优化模型与软件工具

主讲教师：董庆兴

华中师范大学 信息管理学院
qxdong@mail.ccnu.edu.cn

2017 年 12 月 19 日

大纲

1. Lagrange 对偶函数

约束最优化问题

- 考虑标准形式的约束优化问题:

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, i = 1, \cdots, m. \\ & h_j(\mathbf{x}) = 0, j = 1, \cdots, p. \end{aligned} \tag{1}$$

  其中 $\mathbf{x} \in \mathbb{R}^n$。
- 设问题的定义域 $\mathcal{D} = \bigcap_{i=0}^{m} \textbf{dom}\ f_i \cap \bigcap_{j=1}^{p} \textbf{dom}\ h_j$ 是非空集合
- 优化问题的最优值为 p^*
- 我们目前并没有假设优化问题 (1) 是凸优化问题

Lagrange 乘子

- Lagrange 对偶的基本思想是在目标函数中考虑约束条件，即添加约束条件的加权和，得到增广的目标函数
- 定义优化问题 (1) 的Lagrange 函数$L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ 为

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x}) \tag{2}$$

- 其中定义域为 $\mathbf{dom}\ \ L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$
- λ_i 称为第 i 个不等式约束 $f_i(\mathbf{x}) \leq 0$ 对应的Lagrange 乘子
- 向量 $\boldsymbol{\lambda}, \boldsymbol{\mu}$ 称为对偶变量或者是问题 (1) 的Lagrange 乘子向量

Lagrange 对偶函数

- 定义Lagrange 对偶函数$g:\mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ 为 Lagrange 函数关于 $\mathbf{x}$ 取得的最小值：即对 $\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathbb{R}^p$，有

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\mathbf{x}\in\mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\mathbf{x}\in\mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x}) \right) \tag{3}$$

- 如果 Lagrange 函数关于 $\mathbf{x}$ 无下界，则对偶函数取值为 $-\infty$
- 因为对偶函数是一族关于 $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ 的仿射函数的逐点下确界，所以即使原问题 (1) 不是凸的，对偶函数也是凹函数
- 一个函数：$f:\mathbb{R}^n \to \mathbb{R}$ 是仿射的，如果它是一个线性函数和一个常数的和，即 $f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{B}, \mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$

逐点上确界的凸性

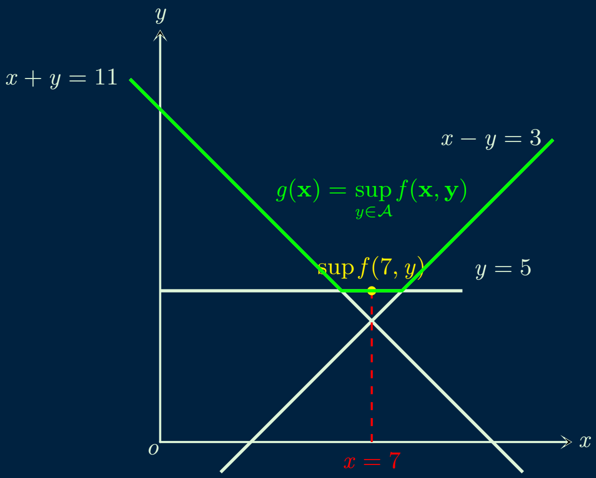
- 一个函数：$f:\mathbb{R}^n\to\mathbb{R}$ 是仿射的，如果它是一个线性函数和一个常数的和，即 $f(\mathbf{x})=\mathbf{A}\mathbf{x}+\mathbf{B},\mathbf{A}\in\mathbb{R}^{m\times n},\mathbf{b}\in\mathbb{R}^m$
- 仿射函数既是凸函数，也是凹函数
- 对于任意 $\mathbf{y}\in\mathcal{A}$，函数 $f(\mathbf{x},\mathbf{y})$ 关于 $\mathbf{x}$ 都是凸的，则函数 g

