

第3章

优化问题

学习要点

1. 理解一般优化问题的有关概念.
2. 理解凸优化问题的概念, 掌握其性质.
3. 理解对偶函数的概念, 会求常见优化问题的对偶函数.
4. 理解优化问题的强对偶性及弱对偶性, 掌握相关定理.
5. 掌握一些常用的最优性条件.
6. 理解并能应用几种不同形式的 Karush-Kuhn-Tucker 定理.

3.1 优化问题

优化问题的一般形式为

$$\min f(x) \quad (3.1)$$

$$\text{s.t.} \quad g_i(x) \leq 0, \quad i = 1, 2, \dots, m, \quad (3.2)$$

$$h_j(x) = 0, \quad j = 1, 2, \dots, l. \quad (3.3)$$

其中, $f: \text{Dom}(f) (\subseteq \mathbb{R}^n) \rightarrow \mathbb{R}$ 称为目标函数 (objective function), 每个 g_i 和 h_i 都是定义在 $\mathbb{R}^n$ 中某个子集上的实值函数. 式 (3.2) 和式 (3.3) 是约束 (constraint), 前者为不等式约束, 后者为等式约束. 称

$$\mathcal{D} := \text{Dom}(f) \cap \left(\bigcap_{i=1}^m \text{Dom}(g_i) \right) \cap \left(\bigcap_{j=1}^l \text{Dom}(h_j) \right) \quad (3.4)$$

为优化问题 (3.1)~(3.3) 的定义域, 它是目标函数和所有约束函数的定义域的交集. 称

$$S_F := \{x \in \mathcal{D} : g_i(x) \leq 0, h_j(x) = 0, i = 1, 2, \dots, m, j = 1, 2, \dots, l\} \quad (3.5)$$

为优化问题 (3.1)~(3.3) 的可行集 (feasible set) 或约束集 (constraint set). 称

$$p^* := \inf\{f(x) : x \in S_F\} \quad (3.6)$$

为优化问题 (3.1)~(3.3) 的最优值 (optimal value), 如果 $S_F = \emptyset$, 则约定 $p^* = \infty$. 如果存在 $x^* \in S_F$ 使得 $f(x^*) = p^*$, 则称 x^* 是优化问题 (3.1)~(3.3) 的最优点 (optimal point) 或最优解 (optimal solution). 最优点可能不是唯一的, 把所有最优点所组成的集合称为最优集 (optimal set).

如果优化问题 (3.1)~(3.3) 的可行集非空, 则称它是可行的 (feasible); 如果它的最优集非空, 则称它是可解的 (solvable), 此时它的最优值是可达到的 (attainable).

对于 $\varepsilon > 0$, 如果可行点 $x \in S_F$ 满足 $f(x) < p^* + \varepsilon$, 则称 x 是优化问题 (3.1)~(3.3) 的 ε -次优点 (ε -suboptimal point), 所有 ε -次优点所组成的集合称为 ε -次优集 (ε -suboptimal set).

对于可行点 $x_0 \in S_F$, 如果存在 $\delta > 0$ 使得

$$f(x_0) = \inf \{f(x) : x \in S_F \cap B(x_0, \delta)\}, \quad (3.7)$$

则称 x_0 是优化问题 (3.1)~(3.3) 的局部最优点 (locally optimal point).

在优化问题 (3.1)~(3.3) 中, 对于可行点 x , 如果 $g_i(x) = 0$, 则称不等式约束 $g_i(x) \leq 0$ 是积极的 (active), 否则称它是非积极的 (inactive). 等式约束总是积极的.

注意: 某个约束条件是积极的还是非积极的依赖于可行点 x , 对于不同的可行点 x , 结论可能不一样.

如果删除某个约束条件不会改变可行集, 则称该约束条件是冗余的 (redundant).

我们说两个优化问题是等价的 (equivalent), 即从其中一个问题的解可以导出另一个问题的解.

例如优化问题

$$\max -f(x), \quad \text{s.t. } g_i(x) \leq 0, \quad i = 1, 2, \dots, m \quad (3.8)$$

与优化问题

$$\min f(x), \quad \text{s.t. } g_i(x) \leq 0, \quad i = 1, 2, \dots, m \quad (3.9)$$

是等价的.

又如, 优化问题

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2 \quad (3.10)$$

与优化问题

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 \quad (3.11)$$

也是等价的.

再如, 优化问题

$$\min_{x \in \mathbb{R}^n} e^{f(x)}, \quad \text{s.t. } Ax \leq b \quad (3.12)$$

与优化问题

$$\min_{x \in \mathbb{R}^n} f(x), \quad \text{s.t. } Ax \leq b \quad (3.13)$$

是等价的.

3.2 凸优化问题

3.2.1 凸优化问题的概念

凸优化问题 (convex optimization problem) 的一般形式为

$$\min f(x) \quad (3.14)$$

$$\text{s.t. } g_i(x) \leq 0, \quad i = 1, 2, \dots, m, \quad (3.15)$$

$$a_j^T x = b_j, \quad j = 1, 2, \dots, l. \quad (3.16)$$

其中, $f, g_i, i = 1, 2, \dots, m$ 是凸函数. 凸优化问题的可行集显然是凸集, 凸优化问题的特点可概括为可行集是凸集, 目标函数是凸函数.

如果 f 是凹函数, 则优化问题

$$\max f(x) \quad (3.17)$$

$$\text{s.t. } g_i(x) \leq 0, \quad i = 1, 2, \dots, m, \quad (3.18)$$

$$a_j^T x = b_j, \quad j = 1, 2, \dots, l \quad (3.19)$$

与凸优化问题

$$\min -f(x) \quad (3.20)$$

$$\text{s.t. } g_i(x) \leq 0, \quad i = 1, 2, \dots, m, \quad (3.21)$$

$$a_j^T x = b_j, \quad j = 1, 2, \dots, l \quad (3.22)$$

等价, 因此我们也称 (3.17)~(3.19) 是凸优化问题.

在优化问题 (3.14)~(3.16) 中, 如果 f 是拟凸函数, 则称之为拟凸优化问题 (quasiconvex problem), 凸优化问题显然也是拟凸优化问题.

3.2.2 凸优化问题的性质

命题 3.1 拟凸优化问题的 ε -次优集一定是凸集, 特别地, 其最优集一定是凸集.

证明 拟凸优化问题的 ε -次优集可表示为

$$X_\varepsilon = \{x \in \mathbb{R}^n : f(x) \leq p^* + \varepsilon\} \cap S_F, \quad (3.23)$$

其中, p^* 是该优化问题的最优值, S_F 是可行集. 由于 f 是拟凸函数, 因此下水平集 $\{x \in \mathbb{R}^n : f(x) \leq p^* + \varepsilon\}$ 是凸集, 又因为 S_F 也是凸集, 因此 X_ε 是凸集. $\square$

对于凸优化问题 (3.14)~(3.16), 根据定理 2.9, 有下列结论.

定理 3.1 凸优化问题 (3.14)~(3.16) 的局部最优点一定是全局最优点; 如果 f 是严格凸的, 则凸优化问题的最优点 (如果存在) 是唯一的.

接下来讨论凸优化问题 (3.14)~(3.16) 的最优性条件. 假设 f 是可微的凸函数, 设 S_F 是该优化问题的可行集, 根据定理 2.2, 对任意 $x, y \in S_F$ 皆有

$$f(y) \geq f(x) + (y - x)^T \nabla f(x). \quad (3.24)$$

如果在 x 点满足

$$(y - x)^T \nabla f(x) \geq 0, \quad \forall y \in S_F, \quad (3.25)$$

则有

$$f(x) \leq f(y) - (y - x)^T \nabla f(x) \leq f(y), \quad \forall y \in S_F, \quad (3.26)$$

因此 x 是最优点.

