$

$$\frac{\partial^2 f}{\partial x_1\partial x_2}=\frac{\partial^2 f}{\partial x_2\partial x_1}=-2$$

$$\frac{\partial^2 f}{\partial x_1\partial x_3}=\frac{\partial^2 f}{\partial x_3\partial x_1}=2$$

$$\frac{\partial^2 f}{\partial x_2\partial x_3}=\frac{\partial^2 f}{\partial x_3\partial x_2}=2$$

所以函数的海塞矩阵为

$$\boldsymbol{H}(\boldsymbol{X})=\begin{bmatrix}\dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\partial x_2} & \dfrac{\partial^2 f}{\partial x_1\partial x_3}\\ \dfrac{\partial^2 f}{\partial x_2\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \dfrac{\partial^2 f}{\partial x_2\partial x_3}\\ \dfrac{\partial^2 f}{\partial x_3\partial x_1} & \dfrac{\partial^2 f}{\partial x_3\partial x_2} & \dfrac{\partial^2 f}{\partial x_3^2}\end{bmatrix}=\begin{bmatrix}6 & -2 & 2\\-2 & 6 & 2\\2 & 2 & 6\end{bmatrix}$$

（2）海塞矩阵的各阶主子式为

$$a_{11}=6>0$$

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} 6 & -2 \\ -2 & 6 \end{vmatrix} = 32 > 0$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} 6 & -2 & 2 \\ -2 & 6 & 2 \\ 2 & 2 & 6 \end{vmatrix} > 0$$

所以其为正定矩阵。

【例 2-4】 判别函数 $f(\boldsymbol{X})=60-x_2+x_1^2+x_2^2+x_1x_2$ 是否为凸函数。

解：用海塞矩阵判别：

$$\frac{\partial f}{\partial x_1} = 2x_1 + x_2, \quad \frac{\partial f}{\partial x_2} = -1 + 2x_2 + x_1$$

$$\frac{\partial^2 f}{\partial x_1^2} = 2, \quad \frac{\partial^2 f}{\partial x_2^2} = 2, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = 1$$

所以

$$\nabla^2 f = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

因为

$$a_{11} = 2 > 0$$

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 4 - 1 > 0$$

所以函数的海塞矩阵正定，故 $f(\boldsymbol{X})$ 为严格的凸函数。

【例 2-5】 判别函数 $f(\boldsymbol{X})=x_1^2+x_2^2-4x_1+4$ 是否为凸函数。

解：函数的海塞矩阵为

$$\boldsymbol{H}(\boldsymbol{X}) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

该海塞矩阵的各阶主子式为

$$a_{11} = 2 > 0$$

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 > 0$$

故海塞矩阵为正定矩阵，所以该函数为严格凸函数。

【例 2-6】 目标函数 $f(\boldsymbol{X})=2x_2+x_2^2+x_1^2-x_1-x_2$，等式约束条件 $h(\boldsymbol{X})=x_1+x_2-2=0$。试求目标函数的机制并判断是极大值还是极小值。

解：该题为等式约束优化问题。首先将 $h(\boldsymbol{X})=0$，改写为 $x_1=2-x_2$，将其代入目标函数 $f(\boldsymbol{X})$，消去 x_1 得

$$f(\boldsymbol{X}) = 2x_2^2 - 2x_2 + 2$$

此为一元函数，求其极值，有

$$\frac{\mathrm{d}f(\boldsymbol{X})}{\mathrm{d}x_2} = 4x_2 - 2 = 0$$

解之得其驻点为

$$x_1 = \frac{3}{2}, \quad x_2 = \frac{1}{2}$$

又因点$\left[\frac{3}{2},\frac{1}{2}\right]^{\mathrm{T}}$处函数的二阶导数为$\frac{\mathrm{d}^2 f(\boldsymbol{X})}{\mathrm{d}x_2^2}=4>0$，所以点$\left[\frac{3}{2},\frac{1}{2}\right]^{\mathrm{T}}$为函数的极小值点，其极小值为$\frac{3}{2}$。

