DCS440　最优化理论

第二章：对偶理论

杨 磊
yanglei39@mail.sysu.edu.cn

计算机学院，2024 秋

§ Lagrange 函数与对偶函数

§ Lagrange 对偶问题

§ Lagrange 对偶的几种解释

§ 最优性条件

§ 敏感性分析（Perturbation Analysis）

§ 利用对偶理解“启发式”方法

一般优化问题

考虑一般优化问题（可能非凸）

$$
\begin{aligned}
\min \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \leqslant 0, \quad i = 1, \cdots, m, \\
& h_j(x) = 0, \quad j = 1, \cdots, p.
\end{aligned}
$$

对于上面这个一般形式的优化问题，假设：

- 问题的定义域 $\mathcal{D} := \bigcap_{i=0}^{m} \operatorname{dom} f_i \cap \bigcap_{j=1}^{p} \operatorname{dom} h_j$ 非空；
- 问题的最优解 x^* 和 最优值 p^* 存在。

问题：如何研究上述形式优化问题的最优性条件？

一个思路：利用**拉格朗日(Lagrange)对偶**！粗略地说，其基本思想：将原约束优化问题的目标和约束放在同一个函数（即拉格朗日函数）中来研究。

拉格朗日(Lagrange)函数

为此，我们先引入拉格朗日(Lagrange)函数的概念。具体来说，对于上一页的一般优化问题，它的 Lagrange 函数 $L: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ 定义为

$$L(x, \lambda, \nu) := f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x),$$

其中：

- $L(x, \lambda, \nu)$ 的定义域为 $\operatorname{dom} L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$；
- λ_i 称为第 i 个不等式约束 $f_i(x) \leqslant 0$ 对应的 Lagrange 乘子；
- ν_j 称为第 j 个等式约束 $h_j(x) = 0$ 对应的 Lagrange 乘子；
- 向量 λ 和 ν 是分别由 λ_i 和 ν_j 组成的向量，称作优化问题的对偶变量 (dual variable) 或者 Lagrange 乘子；
- 此时，将决策变量 x 称作原变量 (primal variable)。

定义 Lagrange 对偶函数（简称对偶函数，dual function）如下：

$$\begin{aligned} g(\lambda,\nu) &:= \inf_{x\in\mathcal{D}} L(x,\lambda,\nu) \\ &= \inf_{x\in\mathcal{D}} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x) \right\}. \end{aligned}$$

性质一：$g(\lambda,\nu)$ 是关于 λ 和 ν 的凹函数！ （思考：为什么？）

答：$g(\lambda,\nu)$ 是一族关于 (λ,ν) 的仿射函数的逐点下确界，因此即使优化问题非凸，其对偶函数 $g(\lambda,\nu)$ 也是凹函数。

性质二：对 $\forall\lambda \geq 0$ 和 $\forall\nu$，可以推出 $g(\lambda,\nu) \leq p^*$（见下一页），即 $g(\lambda,\nu)$ 给出了原问题最优值的下界。

启发：可通过计算以下优化问题（称作对偶问题）得到原约束优化问题最优值 p^* 的**最紧下界**：

$$\sup_{\lambda,\nu} \quad g(\lambda,\nu) \quad \text{s.t.} \quad \lambda \geq 0.$$

定理 1 (弱对偶定理)

对 $\forall\lambda \geq 0$ 和 $\forall\nu$，有 $g(\lambda,\nu) \leq p^*$。

证明： 设 $x^* \in \mathcal{D}$ 是原问题的最优解，则有 $f_i(x^*) \leq 0$, $h_j(x^*) = 0$（即满足可行性）。于是，对 $\forall\lambda \geq 0$ 和 $\forall\nu$，有

$$\underbrace{\sum_{i=1}^{m} \lambda_i f_i(x^*)}_{\leq 0} + \underbrace{\sum_{j=1}^{p} \nu_j h_j(x^*)}_{=0} \leq 0.$$

因此

$$L(x^*,\lambda,\nu) = f_0(x^*) + \sum_{i=1}^{m} \lambda_i f_i(x^*) + \sum_{j=1}^{p} \nu_j h_j(x^*) \leqslant f_0(x^*) = p^*.$$