反之, 如果 $x \in S_F$ 是最优点, 则有

$$f(x) \leq f((1 - t)x + ty), \quad \forall y \in S_F, 0 \leq t \leq 1. \quad (3.27)$$

令 $\varphi(t) = f((1 - t)x + ty)$, $0 \leq t \leq 1$, 则 φ 是可微的, 且 $\varphi(t) \geq \varphi(0)$, $\forall 0 \leq t \leq 1$, 因此有

$$\varphi'(0) = \lim_{t \rightarrow 0^+} \frac{\varphi(t) - \varphi(0)}{t} \geq 0, \quad (3.28)$$

由此立刻得到

$$(y - x)^T \nabla f(x) = \varphi'(0) \geq 0. \quad (3.29)$$

综上所述, 我们证明了如下定理.

定理 3.2 设 f 是可微的凸函数, S_F 是凸优化问题 (3.14)~(3.16) 的可行集. 则 $x \in S_F$ 是最优点当且仅当条件 (3.25) 成立.

推论 3.1 设 $C \subseteq \mathbb{R}^n$ 是开凸集, f 是定义在 C 上的可微凸函数, 考虑优化问题

$$\min_{x \in C} f(x), \quad (3.30)$$

$x \in C$ 是它的最优点的充分必要条件是 $\nabla f(x) = 0$.

证明 先证必要性. 设 $x \in C$ 是最优点, 由于 C 是开集, 因此存在 $r > 0$ 使得 $\bar{B}(x, r) \subseteq C$, 于是

$$f(x + tu) \geq f(x), \quad \forall u \in \mathbb{R}^n, \|u\|_2 = 1, -r \leq t \leq r. \quad (3.31)$$

令 $\varphi(t) = f(x + tu)$, $-r \leq t \leq r$, 则 φ 是可微的, 且 φ 在 $t = 0$ 处取得最小值, 因此有 $\varphi'(0) = 0$, 由此得到

$$u^T \nabla f(x) = \varphi'(0) = 0, \quad \forall u \in \mathbb{R}^n, \|u\|_2 = 1, \quad (3.32)$$

因此必有 $\nabla f(x) = 0$, 否则取 $u = \nabla f(x)/\|\nabla f(x)\|_2$, 则有 $\|u\|_2 = 1$ 且 $u^T \nabla f(x) = \|\nabla f(x)\|_2 \neq 0$, 矛盾.

再证充分性. 由于 f 是 C 上的凸函数, 根据定理 2.9 得

$$f(y) \geq f(x) + (y - x)^T \nabla f(x), \quad \forall y \in C, \quad (3.33)$$

如果 $\nabla f(x) = 0$, 则有 $f(x) \leq f(y), \forall y \in C$, 因此 x 是最优点. $\square$

例 3.1 设

$$f(x) = \frac{1}{2} x^T P x + q^T x + r, \quad x \in \mathbb{R}^n, \quad (3.34)$$

其中 $P \succeq 0$. 考察优化问题

$$\min_{x \in \mathbb{R}^n} f(x), \quad (3.35)$$

由于 $\nabla^2 f(x) = P, \forall x \in \mathbb{R}^n$, 因此 f 是凸函数. 根据推论 3.1, $x \in \mathbb{R}^n$ 是最优点当且仅当 $\nabla f(x) = 0$, 即 $Px = -q$. 当 $P \succ 0$ 时, 方程 $Px = -q$ 有唯一解, 此时优化问题有唯一最优点; 当 P 是奇异矩阵且 $q \in \mathcal{R}(P)$ 时, 其中 $\mathcal{R}(P)$ 表示 P 的像空间 (image space), 方程 $Px = -q$ 有无穷多个解, 此时优化问题有无穷多个最优点; 当 P 是奇异矩阵且 $q \notin \mathcal{R}(P)$ 时, 方程 $Px = -q$ 无解, 此时优化问题没有最优点.

接下来考虑只带有等式约束的凸优化问题

$$\min f(x), \quad \text{s.t.} \quad Ax = b, \quad (3.36)$$

其中, $A = (a_{ij})_{m \times n}, b \in \mathbb{R}^m$. 该优化问题的可行集是方程 $Ax = b$ 的解集, 是一个仿射集. 根据定理 3.2, 可行点 x 是最优点当且仅当

$$(y - x)^T \nabla f(x) \geq 0, \quad \forall y \in \{y \in \mathbb{R}^n : Ay = b\}, \quad (3.37)$$

由于仿射集 $\{y \in \mathbb{R}^n : Ay = b\}$ 中的元素都可表示为

$$y = x + v, \quad v \in \mathcal{N}(A), \quad (3.38)$$

其中, $\mathcal{N}(A)$ 表示 A 的零空间 (null space), 因此式 (3.37) 等价于

$$v^T \nabla f(x) \geq 0, \quad \forall v \in \mathcal{N}(A), \quad (3.39)$$

由于 $\mathcal{N}(A)$ 是一个线性空间, 因此上述条件等价于

$$v^T \nabla f(x) = 0, \quad \forall v \in \mathcal{N}(A), \quad (3.40)$$

即 $\nabla f(x) \in \mathcal{N}(A)^\perp$. 注意到 $\mathcal{N}(A)^\perp = \mathcal{R}(A^T)$, 因此式 (3.40) 又等价于存在 $\nu \in \mathbb{R}^m$ 使得

$$\nabla f(x) + A^T \nu = 0, \quad (3.41)$$

它与可行性条件 $Ax = b$ 一起构成了 Lagrange 乘数法的最优性条件.

综上所述, 我们证明了如下定理.

定理 3.3 $x \in \mathbb{R}^n$ 是凸优化问题 (3.36) 的最优点的充分必要条件是存在 $\nu \in \mathbb{R}^m$, 使得它满足下列 Lagrange 乘数法最优性条件:

$$Ax = b, \quad \nabla f(x) + A^T \nu = 0. \quad (3.42)$$

再来考虑下列凸优化问题:

$$\min f(x), \quad \text{s.t.} \quad x \succeq 0, \quad (3.43)$$

其中, f 是凸函数. 此时最优性条件为

$$x \succeq 0, \quad (y - x)^T \nabla f(x) \geq 0, \quad \forall y \succeq 0. \quad (3.44)$$

从以上条件不难推出 $\nabla f(x) \succeq 0$, 否则 $(y - x)^T \nabla f(x)$ 作为 y 的函数在非负象限上无下界, 违背了条件 $(y - x)^T \nabla f(x) \geq 0$. 由第二个条件得到

$$y^T \nabla f(x) \geq x^T \nabla f(x), \quad \forall y \succeq 0, \quad (3.45)$$

取 $y = 0$ 便得到 $x^T \nabla f(x) \leq 0$, 又因为 $x \succeq 0, \nabla f(x) \succeq 0$, 因此 $x^T \nabla f(x) \geq 0$, 从而只能是 $x^T \nabla f(x) = 0$. 再注意到

$$x^T \nabla f(x) = \sum_{i=1}^n x_i (\nabla f(x))_i, \quad (3.46)$$

其中, $(\nabla f(x))_i$ 表示梯度向量 $\nabla f(x)$ 的第 i 个分量. 上式中每一项都是非负的, 因此 $x^T \nabla f(x) = 0$ 当且仅当

$$x_i (\nabla f(x))_i = 0, \quad i = 1, 2, \dots, n. \quad (3.47)$$

这个条件称为互补性 (complementarity).

综上所述, 我们证明了下列结果.