【例 2-7】 求函数 $f(\boldsymbol{X})=x_1^2+x_1x_2+x_2^2-6x_1-3x_2$ 的极值和极值点。

解：求函数 $f(\boldsymbol{X})$的一阶偏导数，并令其为零，即

$$\begin{cases} \dfrac{\partial f}{\partial x_1} = 2x_1 + x_2 - 6 = 0 \\ \dfrac{\partial f}{\partial x_2} = 2x_2 + x_1 - 3 = 0 \end{cases}$$

解该方程组得

$$\boldsymbol{X}^* = [3,0]^{\mathrm{T}}$$

又因

$$\boldsymbol{H}(\boldsymbol{X}^*) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

其顺序主子式为

$$a_{11} = 2 > 0$$

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 3 > 0$$

所以海塞矩阵正定。所以 $\boldsymbol{X}^*=[3,0]^{\mathrm{T}}$ 为函数 $f(\boldsymbol{X})$的极值点，且为极小值点，函数的极小值为 $f(\boldsymbol{X}^*)=-9$。

【例 2-8】 求函数 $f(\boldsymbol{X})=x_1^2+x_2^2-4x_1-2x_2+5$ 的极值。

解：根据极值的必要条件求驻点

$$\begin{cases} \dfrac{\partial f(\boldsymbol{X})}{\partial x_1} = 2x_1 - 4 = 0 \\ \dfrac{\partial f(\boldsymbol{X})}{\partial x_2} = 2x_2 - 2 = 0 \end{cases}$$

得驻点 $\boldsymbol{X}^*=[2,1]^{\mathrm{T}}$。

然后根据极值的充分条件，判断此点是否为极值点。由于

$$\boldsymbol{H}(\boldsymbol{X}^*) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad 2 > 0, \quad \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 > 0$$

其各阶主子式均大于零，$\boldsymbol{H}(\boldsymbol{X}^*)$为正定矩阵，故 $\boldsymbol{X}^*=[2,1]^{\mathrm{T}}$ 为极小值点，极小值为 $f(\boldsymbol{X}^*)=0$。

【例 2-9】 试证明函数 $f(\boldsymbol{X})=x_1^4-2x_1^2x_2+x_1^2+x_2^2-4x_1+5$ 在点$[2,4]^{\mathrm{T}}$ 处具有极小值。

解：首先计算函数 $f(\boldsymbol{X})$的梯度$\nabla f(\boldsymbol{X})$，将点$[2,4]^{\mathrm{T}}$ 代入$\nabla f(\boldsymbol{X})$，若$\nabla f(\boldsymbol{X})=\boldsymbol{0}$，则该点有可能为该函数的极值点（因为$\nabla f(\boldsymbol{X})=\boldsymbol{0}$ 是该点为极值点的必要条件），再计算在该点处

的海塞矩阵。如海塞矩阵正定，则证明该点为极小点，相应的函数值为极小值。

首先函数 $f(\boldsymbol{X})$ 的梯度为

$$\nabla f(\boldsymbol{X})=\begin{bmatrix}\dfrac{\partial f}{\partial x_1}\\ \dfrac{\partial f}{\partial x_2}\end{bmatrix}=\begin{bmatrix}4x_1^3-4x_1x_2+2x_1-4\\ -2x_1^2+2x_2\end{bmatrix}$$

将点 $[2,4]^{\mathrm{T}}$ 代入，可得 $\nabla f(\boldsymbol{X})=\boldsymbol{0}$。该点为 $f(\boldsymbol{X})$ 的驻点。

计算函数 $f(\boldsymbol{X})$ 在点 $[2,4]$ 处的海塞矩阵为

$$\boldsymbol{H}(\boldsymbol{X})=\begin{bmatrix}\dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\partial x_2}\\ \dfrac{\partial^2 f}{\partial x_2\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2}\end{bmatrix}=\begin{bmatrix}12x_1^2-4x_2+2 & -4x_1\\ -4x_1 & 2\end{bmatrix}_{\substack{x_1=2\\ x_2=4}}=\begin{bmatrix}34 & -8\\ -8 & 2\end{bmatrix}$$

其各阶主子式值为

$$a_{11}=34>0$$

$$\begin{vmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{vmatrix}=\begin{vmatrix}34 & -8\\ -8 & 2\end{vmatrix}=4>0$$