进一步，有

$$g(\lambda,\nu) = \inf_{x\in\mathcal{D}} L(x,\lambda,\nu) \leqslant L(x^*,\lambda,\nu) \leqslant f_0(x^*) = p^*.$$

对偶函数计算：线性规划

线性规划问题（标准型）

$$
\begin{aligned}
\min_{x} \quad & c^\top x \\
\text{s.t.} \quad & Ax = b \qquad \Rightarrow \quad Ax - b = 0 \\
& x \geq 0 \qquad\; \Rightarrow \quad -x \leq 0
\end{aligned}
$$

假设可行域非空。对不等式约束 $-x \leq 0$ 和等式约束 $Ax - b = 0$ 分别引入 Lagrange 乘子 λ 和 ν，写出 Lagrange 函数：

$$L(x, \lambda, \nu) = c^\top x + \nu^\top (Ax - b) + \lambda^\top (-x) = -b^\top \nu + (A^\top \nu - \lambda + c)^\top x.$$

于是，对偶函数为

$$g(\lambda, \nu) = \inf_{x} \; L(x, \lambda, \nu) = -b^\top \nu + \inf_{x} \left\{ (A^\top \nu - \lambda + c)^\top x \right\}$$

对偶函数计算：线性规划

容易分析得到

$$g(\lambda,\nu) = \inf_x \ L(x,\lambda,\nu) = -b^\top \nu + \inf_x \left\{ (A^\top \nu - \lambda + c)^\top x \right\}$$

$$= \begin{cases} -b^\top \nu, & A^\top \nu - \lambda + c = 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

于是，由弱对偶定理可知

$$p^* \text{ 的下界是：} \begin{cases} -b^\top \nu, & A^\top \nu - \lambda + c = 0, \ \ \lambda \geq 0, \\ -\infty, & \text{otherwise.} \quad (\text{显然，该下界无意义}) \end{cases}$$

相应的对偶问题如下：

$$\sup_{\lambda,\nu} \quad -b^\top \nu \quad \text{s.t.} \quad A^\top \nu - \lambda + c = 0, \ \ \lambda \geq 0.$$

上述问题进一步等价于

$$\sup_{\nu} \quad -b^\top \nu \quad \text{s.t.} \quad A^\top \nu + c \geq 0.$$

线性约束的最小二乘解问题

$$\begin{aligned} \min \quad & x^\top x \\ \text{s.t.} \quad & Ax = b \qquad \Rightarrow \quad Ax - b = 0 \end{aligned}$$

为等式约束 $Ax - b = 0$ 引入 Lagrange 乘子 ν，写出 Lagrange 函数

$$L(x, \nu) = x^\top x + \nu^\top (Ax - b).$$

于是，其对偶函数为

$$g(\nu) = \inf_x \; L(x, \nu)$$

这是一个无约束凸优化问题，且 $L(x, \nu)$ 关于 x 是连续可微的，因此关于 x 的最优解可通过一阶最优性条件得到

$$\nabla_x L(x, \nu) = 2x + A^\top \nu = 0 \quad \Rightarrow \quad x_\nu^* = -\frac{1}{2} A^\top \nu.$$

需注意：这里的最优解 x_ν^* 依赖于乘子 ν!

因此，对偶函数为

$$g(\nu) = \inf_x L(x, \nu) = L(x_\nu^*, \nu) = L\left(-\frac{1}{2}A^\top\nu, \nu\right) = -\frac{1}{4}\nu^\top AA^\top\nu - b^\top\nu.$$

该对偶函数显然是凹函数

进一步，由弱对偶定理可知：

$$1.\quad \inf_{x\in\mathbb{R}^n}\{x^\top x \mid Ax = b\} \geqslant -\frac{1}{4}\nu^\top AA^\top\nu - b^\top\nu, \quad \forall\nu$$

$$2.\quad \inf_{x\in\mathbb{R}^n}\{x^\top x \mid Ax = b\} \geqslant \underbrace{\sup_\nu\left\{-\frac{1}{4}\nu^\top AA^\top\nu - b^\top\nu\right\}}_{\text{对偶问题}}$$

关于第 2 点，我们将在之后说明，当 $Ax = b$ 解集非空时，不等式取等号（此时，强对偶性成立）。

双向划分问题

$$
\begin{aligned}
\min_{x} \quad & x^\top W x \\
\text{s.t.} \quad & x_i^2 = 1, \quad i = 1, 2, \cdots, n,
\end{aligned}
$$