定理 3.4 $x \in \mathbb{R}^n$ 是凸优化问题 (3.43) 的最优点的充分必要条件是

$$x \succeq 0, \quad \nabla f(x) \succeq 0, \quad x_i (\nabla f(x))_i = 0, \quad i = 1, 2, \dots, n. \quad (3.48)$$

3.3 Lagrange 对偶函数

3.3.1 Lagrange 对偶函数的定义

现在回到一般的优化问题 (3.1)~(3.3), 假设该问题的定义域 $\mathcal{D}$ 非空, 并设 p^* 是该问题的最优值.

定义 $L: \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^l \rightarrow \mathbb{R}$ 为

$$L(x, \lambda, \nu) := f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^l \nu_j h_j(x), \quad (3.49)$$

称之为优化问题 (3.1)~(3.3) 的 Lagrange 函数, 其中, λ 和 ν 称为 **Lagrange 乘数 (Lagrange multiplier)** 或对偶变量 (**dual variable**).

定义 3.1 设 $L(x, \lambda, \nu)$ 是优化问题 (3.1)~(3.3) 的 Lagrange 函数, 由式 (3.49) 定义, 令

$$L_D(\lambda, \nu) := \inf_{x \in \mathcal{D}} L(x, \lambda, \nu), \quad \lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^l, \quad (3.50)$$

称之为优化问题 (3.1)~(3.3) 的 **Lagrange 对偶函数 (Lagrange dual function)** 或简称为**对偶函数 (dual function)**.

对于给定的 $x \in \mathcal{D}$, $L(x, \lambda, \nu)$ 是 λ 和 ν 的线性函数, 因此对偶函数 L_D 是一族线性函数的逐点下确界, 从而是凸函数, 无论原优化问题是不是凸优化问题.

例 3.2 考虑优化问题

$$\min x^T x, \quad \text{s.t.} \quad Ax = b, \quad (3.51)$$

其中, A 是 $l \times n$ 的实矩阵, $l \leq n$. 此优化问题的 Lagrange 函数为

$$L(x, \nu) = x^T x + \sum_{j=1}^l \nu_j (x^T a_j - b_j) = x^T x + \nu^T (Ax - b), \quad (3.52)$$

其中, a_j 表示 A 的第 j 个行向量, b_j 表示 b 的第 j 个分量. 为了找到对偶函数, 需要固定 ν 找 $L(x, \nu)$ 的最小值. 注意到 $\nabla_x^2 L = 2I \succ 0$, 因此 $L(\cdot, \nu)$ 是一个严格凸函数, 它取最小值的充分必要条件是 $\nabla_x L(x, \nu) = 0$, 即

$$2x + A^T \nu = 0, \quad (3.53)$$

由此解得最小值点为 $x = -(A^T \nu)/2$, 因此有

$$L_D(\nu) = \inf_{x \in \mathbb{R}^n} L(x, \nu) = \frac{1}{4} \nu^T A A^T \nu + \nu^T \left(-\frac{1}{2} A A^T \nu - b \right) = -\frac{1}{4} \nu^T A A^T \nu - \nu^T b, \quad (3.54)$$

这就是优化问题 (3.51) 的 Lagrange 对偶函数.

例 3.3 考察线性规划问题

$$\min c^T x, \quad \text{s.t.} \quad Ax = b, \quad x \succeq 0. \quad (3.55)$$

其中, $c \in \mathbb{R}^n, b \in \mathbb{R}^l$, A 是 $l \times n$ 的实矩阵, $l \leq n$. 首先将不等式约束等价变形为 $-x \preceq 0$, 不难写出此优化问题的 Lagrange 函数为

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x - \lambda^T x + \nu^T (Ax - b) \\ &= (c - \lambda + A^T \nu)^T x - \nu^T b, \quad x, \lambda \in \mathbb{R}^n, \nu \in \mathbb{R}^l. \end{aligned} \quad (3.56)$$

不难看出, 只有当 $c - \lambda + A^T \nu = 0$ 时 $L(x, \lambda, \nu)$ 有下界 $-\nu^T b$, 因此对偶函数为

$$L_D(\lambda, \nu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu) = \begin{cases} -\nu^T b, & c - \lambda + A^T \nu = 0 \\ -\infty, & \text{其他} \end{cases}. \quad (3.57)$$

3.3.2 最优值的下界估计

利用 Lagrange 对偶函数可以给出原优化问题的最优值 p^* 的下界估计. 事实上, 有如下结果.

定理 3.5 设 $L_D(\lambda, \nu)$ 是优化问题 (3.1)~(3.3) 的 Lagrange 对偶函数, p^* 是该问题的最优值, 则有

$$L_D(\lambda, \nu) \leq p^*, \quad \forall \lambda \geq 0, \nu \in \mathbb{R}^l. \quad (3.58)$$

证明 设 $\mathcal{D}$ 和 S_F 是优化问题 (3.1)~(3.3) 的定义域和可行集, $L(x, \lambda, \nu)$ 是此优化问题的 Lagrange 函数, 则对任意 $x \in S_F$ 皆有

$$g_i(x) \leq 0, \quad i = 1, 2, \dots, m, \quad h_j(x) = 0, \quad j = 1, 2, \dots, l, \quad (3.59)$$

于是当 $\lambda \geq 0$ 时, 有

$$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^l \nu_j h_j(x) \leq f(x), \quad (3.60)$$

从而得到

$$L_D(\lambda, \nu) = \inf_{z \in \mathcal{D}} L(z, \lambda, \nu) \leq L(x, \lambda, \nu) \leq f(x), \quad (3.61)$$

由于上式对于任意 $x \in S_F$ 皆成立, 因此有

$$L_D(\lambda, \nu) \leq \inf_{x \in S_F} f(x) = p^*. \quad (3.62)$$

□

在例 3.2 中, 我们已经求出了优化问题 (3.51) 的对偶函数

$$L_D(\nu) = -\frac{1}{4} \nu^T A A^T \nu - \nu^T b, \quad (3.63)$$

设 p^* 是该优化问题的最优值, 根据定理 3.5 得

$$L_D(\nu) \leq p^*, \quad \forall \nu \in \mathbb{R}^l, \quad (3.64)$$

因此有

$$p^* \geq \sup_{\nu \in \mathbb{R}^l} L_D(\nu), \quad (3.65)$$

如果 A 是行满秩的, 则 $A A^T$ 可逆, $L_D(\nu)$ 取最大值的充分必要条件为

$$\nabla L_D(\nu) = -\frac{1}{2} A A^T \nu - b = 0, \quad (3.66)$$

由此推出最大值点为 $\nu = -2(AA^T)^{-1}b$, 将其代入 L_D 的表达式得到 L_D 的最大值为 $b^T(AA^T)^{-1}b$, 因此有

$$p^* \geq \sup_{\nu \in \mathbb{R}^p} L_D(\nu) = b^T(AA^T)^{-1}b. \quad (3.67)$$

在例 3.3 中, 我们求出了线性规划问题 (3.55) 的对偶函数

$$L_D(\lambda, \nu) = \begin{cases} -\nu^T b, & \lambda + A^T \nu = c \\ -\infty, & \text{其他} \end{cases}, \quad (3.68)$$

因此该问题的最优值 p^* 满足

$$p^* \geq \sup_{\lambda \geq 0, \nu \in \mathbb{R}^p} L_D(\lambda, \nu) = \sup \{-\nu^T b : \lambda + A^T \nu = c, \lambda \geq 0\}. \quad (3.69)$$

3.3.3 Lagrange 对偶函数与共轭函数的关系

Lagrange 对偶与第 2 章讲的共轭函数有密切联系, 本小节给出一些例子. 先回忆一下共轭函数的定义为

$$f^*(y) = \sup_{x \in \text{Dom}(f)} (y^T x - f(x)). \quad (3.70)$$

例 3.4 考察优化问题

$$\min f(x), \quad \text{s.t.} \quad Ax \preceq b, \quad Cx = d, \quad (3.71)$$

其 Lagrange 函数为

$$L(x, \lambda, \nu) = f(x) + \lambda^T (Ax - b) + \nu^T (Cx - d), \quad (3.72)$$