可见海塞矩阵为正定矩阵，存在极小值的充分必要条件成立。所以点 $[2,4]^{\mathrm{T}}$ 为函数的极小点，将该点代入 $f(\boldsymbol{X})$，有 $\min f(\boldsymbol{X})=1$。

【例 2-10】 求二元函数 $f(\boldsymbol{X})=x_1^2+x_2^2-2x_1x_2+8x_1+20$ 在点 $\boldsymbol{X}^0=[1,2]^{\mathrm{T}}$ 处沿 $\boldsymbol{S}_1$ 和 $\boldsymbol{S}_2$ 两方向的方向导数。向量 $\boldsymbol{S}_1$ 的方向：$\alpha_1=\alpha_2=\dfrac{\pi}{4}$，向量 $\boldsymbol{S}_2$ 的方向：$\alpha_1=\dfrac{\pi}{3}$，$\alpha_2=\dfrac{\pi}{6}$。并说明函数沿哪个方向变化较快。

解： $\dfrac{\partial f(\boldsymbol{X})}{\partial x_1}=2x_1-2x_2+8$，$\dfrac{\partial f(\boldsymbol{X})}{\partial x_2}=2x_2-2x_1$

梯度 $\nabla f(\boldsymbol{X})=\begin{bmatrix}2x_1-2x_2+8\\ 2x_2-2x_1\end{bmatrix}$

$f(\boldsymbol{X})$ 在 $\boldsymbol{X}^0=[1,2]^{\mathrm{T}}$ 处的梯度值为

$$\nabla f(\boldsymbol{X})=\begin{bmatrix}2x_1-2x_2+8\\ 2x_2-2x_1\end{bmatrix}_{\substack{x_1=1\\ x_2=2}}=\begin{bmatrix}6\\ 2\end{bmatrix}$$

于是 $f(\boldsymbol{X}^0)$ 在 $\boldsymbol{X}^0=[1,2]^{\mathrm{T}}$ 处沿 $\boldsymbol{S}_1$ 和 $\boldsymbol{S}_2$ 两方向的方向导数分别为

$$\begin{aligned}\frac{\partial f(\boldsymbol{X}^0)}{\partial \boldsymbol{S}_1}&=\frac{\partial f(\boldsymbol{X}^0)}{\partial x_1}\cos\alpha_1+\frac{\partial f(\boldsymbol{X}^0)}{\partial x_2}\cos\alpha_2\\ &=6\cos\frac{\pi}{4}+2\cos\frac{\pi}{4}\\ &=5.657\end{aligned}$$

$$\begin{aligned}\frac{\partial f(\boldsymbol{X}^0)}{\partial \boldsymbol{S}_2}&=\frac{\partial f(\boldsymbol{X}^0)}{\partial x_1}\cos\alpha_1+\frac{\partial f(\boldsymbol{X}^0)}{\partial x_2}\cos\alpha_2\\ &=6\cos\frac{\pi}{3}+2\cos\frac{\pi}{6}\\ &=4.732\end{aligned}$$

由此可见，函数 $f(\boldsymbol{X})$ 在同一点沿不同方向的方向导数不同，在点 $\boldsymbol{X}^0=[1,2]^{\mathrm{T}}$ 处沿 $\boldsymbol{S}_1$ 方向的变化率比沿 $\boldsymbol{S}_2$ 方向大，即函数沿 $\boldsymbol{S}_1$ 方向增长较快。

【例 2-11】 判别函数 $f(\boldsymbol{X})=4x_1^2+8x_2^2+x_3^2-2x_2-x_1x_2+30$ 是凸集 D 上的一个凸函数($D=\{X|-\infty<x_i<+\infty,i=1,2,3\}$)。