其中 $W \in S^n$，$x \in \mathbb{R}^n$，且其分量 x_i 取值为 1 或 -1。

为等式约束 $x_i^2 - 1 = 0$ 引入拉格朗日乘子 ν_i，写出 Lagrange 函数：

$$
L(x, \nu) = x^\top W x + \sum_{i=1}^{n} \nu_i \big(x_i^2 - 1\big) = x^\top \big(W + \mathrm{diag}(\nu)\big) x - \mathbf{1}^\top \nu,
$$

其中 $\mathbf{1} := (1, 1, \cdots, 1)^\top$，$\mathrm{diag}(\nu)$ 表示以向量 ν 为对角元素的矩阵。

进一步，对 x 求极小，得到对偶函数：

$$
g(\nu) = \inf_{x} \left\{ x^\top (W + \mathrm{diag}(\nu)) x \right\} - \mathbf{1}^\top \nu = \begin{cases} -\mathbf{1}^\top \nu, & W + \mathrm{diag}(\nu) \succeq 0, \\ -\infty, & \text{otherwise}. \end{cases}
$$

Lagrange 对偶与共轭函数

仿射约束的优化问题

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & Ax \le b \\ & Cx = d \end{aligned}$$

回顾凸函数 $f:\mathbb{R}^n \to \mathbb{R}$ 的共轭函数：$f^*(y) = \sup_{x\in \mathrm{dom}\, f} \left\{ y^\top x - f(x) \right\}$

下面，利用函数 f_0 的共轭函数 f_0^* 表述问题的对偶函数：

$$\begin{aligned} g(\lambda,\nu) = \inf_{x\in\mathrm{dom}\, f_0} L(x,\lambda,\nu) &= \inf_{x\in\mathrm{dom}\, f_0} \left\{ f_0(x) + \lambda^\top (Ax-b) + \nu^\top (Cx-d) \right\} \\ &= \inf_{x\in\mathrm{dom}\, f_0} \left\{ f_0(x) + \left(A^\top\lambda + C^\top\nu\right)^\top x \right\} - \lambda^\top b - \nu^\top d \\ &= -f_0^*\left(-A^\top\lambda - C^\top\nu\right) - \lambda^\top b - \nu^\top d \end{aligned}$$

对偶函数计算：最大熵问题

最大熵问题

$$\begin{array}{ll} \min & f_0(x) = \sum_{i=1}^n x_i \log x_i \\ \text{s.t.} & Ax \le b \\ & \mathbf{1}^\top x = 1 \end{array}$$

对于 f_0，其定义域为 $\operatorname{dom} f_0 = \mathbb{R}_+^n$；

f_0 的共轭函数为 $f_0^*(y) = \displaystyle\sum_{i=1}^n e^{y_i - 1}$，其定义域为 $\operatorname{dom} f_0^* = \mathbb{R}^n$ 。

于是，最大熵问题的对偶函数为

$$\begin{aligned} g(\lambda, \nu) &= -f_0^*\left(-A^\top \lambda - \nu \mathbf{1}\right) - \lambda^\top b - \nu \\ &= -\lambda^\top b - \nu - \sum_{i=1}^n e^{-a_i^\top \lambda - \nu - 1} = -\lambda^\top b - \nu - e^{-\nu - 1} \sum_{i=1}^n e^{-a_i^\top \lambda}, \end{aligned}$$

其中 a_i 表示矩阵 A 的第 i 列向量。

- § Lagrange 函数与对偶函数
- § Lagrange 对偶问题
- § Lagrange 对偶的几种解释
- § 最优性条件
- § 敏感性分析（Perturbation Analysis）
- § 利用对偶理解“启发式”方法

Lagrange 对偶问题

回顾**弱对偶定理**：对于 Lagrange 对偶函数 $g(\lambda,\nu)$，对任意 $\forall\lambda \geq 0$ 和 $\forall\nu$，$g(\lambda,\nu)$ 给出了原优化问题最优值 p^* 的一个下界。

显然，该下界和乘子 λ 和 ν 的选取相关。而当我们极大化对偶函数 $g(\lambda,\nu)$，试图计算 p^* 的最紧下界时，就引出了原问题的 **Lagrange 对偶问题**：