因此其对偶函数为

$$\begin{aligned} L_D(\lambda, \nu) &= \inf_{x \in \text{Dom}(f)} L(x, \lambda, \nu) = \inf_{x \in \text{Dom}(f)} \{f(x) + (C^T \nu + A^T \lambda)^T x - \lambda^T b - \nu^T d\} \\ &= -\lambda^T b - \nu^T d - \sup_{x \in \text{Dom}(f)} \{-(C^T \nu + A^T \lambda)^T x - f(x)\} \\ &= -\lambda^T b - \nu^T d - f^*(-C^T \nu - A^T \lambda). \end{aligned} \quad (3.73)$$

例 3.5 考察优化问题

$$\min \|x\|, \quad \text{s.t.} \quad Ax = b. \quad (3.74)$$

其中, $\|\cdot\|$ 是 $\mathbb{R}^n$ 上的一个一般的范数, 其 Lagrange 函数为

$$L(x, \nu) = \|x\| + \nu^T (Ax - b), \quad (3.75)$$

因此其对偶函数为

$$\begin{aligned} L_D(\nu) &= \inf_{x \in \mathbb{R}^n} L(x, \nu) = -\nu^T b - \sup_{x \in \mathbb{R}^n} (-(A^T \nu)^T x - \|x\|) \\ &= -\nu^T b - f^*(-A^T \nu), \end{aligned} \quad (3.76)$$

其中, f^* 是函数 $f(x) = \|x\|$ 的共轭函数. 利用例 2.20 的结论得到

$$L_D(\nu) = \begin{cases} -\nu^T b, & \|A^T \nu\|_* \leq 1 \\ -\infty, & \text{其他} \end{cases}. \quad (3.77)$$

例 3.6 考察熵最大化问题

$$\min f(x) = \sum_{i=1}^n x_i \ln x_i, \quad \text{s.t.} \quad Ax \preceq b, \quad \sum_{i=1}^n x_i = 1. \quad (3.78)$$

其中, $-f(x)$ 代表概率分布 $P\{X = i\} = x_i, i = 1, 2, \dots, n$ 的熵. 我们先来计算 f 的共轭函数.

$$\begin{aligned} f^*(y) &= \sup_{x \succ 0} \left(y^T x - \sum_{i=1}^n x_i \ln x_i \right) = \sup_{x \succ 0} \left(\sum_{i=1}^n (x_i y_i - x_i \ln x_i) \right) \\ &= \sum_{i=1}^n \sup_{x_i > 0} (x_i y_i - x_i \ln x_i) \\ &= \sum_{i=1}^n e^{y_i - 1}. \end{aligned} \quad (3.79)$$

现在利用例 3.4 的结论得到

$$\begin{aligned} L_D(\lambda, \nu) &= -\nu - \lambda^T b - f^*(-\nu \mathbb{1}_{n \times 1} - A^T \lambda) \\ &= -\nu - \lambda^T b - \sum_{i=1}^n e^{-\nu - a_i^T \lambda - 1} \\ &= -\nu - \lambda^T b - e^{-\nu - 1} \sum_{i=1}^n e^{-a_i^T \lambda}, \end{aligned} \quad (3.80)$$

其中, $\mathbb{1}_{n \times 1}$ 表示元素全是 1 的 $n \times 1$ 矩阵, a_i 表示 A 的第 i 个列向量.

例 3.7 考察优化问题

$$\min f(X) = \ln(\det X^{-1}), \quad \text{s.t.} \quad a_i^T X a_i \leq 1, \quad i = 1, 2, \dots, m, \quad (3.81)$$

其中, 矩阵 X 在 $\mathbf{S}_{++}^n$ 中取值. 注意到 $\det X^{-1}$ 与椭球 $a^T X a \leq 1$ 的体积呈比例, 因此这个问题的本质是要找一个中心在坐标原点的椭球 $a^T X a \leq 1$ 以覆盖住给定点 $a_i, i = 1, 2, \dots, m$, 并使其具有最小体积. 约束条件 $a_i^T X a_i \leq 1$ 是线性的, 可以等价地写成

$$\text{tr}((a_i a_i^T) X) \leq 1, \quad i = 1, 2, \dots, m. \quad (3.82)$$

利用例 2.18 的结论得到

$$f^*(Y) = \sup_{X \in \mathbf{S}_{Y++}^n} g_Y(X) = \begin{cases} -n + \ln \det(-Y^{-1}), & Y \prec 0 \\ \infty, & Y \succeq 0 \end{cases}, \quad (3.83)$$

再利用例 3.4 的结论得到

$$\begin{aligned} L_D(\lambda) &= -\lambda^T \mathbb{1}_{m \times 1} - f^* \left(-\sum_{i=1}^m \lambda_i a_i a_i^T \right) \\ &= \begin{cases} -\sum_{i=1}^m \lambda_i + n + \ln \left(\det \left(\sum_{i=1}^m \lambda_i a_i a_i^T \right) \right), & \sum_{i=1}^m \lambda_i a_i a_i^T \succ 0 \\ -\infty, & \text{其他} \end{cases}. \end{aligned} \quad (3.84)$$

3.4 Lagrange 对偶问题

3.4.1 对偶问题的概念及例子

回到一般的优化问题 (3.1)~(3.3), 设它的最优值为 p^* , Lagrange 对偶函数为 $L_D(\lambda, \nu)$. 根据定理 3.5 得

$$L_D(\lambda, \nu) \leq p^*, \quad \forall \lambda \succeq 0, \nu \in \mathbb{R}^l, \quad (3.85)$$

即任意满足 $\lambda \succeq 0$ 的有序对 (λ, ν) 都给出了 p^* 的一个下界 $L_D(\lambda, \nu)$. 现在想要找一个最优的下界, 就要求解优化问题

$$\max L_D(\lambda, \nu), \quad \text{s.t.} \quad \lambda \succeq 0, \quad (3.86)$$

我们称式 (3.86) 为优化问题 (3.1)~(3.3) 的 **Lagrange 对偶问题 (Lagrange dual problem)**, 或简称之为**对偶问题 (dual problem)**. 如果 (λ, ν) 满足 $\lambda \succeq 0$ 且 $L_D(\lambda, \nu) > -\infty$, 则称之为**对偶可行点 (dual feasible point)**. 所有对偶可行点的集合称为**对偶可行集 (dual feasible set)**. 称对偶问题 (3.86) 的最优点 (λ^*, ν^*) 为**对偶最优点 (dual optimal point)**.

须指出的是, 无论原问题是否为凸优化问题, 其对偶问题一定是凸优化问题.

例 3.8 考察例 3.3 中的线性规划问题 (3.55), 我们已经求出了其对偶函数为

$$L_D(\lambda, \nu) = \begin{cases} -\nu^T b, & c - \lambda + A^T \nu = 0 \\ -\infty, & \text{其他} \end{cases}, \quad (3.87)$$

因此其对偶问题是

$$\max L_D(\lambda, \nu) = \begin{cases} -\nu^T b, & c - \lambda + A^T \nu = 0 \\ -\infty, & \text{其他} \end{cases}. \quad (3.88)$$

$$\text{s.t. } \lambda \succeq 0, \quad (3.89)$$

注意目标函数中隐含了约束条件 $c - \lambda + A^T \nu = 0$, 因此上述问题等价于

$$\max -\nu^T b, \quad \text{s.t. } \lambda \succeq 0, \quad c - \lambda + A^T \nu = 0. \quad (3.90)$$

利用等式约束条件消掉 λ 后得到等价问题

$$\max -\nu^T b, \quad \text{s.t. } A^T \nu + c \succeq 0. \quad (3.91)$$

上述问题也被称作线性规划问题 (3.55) 的对偶问题.

