第 4 章对偶理论

SMaLL

[1] 中国石油大学（华东）
SMaLL 课题组
small.sem.upc.edu.cn
liangxijunsd@163.com

2023

对偶理论

拉格朗日函数

标准形式的优化问题 (不一定是凸问题)

$$
\begin{array}{lll}
\text{minimize} & f_0(x) & \\
\text{subject to} & f_i(x) \leq 0, & i = 1, \ldots, m \\
& h_i(x) = 0, & i = 1, \ldots, p
\end{array} \tag{1}
$$

变量 $x \in \mathbf{R}^n$, 定义域 $\mathcal{D}$, 最优值 $p^\star$

拉格朗日函数: $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$, 其中 $\mathrm{dom} L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

$$
L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \tag{2}
$$

- 目标函数和约束函数的加权和
- λ_i 是对应于不等式约束 $f_i(x) \leq 0$ 的 Lagrange 乘子
- ν_i 是对应于等式约束 $h_i(x) = 0$ 的 Lagrange 乘子

Lagrange 对偶函数

Lagrange 对偶函数: $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

$$g(\lambda, \boldsymbol{v}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\boldsymbol{x}, \lambda, \boldsymbol{v}) = \inf_{\mathbf{x} \in \mathcal{D}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_1 f_i(\boldsymbol{x}) + \sum_{i=1}^{p} v_i h_i(\boldsymbol{x}) \right) \quad (3)$$

g 是凹函数, 对一些 λ, ν 其取值可以为 $-\infty$

分析. g 是 (λ, ν) 的仿射函数族的逐点下确界 $\Rightarrow$ 它是凹的，即使原始问题不是凸的.

下界属性: 如果 $\lambda \succeq 0$, 那么 $g(\lambda, \nu) \leq p^\star$

Lagrange 对偶函数

证明: 如果 $\tilde{x}$ 是一个可行点, 即满足 $f_i(\tilde{x}) \leqslant 0$ $h_i(\tilde{x}) = 0$，并且 $\lambda \succeq 0$, 可得

$$\sum_i^m \lambda_i f_i(x) + \sum_i^p v_i h_i(x) \leqslant 0 \tag{4}$$

因为第一项求和中的每一项都是非正的，并且第二项求和中的每一项都为零,

$$L(\tilde{x}, \lambda, \mathbf{v}) = f_0(\tilde{\mathbf{x}}) + \sum_i^m \lambda_i f_i(\tilde{\mathbf{x}}) + \sum_i^p \nu_i h_i(\tilde{\mathbf{x}}) \leqslant f_0(\tilde{\mathbf{x}}) \tag{5}$$

因此

$$g(\boldsymbol{\lambda}, \mathbf{v}) = \inf_{x \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) \leqslant L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \mathbf{v}) \leqslant f_0(\tilde{\mathbf{x}}) \tag{6}$$

因为 $g(\lambda, v) \leqslant f_0(\tilde{x})$ 对所有的可行点 $\tilde{x}$ 都成立，故不等式 $g(\lambda, \nu) \leq p^\star$ 成立。

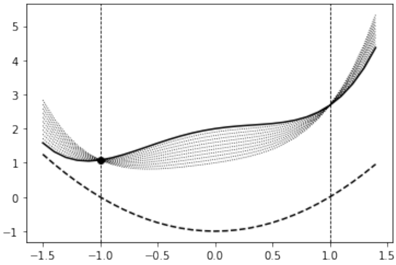

图: 1 一个对偶可行点的下界

图 1 用 $x \in \mathbf{R}$ 并且有一个不等式约束的简单问题解释了不等式 $g(\lambda, \nu) \leq p^\star$ 中的下界，实线为目标函数 f_0, 虚数为约束函数 f_1。可行集合为区间 [-1,1]。最优点和最优值为 $x^* = -1, p^* = 1.08$。
由于在可行集合上对 $\lambda \leqslant 0$ 有$L(x, \lambda) \leqslant f_0(x)$，因此每一个函数 $L(x, \lambda)$ 的最小值都比 p^* 小。

线性方程的最小范数解

$$
\begin{array}{ll}
\text{minimize} & x^T x \\
\text{subject to} & Ax = b
\end{array} \tag{7}
$$

对偶函数

- lagrange 函数为 $L(x,\nu) = x^T x + \nu^T(Ax - b)$
- 为了最小化 L，将梯度设置为零:

$$
\nabla_x L(x,\nu) = 2x + A^T\nu = 0 \quad \Longrightarrow \quad x = -(1/2)A^T\nu \tag{8}
$$