$$g(\mathbf{x})=\sup_{y\in\mathcal{A}} f(\mathbf{x},\mathbf{y}) \tag{4}$$

  关于 $\mathbf{x}$ 也是凸的，且此时 g 的定义域为
  $\textbf{dom}\ \ g=\{\mathbf{x}|(\mathbf{x},\mathbf{y})\in\textbf{dom}\ \ f,\forall\mathbf{y}\in\mathcal{A},\sup_{\mathbf{y}\in\mathcal{A}}f(\mathbf{x},\mathbf{y})<\infty\}$
- 同样地，我们可知一系列凹函数的逐点下确界仍然是凹函数
- 从上境图的角度理解，一系列函数的逐点上确界函数对应着这些函数上境图的交集

$$\textbf{epi}\ \ g=\bigcap_{\mathbf{y}\in\mathcal{A}}\textbf{epi}\ \ f(\cdot\,,\mathbf{y}) \tag{5}$$

逐点上确界的凸性


y
$x+y=11$
$x-y=3$
$g(\mathbf{x})=\sup_{y\in\mathcal{A}} f(\mathbf{x},\mathbf{y})$
$\sup f(7,y)$
$y=5$
o
x
$x=7$

最优值的下界

- 由

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\mathbf{x} \in \mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x}) \right)$$

  可知，对偶函数构成了原问题 (1) 最优解 p^* 的下界，即对任意 $\boldsymbol{\lambda} \succeq \mathbf{0}$ 和 $\boldsymbol{\mu}$，有

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq p^* \tag{6}$$

- 很容易验证式 (6)。设 $\tilde{\mathbf{x}}$ 是原问题 (1) 的一个可行解，即 $f_i(\tilde{\mathbf{x}}) \leq 0, h_j(\tilde{\mathbf{x}}) = 0$，由假设 $\boldsymbol{\lambda} \succeq \mathbf{0}$，可知

$$\sum_{i=1}^{m} \lambda_i f_i(\tilde{\mathbf{x}}) + \sum_{j=1}^{p} \mu_j h_j(\tilde{\mathbf{x}}) \leq 0$$

最优值的下界

- 从而有

$$L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{\mathbf{x}}) + \sum_{j=1}^{p} \mu_j h_j(\tilde{\mathbf{x}}) \leq f_0(\tilde{\mathbf{x}})$$

- 因此

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq f_0(\tilde{\mathbf{x}})$$

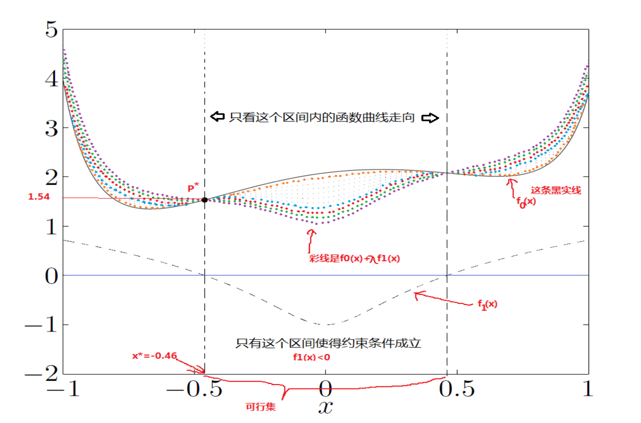
- 因此对于最优值同样成立

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq p^*$$

- 虽然上式成立，但当 $g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\infty$ 时意义不大
- 只有当 $\boldsymbol{\lambda} \succeq \mathbf{0}$ 且 $(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbf{dom}\ g$ 即 $g(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty$ 时，对偶函数才能给出 p^* 的一个非平凡下界
- 此时称满足 $\boldsymbol{\lambda} \succeq \mathbf{0}$ 且 $(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbf{dom}\ g$ 的 $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ 是对偶可行的