设 p^* 是原问题 (3.1)~(3.3) 的最优值, d^* 是对偶问题 (3.86) 的最优值, 则有

$$d^* \leq p^*. \quad (3.92)$$

这个性质称为弱对偶性 (weak duality). 我们称 $p^* - d^*$ 为最优对偶间隙 (optimal duality gap).

例 3.9 我们来考虑一个集合划分的问题. 设 $E = \{v_i : i = 1, 2, \dots, n\}$, 我们需要将其中的元素分成两类, 如果 v_i 被分到第一类, 则给其赋予标签 $x_i = 1$, 如果 v_i 被分到第二类, 则给其赋予标签 $x_i = -1$, 因此将 E 中的元素分类就相当于给出一个标签向量 $x = (x_1, x_2, \dots, x_n)^T$, 它的每一个分量 x_i 取 1 或 -1 . 现在来考虑分类代价函数, 如果 x_i 与 x_j 被划分在同一类, 则分类代价为 $2w_{ij}$, 否则分类代价为 $-2w_{ij}$, 则总的分类代价为

$$f(x) = \sum_{i=1}^n \sum_{j=1}^n x_i x_j w_{ij} = x^T W x, \quad (3.93)$$

其中, $W = (w_{ij})_{n \times n}$ 是一个 n 阶实对称矩阵. 集合划分问题就是要找一个代价最小的划分, 即

$$\min f(x) = x^T W x, \quad \text{s.t. } x_i^2 = 1, \quad i = 1, 2, \dots, n. \quad (3.94)$$

这是一个非凸的优化问题, 下面找其对偶问题, 其 Lagrange 函数为

$$L(x, \nu) = x^T W x + \sum_{i=1}^n \nu_i (x_i^2 - 1) = x^T (W + D_\nu) x - \sum_{i=1}^n \nu_i, \quad (3.95)$$

其中, D_ν 表示以 $\nu_1, \nu_2, \dots, \nu_n$ 为对角元素的对角矩阵. 因此对偶函数为

$$L_D(\nu) = \inf_{x \in \mathbb{R}^n} L(x, \nu) = \begin{cases} -\sum_{i=1}^n \nu_i, & W + D_\nu \succeq 0 \\ -\infty, & \text{其他} \end{cases}, \quad (3.96)$$

因此对偶问题为

$$\max_{\nu \in \mathbb{R}^n} L_D(\nu) = \begin{cases} -\sum_{i=1}^n \nu_i, & W + D_\nu \succeq 0 \\ -\infty, & \text{其他} \end{cases}. \quad (3.97)$$

由于目标函数中隐含了约束条件 $W + D_\nu \geq 0$, 因此上述问题又等价于

$$\max - \sum_{i=1}^n \nu_i, \quad \text{s.t.} \quad W + D_\nu \geq 0. \quad (3.98)$$

这是一个凸优化问题, 有非常有效的算法可以求解这类问题.

3.4.2 强对偶性

弱对偶性表明优化问题的最优值 p^* 与对偶最优值 d^* 满足 $d^* \leq p^*$. 那么等式是否成立呢? 对于一般的优化问题, 并不能保证 $d^* = p^*$. 例如优化问题

$$\min f(x) = e^{-x_1}, \quad \text{s.t.} \quad \frac{x_1^2}{x_2} \leq 0, \quad (3.99)$$

其中, $x = (x_1, x_2)^T$ 在 $\mathcal{D} := \{(x_1, x_2)^T : x_2 > 0\}$ 中取值. 这个问题的可行集是 $S_F = \{(0, x_2)^T : x_2 > 0\}$, 因此最优值为 $p^* = 1$, 其对偶函数为

$$L_D(\lambda) = \inf_{x \in \mathcal{D}} \left(e^{-x_1} + \lambda \frac{x_1^2}{x_2} \right) = \begin{cases} 0, & \lambda \geq 0 \\ -\infty, & \lambda < 0. \end{cases} \quad (3.100)$$

因此对偶问题为

$$\max_{\lambda \geq 0} L_D(\lambda), \quad (3.101)$$

对偶最优值为 $d^* = 0$, 从而有 $d^* < p^*$.

如果一个优化问题的最优值 p^* 和对偶最优值 d^* 相等, 即 $p^* = d^*$, 则称它满足**强对偶性 (strong duality)**.

接下来研究保证强对偶性的条件.

对于 $\mathbb{R}^n$ 的子集 E , 其仿射包为 $\text{aff}(E)$, 如果 E 的仿射维数小于 n , 则 $\text{aff}(E)$ 是一个维数小于 n 的仿射集. 对于 $x \in E$, 如果存在开球 $B(x, r) := \{z \in \mathbb{R}^n : \|z - x\|_2 < r\}$ 使得

$$B(x, r) \cap \text{aff}(E) \subseteq E, \quad (3.102)$$

则称 x 是 E 的**相对内点 (relative interior point)**. 称 E 的所有相对内点所构成的集合为 E 的**相对内部 (relative interior)**, 记为 $\text{relint}(E)$.

回到凸优化问题 (3.14)~(3.16), 设其定义域为 $\mathcal{D}$, 如果存在 $x \in \text{relint}(\mathcal{D})$ 使得

$$g_i(x) < 0, \quad i = 1, 2, \dots, m, \quad a_j^T x - b_j = 0, \quad j = 1, 2, \dots, l, \quad (3.103)$$

则称该优化问题满足 **Slater 条件 (Slater's condition)**, 称这样的点 x 为**严格可行点 (strictly feasible point)**.

如果 $g_i, i = 1, 2, \dots, m$ 中含有仿射函数, 不妨设前 $k(k \leq m)$ 个是仿射函数, 则可以将 Slater 条件减弱, 定义弱 Slater 条件 (weak Slater's condition) 为存在 $x \in \text{relint}(\mathcal{D})$ 使得

$$\begin{aligned} g_i(x) &\leq 0, \quad i = 1, 2, \dots, k, & g_i(x) &< 0, \quad i = k+1, k+2, \dots, m, \\ a_j^T x - b_j &= 0, \quad j = 1, 2, \dots, l. \end{aligned} \quad (3.104)$$

定理 3.6 对于凸优化问题 (3.14)~(3.16), 如果它满足 Slater 条件, 则它满足强对偶性, 而且对偶最优值是可达到的. 如果不等式约束函数中有仿射函数, 则只需满足弱 Slater 条件, 上述结论就成立.