解：首先求出海塞矩阵 $\boldsymbol{H}(\boldsymbol{X})$：

$$\boldsymbol{H}(\boldsymbol{X})=\begin{bmatrix}\dfrac{\partial^2 f(\boldsymbol{X})}{\partial x_1^2} & \dfrac{\partial^2 f(\boldsymbol{X})}{\partial x_1\partial x_2} & \dfrac{\partial^2 f(\boldsymbol{X})}{\partial x_1\partial x_3}\\ \dfrac{\partial^2 f(\boldsymbol{X})}{\partial x_2\partial x_1} & \dfrac{\partial^2 f(\boldsymbol{X})}{\partial x_2^2} & \dfrac{\partial^2 f(\boldsymbol{X})}{\partial x_2\partial x_3}\\ \dfrac{\partial^2 f(\boldsymbol{X})}{\partial x_3\partial x_1} & \dfrac{\partial^2 f(\boldsymbol{X})}{\partial x_3\partial x_2} & \dfrac{\partial^2 f(\boldsymbol{X})}{\partial x_3^2}\end{bmatrix}=\begin{bmatrix}8 & -1 & 0\\ -1 & 16 & 0\\ 0 & 0 & 2\end{bmatrix}$$

因为 $\boldsymbol{H}(\boldsymbol{X})$的行列式$|\boldsymbol{A}|$的各阶主子式为

$$8>0,\quad \begin{vmatrix}8 & -1\\ -1 & 16\end{vmatrix}=127>0,\quad \begin{vmatrix}8 & -1 & 0\\ -1 & 16 & 0\\ 0 & 0 & 2\end{vmatrix}=254>0$$

所以，矩阵 $\boldsymbol{H}(\boldsymbol{X})$为正定矩阵。因此，函数 $f(\boldsymbol{X})$为凸函数，而且为严格的凸函数。

【例 2-12】 将以下函数在指定点上简化为线性函数和二次函数。

(1) $f(\boldsymbol{X})=x_1(x_1-2)^2+x_2(x_2+1)^2$，$\boldsymbol{X}^1=[1,2]^{\mathrm{T}}$，$\boldsymbol{X}^2=[2,1]^{\mathrm{T}}$

解：由题知

$$f(\boldsymbol{X})=x_1^3-4x_1^2+4x_1+x_2^3+2x_2^2+x_2$$

在 $\boldsymbol{X}^1$ 点

$$f(\boldsymbol{X}^1)=19$$

$$\nabla f(\boldsymbol{X}^1)=\begin{bmatrix}3x_1^2-8x_1+4\\ 3x_2^2+4x_2+1\end{bmatrix}_{\begin{bmatrix}1\\2\end{bmatrix}}=\begin{bmatrix}-1\\ 21\end{bmatrix}$$

$$\nabla^2 f(\boldsymbol{X}^1)=\begin{bmatrix}6x_1-8 & 0\\ 0 & 6x_2+4\end{bmatrix}_{\begin{bmatrix}1\\2\end{bmatrix}}=\begin{bmatrix}-2 & 0\\ 0 & 16\end{bmatrix}$$

$$\boldsymbol{X}-\boldsymbol{X}^1=\begin{bmatrix}x_1-1\\ x_2-2\end{bmatrix}$$

展开的线性函数为

$$\begin{aligned}f(\boldsymbol{X})&=f(\boldsymbol{X}^1)+\nabla f(\boldsymbol{X}^1)^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{X}^1)\\&=19+[-1,21]\begin{bmatrix}x_1-1\\ x_2-2\end{bmatrix}\\&=-x_1+21x_2-22\end{aligned}$$

展开的二次项为

$$\begin{aligned}\frac{1}{2}(\boldsymbol{X}-\boldsymbol{X}^1)^{\mathrm{T}}\nabla^2 f(\boldsymbol{X}^1)(\boldsymbol{X}-\boldsymbol{X}^1)&=\frac{1}{2}[x_1-1,x_2-2]\begin{bmatrix}-2 & 0\\ 0 & 16\end{bmatrix}\begin{bmatrix}x_1-1\\ x_2-2\end{bmatrix}\\&=-(x_1-1)^2+8(x_2-2)^2\end{aligned}$$

展开的二次函数为

$$\begin{aligned}f(\boldsymbol{X})&=-x_1+21x_2-22-(x_1-1)^2+8(x_2-2)^2\\&=-x_1^2+8x_2^2+x_1-11x_2+9\end{aligned}$$