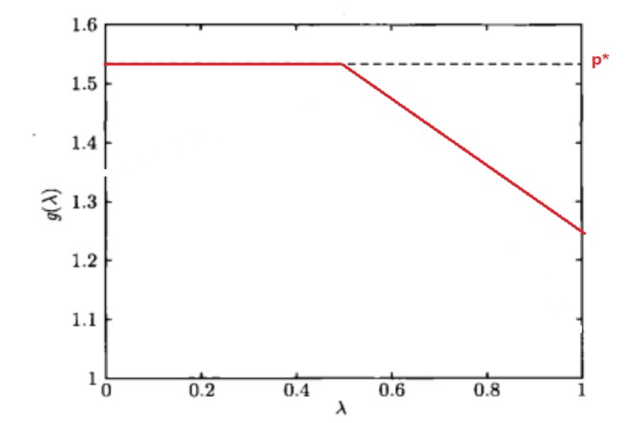
最优值的下界图示

实线表示目标函数 $f_0(x)$，虚线表示约束函数 $f_1(x)$．可行集是区间 $[-0.46, 0.46]$, 如图中两条垂直点线所示.



对偶函数的图像

虽然目标函数 $f_0(x)$ 和约束函数 $f_1(x)$ 都不是凸函数，但是对偶函数 g 是凹函数



线性逼近：硬约束

- 可以通过对集合 $\{0\}$ 和 $-\mathbb{R}^+$ 的示性函数进行线性逼近来理解 Lagrange 函数和其性质
- 首先原问题 (1) 可以重新描述为一个无约束问题

$$\text{minimize} \quad f_0(\mathbf{x}) + \sum_{i=1}^{m} I_-(f_i(\mathbf{x})) + \sum_{j=1}^{p} I_0(h_j(\mathbf{x})) \tag{7}$$

其中 $I_- : \mathbb{R} \to \mathbb{R}$ 为非正实数集的示性函数，I_0 为集合 $\{0\}$ 的示性函数

$$I_-(u) = \begin{cases} 0, & u \le 0 \\ \infty, & u > 0 \end{cases}, \quad I_0(u) = \begin{cases} 0, & u = 0 \\ \infty, & u \neq 0 \end{cases}$$

- 在式 (7) 中，示性函数可以理解为对相应约束函数满足程度的一种约束，以 $I_-(u)$ 为例，随着 $u = f_i(\mathbf{x})$ 从非正数变为正数，我们的不满意度由零升值无穷大

线性逼近：软约束

- 设在式 (7) 中，用线性函数 $\lambda_i u$ 代替 $I_-(u)$, 其中 $\lambda_i \geq 0$，用函数 $\mu_j u$ 代替 $I_0(u)$. 则目标函数变为 Lagrange 函数 $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$，且对偶函数值 $g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ 为问题的最优值

$$\text{minimize} \quad L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x}) \tag{8}$$

- 在式 (8) 中，我们用线性不满意函数替换了示性函数：
  - 对于不等式约束，如果 $f_i(\mathbf{x}) \leq 0$，我们的不满意度为零
  - 当 $f_i(\mathbf{x}) > 0$ 时，不满意度大于零
  - 当 $f_i(\mathbf{x}) > 0$ 时，随着 $f_i(\mathbf{x})$ 增大，我们越来越不满意
- 显然用线性函数逼近示性函数还是远远不够的，但是线性函数至少可以看成示性函数的一个下估计，因为：$\lambda_i u \leq I_-(u)$, $\quad \mu_j u \leq I_0(u)$
- 所以对偶函数是原问题最优函数值的一个下界

线性方程组的最小二乘问题

- 考虑问题:
$$\begin{array}{ll} \text{minimize} & \mathbf{x}^{\mathrm{T}}\mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array} \tag{9}$$
  其中 $\mathbf{A} \in \mathbb{R}^{p\times n}$。该问题无不等式约束，有 p 个线性等式约束
- 其 Lagarange 函数为
$$L(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{x}^{\mathrm{T}}\mathbf{x} + \boldsymbol{\mu}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{b})$$
  $\mathbf{dom}\ \ L : \mathbb{R}^n \times \mathbb{R}^p$
- 对偶函数为
$$g(\boldsymbol{\mu}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu})$$
- 由于 $L(\mathbf{x}, \boldsymbol{\mu})$ 为关于 $\mathbf{x}$ 的二次凸函数，因此可以求解一阶最优条件
$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu}) = 2\mathbf{x} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\mu} = 0$$