证明 我们只证 $\mathcal{D}$ 的仿射维数等于 n 的情形, 此时 $\text{relint}(\mathcal{D}) = D^\circ$. 先证明定理的第一部分. 首先, 我们可以假设 $a_1, a_2, \dots, a_l$ 线性无关, 否则可以去掉一些冗余的等式约束直至剩余向量线性无关, 这样做不会改变问题的可行集, 最优值 p^* 和对偶最优值 d^* 都不会改变. 为了表示简洁, 令

$$g = (g_1, g_2, \dots, g_m)^T, \quad h_j(x) = a_j^T x - b_j, \quad h = (h_1, h_2, \dots, h_l)^T. \quad (3.105)$$

采用这些记号可将凸优化问题 (3.14)~(3.16) 表示为

$$\min f(x), \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \quad (3.106)$$

原问题的 Lagrange 函数为

$$L(x, \lambda, \nu) = f(x) + \lambda^T g(x) + \nu^T h(x). \quad (3.107)$$

现在定义 $\mathbb{R}^m \times \mathbb{R}^l \times \mathbb{R}$ 的子集

$$\mathcal{A} := \{(u, v, t) : \exists x \in \mathcal{D}, g(x) \leq u, h(x) = v, f(x) \leq t\}, \quad (3.108)$$

如果原问题是可行的, 则 $\mathcal{A}$ 非空. 这是因为如果 x_0 是原问题的可行点, 则对任意满足 $s \geq f(x_0)$ 的实数 s 皆有 $(0, 0, s) \in \mathcal{A}$. 此外, $\mathcal{A}$ 还是凸集. 这是因为对任意 $(u, v, t), (\tilde{u}, \tilde{v}, \tilde{t}) \in \mathcal{A}$, 存在 $x, \tilde{x} \in \mathcal{D}$ 使得

$$g(x) \leq u, \quad g(\tilde{x}) \leq \tilde{u}, \quad h(x) = v, \quad h(\tilde{x}) = \tilde{v}, \quad f(x) \leq t, \quad f(\tilde{x}) \leq \tilde{t}, \quad (3.109)$$

由于 g 的每个分量是凸函数, h 的每个分量是仿射函数, f 是凸函数, 因此对任意 $0 < \theta < 1$ 皆有

$$g((1-\theta)x + \theta\tilde{x}) \leq (1-\theta)g(x) + \theta g(\tilde{x}) \leq (1-\theta)u + \theta\tilde{u}, \quad (3.110)$$

$$h((1-\theta)x + \theta\tilde{x}) = (1-\theta)h(x) + \theta h(\tilde{x}) = (1-\theta)v + \theta\tilde{v}, \quad (3.111)$$

$$f((1-\theta)x + \theta\tilde{x}) \leq (1-\theta)f(x) + \theta f(\tilde{x}) \leq (1-\theta)t + \theta\tilde{t}, \quad (3.112)$$

由于 $\mathcal{D}$ 是凸集, 因此 $(1-\theta)x + \theta\tilde{x} \in \mathcal{D}$, 由此得出

$$(1-\theta)(u, v, t) + \theta(\tilde{u}, \tilde{v}, \tilde{t}) \in \mathcal{A}, \quad (3.113)$$

这就证明了 $\mathcal{A}$ 是凸集.

再定义

$$\mathcal{B} := \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^l \times \mathbb{R} : s < p^*\}, \quad (3.114)$$

则 $\mathcal{B}$ 显然也是凸集, 且 $\mathcal{A} \cap \mathcal{B} = \emptyset$. 这是因为如果 $(0, 0, s) \in \mathcal{A}$, 则存在 $x_0 \in \mathcal{D}$ 使得

$$g(x_0) \preceq 0, \quad h(x_0) = 0, \quad f(x_0) \leq s, \quad (3.115)$$

因此

$$s \geq f(x_0) \geq \inf_{x \in S_F} f(x) = p^*, \quad (3.116)$$

其中, S_F 是原优化问题的可行集, 因此 $(0, 0, s) \notin \mathcal{B}$.

如果原凸优化问题满足 Slater 条件, 则原问题至少有一个可行点, 因此 $p^* < \infty$, 如果 $p^* = -\infty$, 则根据弱对偶性, $d^* = -\infty$, 定理结论成立. 如果 $p^* > -\infty$, 则 $\mathcal{A}$ 和 $\mathcal{B}$ 都是非空凸集, 根据凸集分离定理, 存在 $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$ 使得

$$\tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \geq \alpha, \quad \forall (u, v, t) \in \mathcal{A}, \quad (3.117)$$

$$\mu s \leq \alpha, \quad \forall (0, 0, s) \in \mathcal{B}, \quad (3.118)$$

由式 (3.117) 得 $\tilde{\lambda} \succeq 0, \mu \geq 0$, 否则 $\tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t$ 在 $\mathcal{A}$ 上无下界. 再由式 (3.118) 得 $\mu s \leq \alpha, \forall s < p^*$, 由此推出 $\mu p^* \leq \alpha$. 综合以上推论得

$$\tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \geq \alpha, \quad \tilde{\lambda} \succeq 0, \quad \mu \geq 0, \quad \mu p^* \leq \alpha, \quad \forall (u, v, t) \in \mathcal{A}. \quad (3.119)$$

对任意 $x \in \mathcal{D}$ 皆有 $(g(x), h(x), f(x)) \in \mathcal{A}$, 因此有

$$\tilde{\lambda}^T g(x) + \tilde{\nu}^T h(x) + \mu f(x) \geq \alpha \geq \mu p^*. \quad (3.120)$$

如果 $\mu > 0$, 则由式 (3.120) 得到

$$\frac{1}{\mu} \tilde{\lambda}^T g(x) + \frac{1}{\mu} \tilde{\nu}^T h(x) + f(x) \geq p^*, \quad \forall x \in \mathcal{D}, \quad (3.121)$$

即

$$L\left(x, \frac{\tilde{\lambda}}{\mu}, \frac{\tilde{\nu}}{\mu}\right) \geq p^*, \quad \forall x \in \mathcal{D}, \quad (3.122)$$

因此有

$$L_D\left(\frac{\tilde{\lambda}}{\mu}, \frac{\tilde{\nu}}{\mu}\right) = \inf_{x \in \mathcal{D}} L\left(x, \frac{\tilde{\lambda}}{\mu}, \frac{\tilde{\nu}}{\mu}\right) \geq p^*, \quad (3.123)$$

由此得到 $d^* = \sup_{\lambda \geq 0, \nu \in \mathbb{R}^l} L_D(\lambda, \nu) \geq p^*$, 结合弱对偶性立刻得到 $d^* = p^*$, 这就证明了强对偶性, 而且证明了对偶最优值是可达到的.

接下来证明 μ 不可能为零. 如果 $\mu = 0$, 则由式 (3.120) 得到

$$\tilde{\lambda}^T g(x) + \tilde{\nu}^T h(x) \geq 0, \quad \forall x \in \mathcal{D}. \quad (3.124)$$

根据 Slater 条件, 存在 $\tilde{x} \in \mathcal{D}^\circ$ 使得 $g(\tilde{x}) \prec 0, h(\tilde{x}) = 0$, 将其代入式 (3.124) 得

$$\tilde{\lambda}^T g(\tilde{x}) \geq 0, \quad (3.125)$$

由于 $\tilde{\lambda} \succeq 0$, 因此必有 $\tilde{\lambda} = 0$, 于是由式 (3.124) 得

$$\tilde{\nu}^T h(x) \geq 0, \quad \forall x \in \mathcal{D}. \quad (3.126)$$

既然 $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$, 必有 $\tilde{\nu} \neq 0$. 由于 $\tilde{x} \in \mathcal{D}^\circ$, 因此存在 $\delta > 0$, 使得当 $y \in B(0, \delta)$ 时恒有 $\tilde{x} + y \in \mathcal{D}$, 于是有

$$\tilde{\nu}^T A^T y = \tilde{\nu}^T h(\tilde{x} + y) \geq 0, \quad \forall \|y\|_2 < \delta, \quad (3.127)$$

其中, $A = (a_1, a_2, \dots, a_l), h(x) = A^T x - b, b = (b_1, b_2, \dots, b_l)^T$. 由此得到

$$A\tilde{\nu} = 0,$$

即

$$\sum_{j=1}^l \tilde{\nu}_j a_j = 0, \quad (3.128)$$

由于已经假设 $a_1, a_2, \dots, a_l$ 线性无关, 因此必有 $\tilde{\nu} = 0$, 矛盾.