- 插入 L 得到 g :

$$
g(\nu) = L((-1/2)A^T\nu, \nu) = -\frac{1}{4}\nu^T AA^T\nu - b^T\nu \tag{9}
$$

它是一个关于 ν 的凹二次函数

下界性质: $p^\star \geq -(1/4)\nu^T AA^T\nu - b^T\nu$ 对于所有 ν

标准形式的 LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \succeq 0 \end{array} \tag{10}$$

对偶函数

- lagrange 函数为

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\ &= -b^T \nu + (c + A^T \nu - \lambda)^T x \end{aligned} \tag{11}$$

- L 是在 x 上的仿射函数, 因此

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases} \tag{12}$$

g 是一条直线，定义域为 $\{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\}$, 因此是凹的

下界性质: $p^\star \geq -b^T \nu$ if $A^T \nu + c \succeq 0$

等式约束范数最小化 (Ex.)

$$\begin{array}{ll} \text{minimize} & \|x\| \\ \text{subject to} & Ax = b \end{array} \tag{13}$$

对偶函数

$$g(\nu) = \inf_x \left(\|x\| - \nu^T Ax + b^T\nu\right) = \begin{cases} b^T\nu & \|A^T\nu\|_* \leq 1 \\ -\infty & 其他 \end{cases} \tag{14}$$

其中 $\|v\|_* = \sup_{\|u\|\leq 1} u^T v$ ， $\|\cdot\|$ 是对偶范数

证明:

- $\inf_x \left(\|x\| - y^T x\right) = \begin{cases} 0 & \|y\|_* \leq 1 \\ -\infty, & 否则 \end{cases}$
- 如果 $\|y\|_* \leq 1$, 那么 $\|x\| - y^T x \geq 0$ 对于任意 x, 如果 $x = 0$, 则相等
- 如果 $\|y\|_* > 1$, 令 $x = tu$ 其中 $\|u\| \leq 1, u^T y = \|y\|_* > 1$

$$\|x\| - y^T x = t\left(\|u\| - \|y\|_*\right) \to -\infty \quad 当 t \to \infty \tag{15}$$

下界性质: $p^\star \geq b^T\nu$ 如果 $\|A^T\nu\|_* \leq 1$

双向分区（Two-way partitioning）

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array} \tag{16}$$

- 非凸问题；可行集包含 2^n 个离散点
- 解释: 将 $\{1, \cdots, n\}$ 分成两个集合; W_{ij} 是将 i, j 分到相同集合的成本; $-W_{ij}$ 是分配到不同集合的成本

对偶函数

$$\begin{aligned} g(\nu) = \inf_x \left(x^T W x + \sum_i \nu_i (x_i^2 - 1) \right) &= \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu \\ &= \begin{cases} -\mathbf{1}^T \nu & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{其他} \end{cases} \end{aligned} \tag{17}$$

下界性质: $p^\star \geq -1^T \nu$ 如果 $W + \mathbf{diag}(\nu) \succeq 0$

例如：$\nu = -\lambda_{\min}(W)\mathbf{1}$ gives bound $p^\star \geq n\lambda_{\min}(W)$

Lagrange 对偶和共轭函数

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & Ax \preceq b, \quad Cx = d
\end{array} \tag{18}
$$

对偶函数

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_{x \in \mathbf{dom} f_0} \left(f_0(x) + \left(A^T\lambda + C^T\nu\right)^T x - b^T\lambda - d^T\nu \right) \\
&= -f_0^*\left(-A^T\lambda - C^T\nu\right) - b^T\lambda - d^T\nu
\end{aligned} \tag{19}
$$

- 共轭的定义：$f^*(y) = \sup_{x \in \mathbf{dom} f}\left(y^T x - f(x)\right)$
- 如果已知 f_0 的共轭，则简化对偶的推导

例 (熵最大化)

$$
f_0(x) = \sum_{i=1}^n x_i \log x_i, \quad f_0^*(y) = \sum_{i=1}^n e^{y_i - 1} \tag{20}
$$

对偶问题

Lagrange 对偶问题

$$\max_{\lambda \succeq 0} \quad g(\lambda, \nu) = \max_{\lambda \succeq 0} \quad \min_x L(x, \lambda, \nu) \tag{21}$$