线性方程组的最小二乘问题

- 从而可以解得，在 $\mathbf{x} = -\frac{1}{2}\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu}$ 处 Lagrange 函数达到最小值，此时对偶函数为

$$
\begin{aligned}
g(\boldsymbol{\mu}) &= L(-\frac{1}{2}\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu}, \boldsymbol{\mu}) \\
&= \left(-\frac{1}{2}\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu}\right)^{\mathrm{T}}\left(-\frac{1}{2}\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu}\right) + \boldsymbol{\mu}^{\mathrm{T}}\left(\mathbf{A}\left(-\frac{1}{2}\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu}\right) - \mathbf{b}\right) \\
&= -\frac{1}{4}\boldsymbol{\mu}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu} - \mathbf{b}^{\mathrm{T}}\boldsymbol{\mu}
\end{aligned}
$$

- 它是一个二次凹函数，根据对偶函数给出原问题下界的性质, 对任意 $\boldsymbol{\mu} \in \mathbb{R}^p$ 有:

$$
-\frac{1}{4}\boldsymbol{\mu}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu} - \mathbf{b}^{\mathrm{T}}\boldsymbol{\mu} \le \inf\left\{\mathbf{x}^{\mathrm{T}}\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{b}\right\}
$$

标准形式的线性规划问题

- 考虑问题：

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \succeq \mathbf{0} \end{aligned} \tag{10}$$

  该问题不等式约束为 $f_i(\mathbf{x}) = -x_i, i = 1, \ldots, n$，有 p 个线性等式约束

- 其 Lagarange 函数为

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \mathbf{c}^{\mathrm{T}}\mathbf{x} - \sum_{i=1}^{n} \lambda_i x_i + \boldsymbol{\mu}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{b}) \\ &= -\mathbf{b}^{\mathrm{T}}\boldsymbol{\mu} + \left(\mathbf{c} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\mu} - \boldsymbol{\lambda}\right)^{\mathrm{T}}\mathbf{x} \end{aligned}$$

- 对偶函数为

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\mathbf{b}^{\mathrm{T}}\boldsymbol{\mu} + \inf_{\mathbf{x}} \left(\mathbf{c} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\mu} - \boldsymbol{\lambda}\right)^{\mathrm{T}}\mathbf{x}$$

标准形式的线性规划问题

- 由于线性函数只有恒为零时才有下界，因此有

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{cases} -\mathbf{b}^{\mathrm{T}}\boldsymbol{\mu}, & \mathbf{c} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\mu} - \boldsymbol{\lambda} = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

- 只有当 $\boldsymbol{\lambda} \succeq \mathbf{0}$ 且 $\mathbf{c} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\mu} - \boldsymbol{\lambda} = 0$ 的情况下，下界性质才是非平凡的，也就是 $-\mathbf{b}^{\mathrm{T}}\boldsymbol{\mu}$ 是原线性规划问题的一个下界

共轭函数

$f : \mathbb{R}^n \to \mathbb{R}$ 的共轭函数为：

$$f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \mathbf{dom}\ f} \left(\mathbf{y}^{\mathrm{T}}\mathbf{x} - f(\mathbf{x})\right) \tag{11}$$

使上述上确界有限，即差值 $\mathbf{y}^{\mathrm{T}}\mathbf{x} - f(\mathbf{x})$ 在 $\mathbf{dom}\ f$ 有上界的所有 $\mathbf{y} \in \mathbb{R}^n$ 构成了共轭函数的定义域

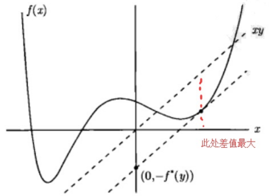


图3.8 函数 $f : \mathbf{R} \to \mathbf{R}$ 以及某一 $y \in \mathbf{R}$。共轭函数 $f^*(y)$ 是线性函数 yx 和 $f(x)$ 之间的最大差值，见图中虚线所示。如果 f 可微，在满足 $f'(x) = y$ 的点 x 处差值最大。