接下来证明定理的第二部分. 如果原问题含有仿射不等式约束, 例如

$$a_0^T x - b_0 \leq 0, \quad (3.129)$$

则加入松弛变量 y 使其转换为等式约束, 原问题等价于

$$\min \varphi(x, y) = f(x), \quad (3.130)$$

$$\text{s.t.} \quad g(x) \leq 0, \quad a_0^T x + y - b_0 = 0, \quad a_j^T x - b_j = 0, \quad j = 1, 2, \dots, l, \quad (3.131)$$

此时问题的定义域为

$$\mathcal{D}' = \{(x, y) : x \in \mathcal{D}, y \in \mathbb{R}\}. \quad (3.132)$$

如果原问题满足弱 Slater 条件, 则加松弛变量后的问题满足 Slater 条件, 根据已经证明的前部分结论, 存在 $\tilde{\lambda} \succeq 0, \tilde{\nu} \in \mathbb{R}^p$ 及 $\tilde{\nu}_0 \in \mathbb{R}$ 使得

$$L'_D(\tilde{\lambda}, \tilde{\nu}, \tilde{\nu}_0) = \inf_{(x, y) \in \mathcal{D}'} L'(x, y, \tilde{\lambda}, \tilde{\nu}, \tilde{\nu}_0)$$

$$:= \inf_{(x,y) \in \mathcal{D}'} f(x) + \tilde{\lambda}^T g(x) + \tilde{\nu}^T h(x) + \tilde{\nu}_0(a_0^T x + y - b_0) = p^*. \quad (3.133)$$

如果 $p^* = -\infty$, 则定理结论显然成立. 如果 p^* 是有限实数, 则必有 $\tilde{\nu}_0 = 0$ (否则 $\tilde{\nu}_0(a_0^T x + y - b_0)$ 无下界), 此时原问题的 Lagrange 函数满足

$$L(x, \tilde{\lambda}, \tilde{\nu}, \tilde{\nu}_0) = f(x) + \tilde{\lambda}^T g(x) + \tilde{\nu}^T h(x) + \tilde{\nu}_0(a_0^T x - b_0), \quad (3.134)$$

因此有

$$\begin{aligned} L_D(\tilde{\lambda}, \tilde{\nu}, \tilde{\nu}_0) &= \inf_{x \in \mathcal{D}} L(x, \tilde{\lambda}, \tilde{\nu}, \tilde{\nu}_0) = \inf_{(x,0) \in \mathcal{D}'} L'(x, 0, \tilde{\lambda}, \tilde{\nu}, \tilde{\nu}_0) \\ &\geq \inf_{(x,y) \in \mathcal{D}'} L'(x, y, \tilde{\lambda}, \tilde{\nu}, \tilde{\nu}_0) = p^*, \end{aligned} \quad (3.135)$$

由此推出 $d^* \geq p^*$, 强对偶性得证. $\square$

例 3.10 考虑仿射约束凸优化问题

$$\min f(x), \quad \text{s.t.} \quad Cx \leq d, \quad Ax = b, \quad (3.136)$$

其中, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是凸函数, C 是 $m \times n$ 的实矩阵, A 是 $l \times n$ 的实矩阵. 只要该问题是可行的, 就一定满足弱 Slater 条件, 根据定理 3.6, 它满足强对偶性, 而且存在 $\tilde{\lambda} \in \mathbb{R}^m, \tilde{\nu} \in \mathbb{R}^l$ 使得 $L_D(\tilde{\lambda}, \tilde{\nu}) = d^* = p^*$.

作为特例, 线性规划问题

$$\min \alpha^T x, \quad \text{s.t.} \quad c_i^T x \leq d_i, \quad i = 1, 2, \dots, m, \quad a_j^T x = b_j, \quad j = 1, 2, \dots, l \quad (3.137)$$

只要是可行的, 就满足强对偶性, 而且对偶最优值是可以取到的.

例 3.11 回到例 3.6 的熵最大化问题 (3.78), 我们已经求出其对偶函数

$$L_D(\lambda, \nu) = -\nu - \lambda^T b - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^T \lambda}, \quad (3.138)$$

因此对偶问题为

$$\max L_D(\lambda, \nu) = -\nu - \lambda^T b - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^T \lambda}, \quad \text{s.t.} \quad \lambda \succeq 0. \quad (3.139)$$

这个问题还可以简化, 对于给定的 $\lambda \succeq 0$, 先将 $L_D(\lambda, \nu)$ 对 ν 取最大值, 得

$$\sup_{\nu \in \mathbb{R}} L_D(\lambda, \nu) = -\ln \left(\sum_{i=1}^n e^{-a_i^T \lambda} \right) - \lambda^T b, \quad (3.140)$$

因此对偶问题 (3.139) 等价于

$$\max -\ln \left(\sum_{i=1}^n e^{-a_i^T \lambda} \right) - \lambda^T b, \quad \text{s.t.} \quad \lambda \succeq 0, \quad (3.141)$$

这是一个带有非负约束的几何规划问题, 而且是凸优化问题. 由于原问题的约束条件都是仿射约束, 根据定理 3.6, 只要存在 $x \succ 0$ 满足 $Ax \preceq b$ 及 $\sum_{i=1}^n x_i = 1$, 它就满足强对偶性, 而且对偶最优值 $d^* = p^*$ 是可取到的.

例 3.12 回到例 3.7 介绍的最小体积椭圆覆盖问题 (3.81), 我们已经求出其对偶函数

$$L_D(\lambda) = \begin{cases} -\sum_{i=1}^m \lambda_i + n + \ln \left(\det \left(\sum_{i=1}^m \lambda_i a_i a_i^T \right) \right), & \sum_{i=1}^m \lambda_i a_i a_i^T \succ 0 \\ -\infty, & \text{其他} \end{cases}, \quad (3.142)$$

因此对偶问题为

$$\max -\sum_{i=1}^m \lambda_i + n + \ln \left(\det \left(\sum_{i=1}^m \lambda_i a_i a_i^T \right) \right), \quad \text{s.t.} \quad \lambda \succ 0. \quad (3.143)$$

原问题的不等式约束 $a_i^T X a_i \leq 1$ 对变量 X 是仿射约束, 因此只要满足弱 Slater 条件, 原问题就具有强对偶性. 弱 Slater 条件相当于存在 $X \in S_{++}$ 使得

$$a_i^T X a_i \leq 1, \quad i = 1, 2, \dots, m, \quad (3.144)$$

即原问题是可行的.

3.5 最优性条件

3.5.1 无约束优化问题的最优性条件

本小节我们考虑无约束优化问题

$$\min_{x \in \mathcal{D}} f(x), \quad (3.145)$$

其中, $\mathcal{D}$ 是 $\mathbb{R}^n$ 中的开集, f 是一阶连续可微的, 这里并不假设 f 是凸函数. 如果 $x^* \in \mathcal{D}$ 是局部最优点, 则对任意 $v \in \mathbb{R}^n$, 当实数 t 的绝对值足够小时必有

$$f(x^*) \leq f(x^* + tv). \quad (3.146)$$

记 $\varphi(t) := f(x^* + tv)$, 则 φ 在 $t = 0$ 处取得局部极小值, 因此有

$$\varphi'(0) = v^T \nabla f(x^*) = 0, \quad (3.147)$$

由 v 的任意性推出 $\nabla f(x^*) = 0$. 于是我们得到了下列必要条件.

定理 3.7 设 $\mathcal{D}$ 是 $\mathbb{R}^n$ 中的开集, f 是一阶连续可微的, 则 x^* 是无约束优化问题 (3.145) 的局部极小点的必要条件是 $\nabla f(x^*) = 0$.

须指出的是, 上述条件是必要的, 但不是充分的. 如果 f 是二阶连续可微的, 则有下列充分条件.