- Lagrange 对偶函数: 找到 $p^\star$ 上的最优下界
- 对偶问题为凸优化问题; 最优值记为 $d^\star$
- 如果 $\lambda \succeq 0$, 则 $(\lambda, \nu) \in \mathrm{dom}\, g$ ，称 λ, ν 是对偶可行的

对偶问题

例 (标准形式 LP 及其对偶)

$$
\begin{array}{ll}
\text{minimize} & c^T x \\
\text{subject to} & Ax = b \\
 & x \succeq 0
\end{array}
\qquad (22)
\qquad\qquad
\begin{array}{ll}
\text{minimize} & -b^T \nu \\
\text{subject to} & A^T \nu + c \succeq 0
\end{array}
\qquad (23)
$$

- 假设 $A \in \mathbb{R}^{m \times n}$, $m << n$
- 原始问题: $x \in \mathbb{R}^n$, 对偶问题 $\nu \in \mathbb{R}^m$;
- $m << n \rightarrow$ 对偶规划显著减少了变量个数
- 何时考虑对偶规划（来减少变量）？
  较少的等式约束：许多变量 $\rightarrow$ 对偶问题

对偶理论

弱对偶和强对偶

弱对偶： $d^\star \leq p^\star$

- 总是成立（对于凸和非凸问题）
- 可以用来寻找难以解决的问题的最优值下界
  例如，解决 SDP（半正定规划）问题

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T\nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array} \tag{24}$$

  给出了一个关于双向分配问题的下界

强对偶： $d^* = p^*$

- 一般不成立
- （通常）对于凸问题成立
- 在凸问题中保证强对偶的条件称为 **约束规范**

Slater 约束条件

Proposition

- 以下凸问题

$$\begin{array}{ll} \textit{minimize} & f_0(x) \\ \textit{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array} \tag{25}$$

-

$$\exists x \in \ \textbf{\textit{int}}\ \mathcal{D}: \quad f_i(x) < 0, \quad i = 1, \ldots, m, \quad Ax = b \tag{26}$$

$\mathcal{D} = \cap_{i=0}^{m} \textbf{\textit{dom}}\, f_i(x)$

$\Longrightarrow$ 强对偶性成立.

注意

- 也保证了达到对偶问题的最优解 (如果 $d^* > -\infty$)
- 仿射不等式不必与严格不等式同时成立: 如果 $f_1, \cdots, f_k$ 是仿射的, 则在下面条件成立时，强对偶性成立:

$$\exists x \in \ \textbf{relint}\ \mathcal{D}: f_i(x) \le 0,\ i = 1, \cdots, m,\ f_i(x) < 0,\ i = k+1, \ldots, m,\ Ax = b$$

不等式形式的 LP 问题

原问题

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array} \tag{27}$$

对偶函数

$$g(\lambda) = \inf_x \left((c + A^T\lambda)^T x - b^T\lambda \right) = \begin{cases} -b^T\lambda & A^T\lambda + c = 0 \\ -\infty & \text{其他} \end{cases} \tag{28}$$

对偶问题

$$\begin{array}{ll} \text{maximize} & -b^T\lambda \\ \text{subject to} & A^T\lambda + c = 0, \quad \lambda \succeq 0 \end{array} \tag{29}$$

- 由 Slater 条件得: 如果对某些 $\tilde{x}$, $A\tilde{x} \prec b$ ，则 $p^* = d^*$
- 事实上, $p^* = d^*$ 除非原问题和对偶问题不可行

二次规划问题 (Ex.)

原问题 (假设 $P \in \mathbf{S}^n_{++}$)

$$
\begin{array}{ll}
\text{minimize} & x^T P x \\
\text{subject to} & Ax \preceq b
\end{array} \tag{30}
$$

对偶函数

$$
g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b)\right) = -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda \tag{31}
$$

对偶问题

$$
\begin{array}{ll}
\text{maximize} & -(1/4)\lambda^T A P^{-1} A^T \lambda - b^T \lambda \\
\text{subject to} & \lambda \succeq 0
\end{array} \tag{32}
$$

- 由 Slatera 条件: 如果对某些 $\tilde{x}$, $A\tilde{x} \prec b$ ，则 $p^* = d^*$
- 事实上, $p^* = d^*$ 总是成立

对偶理论

互补松弛度

假设强对偶成立, $x^\star$ 是原问题的最优解, $(\lambda^\star, \nu^\star)$ 是对偶问题的最优解

$$
\begin{aligned}
f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\
&\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\
&\leq f_0(x^\star)
\end{aligned} \tag{33}
$$