在 $\boldsymbol{X}^2$ 点

$$f(\boldsymbol{X}^2)=4$$

$$\nabla f(\boldsymbol{X}^2)=\begin{bmatrix}3x_1^2-8x_1+4\\3x_2^2+4x_2+1\end{bmatrix}_{\begin{bmatrix}2\\1\end{bmatrix}}=\begin{bmatrix}0\\8\end{bmatrix}$$

$$\nabla^2 f(\boldsymbol{X}^2)=\begin{bmatrix}6x_1-8 & 0\\0 & 6x_2+4\end{bmatrix}_{\begin{bmatrix}2\\1\end{bmatrix}}=\begin{bmatrix}4 & 0\\0 & 10\end{bmatrix}$$

$$\boldsymbol{X}-\boldsymbol{X}^2=\begin{bmatrix}x_1-2\\x_2-1\end{bmatrix}$$

展开的线性函数为

$$\begin{aligned}f(\boldsymbol{X})&=f(\boldsymbol{X}^2)+\nabla f(\boldsymbol{X}^2)^{\mathrm{T}}[\boldsymbol{X}-\boldsymbol{X}^2]\\&=4+[0,8]\begin{bmatrix}x_1-2\\x_2-1\end{bmatrix}=8x_2-4\end{aligned}$$

展开的二次项为

$$\begin{aligned}\frac{1}{2}[\boldsymbol{X}-\boldsymbol{X}^2]^{\mathrm{T}}\nabla^2 f(\boldsymbol{X}^2)[\boldsymbol{X}-\boldsymbol{X}^2]&=\frac{1}{2}[x_1-2,x_2-1]\begin{bmatrix}4 & 0\\0 & 10\end{bmatrix}\begin{bmatrix}x_1-2\\x_2-1\end{bmatrix}\\&=2(x_1-2)^2+5(x_2-1)^2\end{aligned}$$

展开的二次函数为

$$\begin{aligned}f(\boldsymbol{X})&=8x_2-4+2(x_1-2)^2+5(x_2-1)^2\\&=2x_1^2+5x_2^2-8x_1-2x_2+9\end{aligned}$$

(2) $f(\boldsymbol{X})=x_1^3-x_2^3+3x_1^2+3x_2^2-8x_1$，$\boldsymbol{X}^1=[1,2]^{\mathrm{T}}$

解：$f(\boldsymbol{X}^1)=0$

$$\nabla f(\boldsymbol{X}^1)=\begin{bmatrix}3x_1^2+6x_1-8\\-3x_2^2+6x_2\end{bmatrix}_{\begin{bmatrix}1\\2\end{bmatrix}}=\begin{bmatrix}1\\0\end{bmatrix}$$

$$\nabla^2 f(\boldsymbol{X}^1)=\begin{bmatrix}6x_1+6 & 0\\0 & -6x_2+6\end{bmatrix}_{\begin{bmatrix}1\\2\end{bmatrix}}=\begin{bmatrix}12 & 0\\0 & -6\end{bmatrix}$$

$$\boldsymbol{X}-\boldsymbol{X}^1=\begin{bmatrix}x_1-1\\x_2-2\end{bmatrix}$$

展开的线性函数为

$$f(\boldsymbol{X})=f(\boldsymbol{X}^1)+\nabla f(\boldsymbol{X}^1)^{\mathrm{T}}[\boldsymbol{X}-\boldsymbol{X}^1]=0+[1,0]\begin{bmatrix}x_1-1\\x_2-2\end{bmatrix}=x_1-1$$

展开的二次函数为

$$\begin{aligned}f(\boldsymbol{X})&=x_1-1+\frac{1}{2}[\boldsymbol{X}-\boldsymbol{X}^1]^{\mathrm{T}}\nabla^2 f(\boldsymbol{X}^1)[\boldsymbol{X}-\boldsymbol{X}^1]\\&=x_1-1+\frac{1}{2}[x_1-1,x_2-2]\begin{bmatrix}12 & 0\\0 & -6\end{bmatrix}\begin{bmatrix}x_1-1\\x_2-2\end{bmatrix}\\&=6x_1^2-3x_2^2-11x_1+12x_2-7\end{aligned}$$