定理 3.8 设 $\mathcal{D}$ 是 $\mathbb{R}^n$ 中的开集, f 是二阶连续可微的, 则 x^* 是无约束优化问题 (3.145) 的严格局部极小点的充分条件是

$$\nabla f(x^*) = 0, \quad \nabla^2 f(x^*) \succ 0. \quad (3.148)$$

证明 由于 f 二阶连续可微, 因此 $\nabla^2 f(x)$ 是连续依赖于 x 的, 因此存在 $\delta > 0$, 使得当 $x \in B(x^*, \delta)$ 时恒有 $\nabla^2 f(x) \succ 0$. 对任意 $x \in B(x^*, \delta) \setminus \{x^*\}$, 由 Taylor 公式得

$$\begin{aligned} f(x) &= f(x^*) + (x - x^*)^T \nabla f(x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x^* + \theta(x - x^*))(x - x^*) \\ &= f(x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x^* + \theta(x - x^*))(x - x^*), \end{aligned} \quad (3.149)$$

其中, $0 < \theta < 1$. 由于 $x^* + \theta(x - x^*) \in B(x^*, \delta)$, 因此 $\nabla^2 f(x^* + \theta(x - x^*)) \succ 0$, 从而有

$$(x - x^*)^T \nabla^2 f(x^* + \theta(x - x^*))(x - x^*) > 0, \quad (3.150)$$

由此得到 $f(x) > f(x^*)$, $\forall x \in B(x^*, \delta) \setminus \{x^*\}$, 这就证明了 x^* 是 f 的严格局部极小点. $\square$

3.5.2 只含等式约束的优化问题的最优条件

先考虑只含有一个等式约束的优化问题

$$\min f(x), \quad \text{s.t.} \quad h_1(x) = 0, \quad (3.151)$$

我们假设问题的定义域是 $\mathbb{R}^n$ 中的开集, 目标函数 f 和约束函数 h_1 都是一阶连续可微的.

设 x^* 是问题 (3.151) 的局部极小点, v 是与 $\nabla h_1(x^*)$ 正交的向量, 则有

$$f(x^* + tv) = f(x^*) + tv^T \nabla f(x^*) + o(t), \quad (3.152)$$

$$h_1(x^* + tv) = h_1(x^*) + tv^T \nabla h_1(x^*) + o(t) = h_1(x^*) + o(t), \quad (3.153)$$

如果忽略高阶无穷小量 $o(t)$, 则有

$$f(x^* + tv) \approx f(x^*) + tv^T \nabla f(x^*), \quad (3.154)$$

$$h_1(x^* + tv) \approx h_1(x^*) = 0, \quad (3.155)$$

由于 x^* 是问题 (3.151) 的局部极小点, 因此必有 $v^T \nabla f(x^*) = 0$. 换言之, 每一个与 $\nabla h_1(x^*)$ 正交的向量都必须与 $\nabla f(x^*)$ 正交, 因此 $\nabla f(x^*)$ 与 $\nabla h_1(x^*)$ 共线, 或者说存在 $\nu \in \mathbb{R}$ 使得

$$\nabla f(x^*) = \nu \nabla h_1(x^*). \quad (3.156)$$

再来看有多个等式约束的优化问题

$$\min f(x), \quad \text{s.t.} \quad h_j(x) = 0, \quad j = 1, 2, \dots, l. \quad (3.157)$$

设 x^* 是问题 (3.157) 的局部极小点, 记 $W = \text{Span}\{\nabla h_j(x^*) : j = 1, 2, \dots, l\}$, 即由约束函数的梯度张成的线性空间, 设 $v \in W^\perp$, 则有

$$v^T \nabla h_j(x^*) = 0, \quad j = 1, 2, \dots, l. \quad (3.158)$$

于是对于绝对值足够小的 t 有

$$f(x^* + tv) = f(x^*) + tv^T \nabla f(x^*) + o(t), \quad (3.159)$$

$$h_j(x^* + tv) = h_j(x^*) + tv^T \nabla h_j(x^*) + o(t) = h_j(x^*) + o(t), \quad j = 1, 2, \dots, l. \quad (3.160)$$

如果忽略高阶无穷小量 $o(t)$, 则有

$$f(x^* + tv) \approx f(x^*) + tv^T \nabla f(x^*), \quad (3.161)$$

$$h_j(x^* + tv) \approx h_j(x^*) = 0, \quad j = 1, 2, \dots, l. \quad (3.162)$$

由于 x^* 是问题 (3.157) 的局部极小点, 因此必有 $v^T \nabla f(x^*) = 0$. 换言之, $\nabla f(x^*)$ 与 $W^\perp$ 中每一个向量都正交, 因此 $\nabla f(x^*) \in W$, 从而存在实数 $\nu_1, \nu_2, \dots, \nu_l$ 使得

$$\nabla f(x^*) = \sum_{j=1}^l \nu_j \nabla h_j(x^*). \quad (3.163)$$

归纳以上分析的结果, 得到下列定理.

定理 3.9 设优化问题 (3.157) 的定义域 $\mathcal{D}$ 是 $\mathbb{R}^n$ 的开子集, 目标函数 f 和等式约束函数 $h_j, j = 1, 2, \dots, l$ 都是连续可微的, 且 $l < n$. 设 W 是由约束函数的梯度向量 $\nabla h_j(x^*), j = 1, 2, \dots, l$ 张成的向量空间. 则 x^* 是该优化问题的局部极小点的必要条件是 $\nabla f(x^*) \in W$, 即存在 $\nu = (\nu_1, \nu_2, \dots, \nu_l)^T \in \mathbb{R}^l$ 使得式 (3.163) 成立.

前面的分析非常粗略, 不能充当定理的证明. 下面给出定理的严格证明.

证明 首先可以假设约束函数在点 x^* 的梯度向量

$$\nabla h_1(x^*), \nabla h_2(x^*), \dots, \nabla h_l(x^*), \quad (3.164)$$

线性无关, 否则可以去掉 (在该点的) 冗余约束, 使得剩余的梯度向量线性无关.

接下来用反证法证明定理结论成立. 如果 $\nabla f(x^*) \notin W$, 设 $P_W(\nabla f(x^*))$ 是 $\nabla f(x^*)$ 在 W 上的正交投影, 则 $v = \nabla f(x^*) - P_W(\nabla f(x^*))$ 是与 W 正交的非零向量, 且 $v^T \nabla f(x^*) > 0$. 令 $u = v/\|v\|_2$ 并将其扩充为 $W^\perp$ 的规范正交基

$$\{u, u^{(1)}, u^{(2)}, \dots, u^{(n-l-1)}\}, \quad (3.165)$$

考虑关于 x 的方程组

$$\begin{cases} h_j(x) = 0, & j = 1, 2, \dots, l \end{cases} \quad (3.166)$$

$$\begin{cases} (u^{(k)})^T(x - x^*) = 0, & k = 1, 2, \dots, n-l-1, \end{cases} \quad (3.167)$$

$$\begin{cases} u^T(x - x^*) = t \end{cases} \quad (3.168)$$

其中, t 是一个可变参数. 下面用反函数定理 (附录 D 中的定理 D.1) 证明方程组 (3.166)~(3.168) 在局部范围内有解 $x = x(t)$, 而且它是 t 的连续可微函数. 记

$$\begin{cases} \phi_j(x) = h_j(x), & j = 1, 2, \dots, l \end{cases} \quad (3.169)$$

$$\begin{cases} \phi_{l+k}(x) = (u^{(k)})^T(x - x^*), & k = 1, 2, \dots, n-l-1, \end{cases} \quad (3.170)$$

$$\begin{cases} \phi_n(x) = u^T(x - x^*), & \phi = (\phi_1, \phi_2, \dots, \phi_n)^T \end{cases} \quad (3.171)$$