因此, 这两个不等式取等号

- $x^\star$ 最小化 $L(x, \lambda^\star, \nu^\star)$
- $\lambda_i^\star f_i(x^\star) = 0 \;,\; i = 1, \ldots, m$ (称为互补松弛):

$$
\lambda_i^\star > 0 \Longrightarrow f_i(x^\star) = 0, \quad f_i(x^\star) < 0 \Longrightarrow \lambda_i^\star = 0 \tag{34}
$$

- x^* 是 $L(x, \lambda^*, \nu^*)$ 的最小值点 $\Rightarrow \nabla_x L(x, \lambda^*, \nu^*)|_{x=x^*} = 0$

Karush-Kuhn-Tucker (KKT) 条件 **

以下四个条件称为 KKT 条件 (对于可微的 f_i, h_i):

1. 原始约束: $f_i(x) \leq 0, i = 1, \ldots, m,\ h_i(x) = 0, i = 1, \ldots, p$
2. 对偶约束: $\lambda \succeq 0$
3. 互补松弛条件: $\lambda_i f_i(x) = 0, i = 1, \ldots, m$
4. Lagrange 函数在点 x 处的梯度为 0 :

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0 \tag{35}$$

Proposition

如果强对偶成立，且 x, λ, ν 是原问题和对偶问题的最优解 $\Longrightarrow$ *KKT* 条件成立.

凸问题的 KKT 条件

定理

- 如果原问题是凸问题: f_i 是凸函数, h_i 是仿射函数
- 如果 $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ 满足 KKT 条件

$\Rightarrow$ $\tilde{x}$, $(\tilde{\lambda}, \tilde{\nu})$ 是零对偶间隙的原始最优解和对偶最优解.

- 注意：对于任意可微目标和约束函数的凸 优化问题, 满足 KKT 条件的任何点是原始和对偶最优，并且具有零对偶间隙。

证明:

- $\tilde{x}$: 原始可行解, $(\tilde{\lambda}, \tilde{\nu})$ 对偶可行解;
- $\tilde{\lambda}_i \geq 0 \Rightarrow L(x, \tilde{\lambda}, \tilde{\nu})$ 在 x 处是凸的;
- 梯度条件: $\nabla_x L(x, \tilde{\lambda}, \tilde{\nu})|_{x=\tilde{x}} = 0 \Rightarrow \tilde{x} = \mathrm{argmin}_x L(x, \tilde{\lambda}, \tilde{\nu})$
- $$\begin{aligned} g(\tilde{\lambda}, \tilde{\nu}) &= L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \\ &= f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{x}) \\ &= f_0(\tilde{x}) \end{aligned}$$

应用: 支持向量机

- 二分类问题
- 训练集: $\{x_i, y_i\}_{i=1}^n$, $x_i \in \mathbb{R}^d$, $y_i \in \{1, -1\}$
- 线性分类器: $y = \langle w, x \rangle + b$
- 支持向量机分类

$$\begin{array}{ll} \min_{w,b} & \frac{1}{2}\|w\|^2 \\ s.t. & y_i(\langle w, x_i \rangle + b) \geq 1, i = 1 \cdots, n \end{array}$$

所有样本均正确分类.

- 验证其 Lagrange 对偶问题:

$$\begin{array}{ll} \max_{\alpha} & -\frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle + \langle \mathbf{1}, \alpha \rangle \\ s.t. & \langle \alpha, y \rangle = 0 \\ & \alpha_i \geq 0, \quad i = 1, \ldots, n \end{array}$$

应用: 支持向量机 (Ex.)

- 软间隔 (Soft-margin) 支持向量机分类

$$
\begin{array}{cl}
\min_{w,b} & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^n \xi_i \\
s.t. & y_i(\langle w, x_i\rangle + b) \geq 1 - \xi_i, i = 1\cdots, n
\end{array}
$$

允许一些分类错误的样本.

- 验证其 Lagrange 对偶问题:

$$
\begin{array}{cl}
\max_{\alpha} & -\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle + \langle \mathbf{1}, \alpha\rangle \\
s.t. & \langle \alpha, y\rangle = 0 \\
 & C \geq \alpha_i \geq 0, \quad i = 1, \ldots, n
\end{array}
$$

作业 I

1. 考虑如下优化问题

$$\begin{array}{ll} \text{minimize} & x^2 - 2x + 1 \\ \text{subject to} & (x-1)(x-5) \leqslant 0 \end{array} \tag{36}$$

其中 $\mathbf{x} \in \mathbf{R}^n$

（1）分析原问题，给出可行集、最优值以及最优解

（2）画出目标函数关于 x 的图像，并在图像上画出可行集、最优值以及最优解。画出 λ=0.2 时，Lagrange 函数 $L(x, \lambda)$ 关于 x 的图像，验证下界性质：$p^* \geqslant \inf_x L(x, \lambda)$ 对 $\lambda \geqslant 0$ 成立。推导并画出 Lagrange 对偶函数 g.