共轭函数性质

- 可见，$f*$ 是凸函数，是一系列 $\mathbf{y}$ 的凸函数的逐点上确界，无论 f 是不是凸函数，f^* 都是凸函数
- 考虑如下常见凸函数的共轭函数:
  - 仿射函数 $f(x) = ax + b$，当且仅当 $y = a$ 时，$yx - ax - b$ 有界，因此 $\mathbf{dom}\ f^* = \{a\}$ 且 $f^*(y) = b$
  - 指数函数 $f(x) = e^x$，当且仅当 $y > 0$ 时，$xy - e^x$ 有界且在 $x = \log y$ 处达到最大值，因此 $\mathbf{dom}\ f^* = \mathbb{R}_+$ 且 $f^*(y) = y\log y - y$
  - 负熵函数 $f(x) = x\log x$,$\mathbf{dom}\ f = \mathbb{R}_+$, 规定 $(f(0) = 0)$，对所有 y，$xy - x\log x$ 有上界且在 $x = e^{y-1}$ 处，因此 $\mathbf{dom}\ f^* = \mathbb{R}$ 且 $f^*(y) = e^{y-1}$

共轭函数与 Lagrange 对偶函数

考虑一个约束优化问题:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & \mathbf{A}\mathbf{x} \preceq \mathbf{b} \\ & \mathbf{C}\mathbf{x} = \mathbf{d} \end{array} \tag{12}$$

其对偶函数为

$$\begin{aligned} g(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{b}) + \boldsymbol{\mu}^{\mathrm{T}}(\mathbf{C}\mathbf{x} - \mathbf{d}) \right) \\ &= -\mathbf{b}^{\mathrm{T}}\boldsymbol{\lambda} - \mathbf{d}^{\mathrm{T}}\boldsymbol{\mu} + \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} + \mathbf{C}^{\mathrm{T}}\boldsymbol{\mu}\right) \mathbf{x} \right) \\ &= -\mathbf{b}^{\mathrm{T}}\boldsymbol{\lambda} - \mathbf{d}^{\mathrm{T}}\boldsymbol{\mu} - \sup_{\mathbf{x}} \left(-\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} + \mathbf{C}^{\mathrm{T}}\boldsymbol{\mu}\right) \mathbf{x} - f_0(\mathbf{x}) \right) \\ &= -\mathbf{b}^{\mathrm{T}}\boldsymbol{\lambda} - \mathbf{d}^{\mathrm{T}}\boldsymbol{\mu} - f_0^*\left(-\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} - \mathbf{C}^{\mathrm{T}}\boldsymbol{\mu} \right) \end{aligned} \tag{13}$$

从而 $\mathbf{dom}\ g = \{(\boldsymbol{\lambda}, \boldsymbol{\mu}) | -\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} - \mathbf{C}^{\mathrm{T}}\boldsymbol{\mu} \in \mathbf{dom}\ f_0^*\}$

最大熵问题的对偶函数

$$\begin{aligned}\text{minimize}\quad & f_0(x)=\sum_{i=1}^{n}x_i\log x_i\\ \text{subject to}\quad & \mathbf{A}\mathbf{x}\succeq\mathbf{b}\\ & \mathbf{1}^{\mathrm{T}}\mathbf{x}=1\end{aligned}\tag{14}$$

其对偶函数为可由式（13）给出

$$\begin{aligned}g(\boldsymbol{\lambda},\boldsymbol{\mu})&=-\mathbf{b}^{\mathrm{T}}\boldsymbol{\lambda}-\mathbf{d}^{\mathrm{T}}\boldsymbol{\mu}-f_0^*\Big(-\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda}-\mathbf{C}^{\mathrm{T}}\boldsymbol{\mu}\Big)\\ &=-\mathbf{b}^{\mathrm{T}}\boldsymbol{\lambda}-\boldsymbol{\mu}-\sum_{i=1}^{n}e^{(-\mathbf{a}_i^{\mathrm{T}}\boldsymbol{\lambda}-\mu)-1}\\ &=-\mathbf{b}^{\mathrm{T}}\boldsymbol{\lambda}-\mu-e^{-\mu-1}\sum_{i=1}^{n}e^{-\mathbf{a}_i^{\mathrm{T}}\boldsymbol{\lambda}}\end{aligned}$$

其中 $\mathbf{a}_i^{\mathrm{T}}$ 为 $\mathbf{A}$ 的第 i 列向量