作业 II

2. 考虑如下凸分段线性函数最小化问题

$$\text{minimize} \quad \max_{i=1,\ldots,m}(a_i^{\mathrm{T}}\mathbf{x} + b_i) \tag{37}$$

其中 $\mathbf{x} \in \mathbf{R}^n$

(1) 考虑等式问题

$$\begin{aligned} \text{minimize} \quad & \max_{i=1,\cdots,m} y_i \\ \text{subject to} \quad & \mathbf{a}_i^{\mathrm{T}}\mathbf{x} + b_i = y_i, i = 1, \cdots, m \end{aligned} \tag{38}$$

其中 $\mathbf{x} \in \mathbf{R}^n$， $\mathbf{y} \in \mathbf{R}^m$, 推导出 Lagrange 对偶问题。

(2) 将分段线性问题（34）表述为 LP 问题并写出 LP 问题的对偶问题。将 LP 对偶与 (1) 中所求得的对偶问题联系起来。

作业 III

3. 设有若干二分类问题的观测样本 $\{(x_i, y_i)\}_{i=1}^{\mathrm{n}} \subseteq \mathbf{R}^d \times \{+1, -1\}$, 考虑线性支持向量机模型:

$$min_{w\in\mathbf{R}^4, b\in\mathbf{R}} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n} l(x_i, y_i; \mathbf{w}, b) \tag{39}$$

c > 0 为参数, $l(x_i, y_i; \mathbf{w}, b) = \max(0, 1 - y_i(\mathbf{w}^\top x_i + b))$ 为样本点 (x_i, y_i) 的经验损失, 对应 Hinge 损失函数 $l(t) = \max(0, 1 - t)$, 函数 $y(x) = w^\top x + b$ 为判别函数。

（1）将线性支持向量机模型化为凸二次规划模型, 并写出该规划模型的对偶问题。

（2）结合凸二次规划模型的最优化条件, 说明哪些样本对模型起作用, 哪些不起作用。

作业 IV

4. 通过引入新变量 $y_i \in \mathbf{R}^{m_i}$ 以及等式约束 $y_i = A_i x + b_i$, 推导出下式的对偶问题

$$\text{minimize} \quad \sum_{i=1}^{N} \|A_i x + b_i\|_2 \tag{40}$$

其中 $A_i \in \mathbf{R}^{m_i \times n}$,$b_i \in \mathbf{R}^{m_i}$

5. 考虑 Lasso 模型

$$\text{minimize} \quad \|Ax + b\|^2 + \lambda \cdot |x|_1 \tag{41}$$

其中 $A \in \mathbf{R}^{m \times n}$, $\boldsymbol{b} \in \mathbf{R}^{m}$, $\lambda > 0$, 假设模型有唯一解，证明该模型的最优解关于 λ 是分片线性的。

作业 V

6. 考虑 QCQP 问题

$$\begin{array}{ll}\text{minimize} & \frac{1}{2}(x_1^2+x_2^2)-4 \\ \text{subject to} & (x_1-1)^2+(x_2-1)^2 \leqslant 4\end{array} \tag{42}$$

其中变量 $x \in \mathbb{R}^2$

(1) 写出最优点 x^* 和最优值 p^*。

(2) 写出 KKT 条件。

(3) 求解 Lagrange 对偶问题并说明强对偶性性是否成立。

7. 推导出下面问题的 KKT 条件

$$\begin{array}{ll}\text{minimize} & \operatorname{tr} X - \log\det X \\ \text{subject to} & X\boldsymbol{s} = \boldsymbol{y}\end{array} \tag{43}$$

其中 $X \in \mathbf{S}^n$, 定义域 $\mathbf{s}^n_{++}$。已知 $y \in \mathbf{R}^n$, $s \in \mathbf{R}^n$ 且 $s^{\mathrm{T}} y = 1$, 证明最优解为

$$X^* = I + yy^{\mathrm{T}} - \frac{1}{s^{\mathrm{T}} s} ss^{\mathrm{T}} \tag{44}$$