Lagrange 对偶问题

$$
\begin{aligned}
\inf / \min \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \leq 0, \qquad \text{(Primal)} \\
& h_j(x) = 0.
\end{aligned}
\qquad\qquad
\begin{aligned}
\sup / \max \quad & g(\lambda,\nu) \\
\text{s.t.} \quad & \lambda \geq 0
\end{aligned}
\qquad \text{(Dual)}
$$

- 左侧为原问题(Primal)，右侧为 Lagrange 对偶问题(Dual)
- 原问题(Primal)的最优值为 $p^* \quad \geq \quad$ 对偶问题(Dual)的最优值为 d^*
- 对偶问题(Dual)的最优解 (λ^*, ν^*) 称为对偶最优解或最优 Lagrange 乘子
- 无论原问题(Primal)是否为凸，对偶问题(Dual) 总是一个**凹优化问题**
- 为方便，后面出现的对偶问题无论解是否存在，我们都用符号 max

例：线性规划的对偶问题

考虑线性规划问题

$$\begin{aligned} \min_{x} \quad & c^\top x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{Primal}$$

已推导其 Lagrange 对偶函数为

$$g(\lambda, \nu) = \begin{cases} -b^\top \nu, & A^\top \nu - \lambda + c = 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

进一步，写出 Lagrange 对偶问题：

$$\begin{aligned} \max_{\lambda, \nu} \quad & g(\lambda, \nu) = -b^\top \nu \\ \text{s.t.} \quad & A^\top \nu - \lambda + c = 0 \\ & \lambda \geq 0 \end{aligned} \tag{Dual}$$

例：线性规划的对偶问题（续）

解析： 已推导其 Lagrange 对偶函数为

$$g(\lambda, \nu) = \begin{cases} -b^\top \nu, & A^\top \nu - \lambda + c = 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

- 当 $g(\lambda, \nu) = -\infty$ 时，极大化它是没有意义的；
- 因此对偶问题事实上具有隐含的约束条件 $A^\top \nu - \lambda + c = 0$；

$$\begin{aligned} \max_{\lambda, \nu} \quad & g(\lambda, \nu) = -b^\top \nu \\ \text{s.t.} \quad & A^\top \nu - \lambda + c = 0 \\ & \lambda \geq 0 \end{aligned} \tag{Dual}$$

- 进一步地，乘子 λ 可以被隐去：

$$\begin{aligned} \max_{\nu} \quad & g(\lambda, \nu) = -b^\top \nu \\ \text{s.t.} \quad & A^\top \nu + c \geq 0 \quad (A^\top \nu + c = \lambda \geq 0) \end{aligned} \tag{Dual}$$

例：线性约束的最小二乘

回顾线性约束的最小二乘问题

$$\begin{aligned} \min \quad & x^\top x \\ \text{s.t.} \quad & Ax = b \end{aligned} \qquad \text{(Primal)}$$

已推导其 Lagrange 对偶函数为

$$g(\nu) = L\left(-\frac{1}{2}A^\top \nu, \nu\right) = -\frac{1}{4}\nu^\top AA^\top \nu - b^\top \nu.$$

进一步，写出 Lagrange 对偶问题：

$$\begin{aligned} & \max_{\nu} \quad g(\nu) = -\frac{1}{4}\nu^\top AA^\top \nu - b^\top \nu \\ \Longleftrightarrow \quad & \min_{\nu} \quad \frac{1}{4}\nu^\top AA^\top \nu + b^\top \nu \end{aligned} \qquad \text{(Dual)}$$

设原问题(Primal)的最优值为 $p^* = f_0(x^*)$，对偶问题(Dual)的最优值为 $d^* = g(\lambda^*, \nu^*)$。于是，

- 不等式 $d^* \le p^*$ 总是成立的，称为弱对偶性 (weak duality)
- 等式 $d^* = p^*$ 不必然成立！当等式成立时，称为强对偶性(strong duality)[1]
- 差值 $p^* - d^*$ 称为对偶间隙 (duality gap)，根据弱对偶性可知对偶间隙总是非负的

思考：什么情况下强对偶性成立，即对偶间隙为 0?

[1]注意：这里的 p^* 可以为 $-\infty$，此时由弱对偶性可知 $d^* = -\infty$，因此强对偶性成立。为方便，我们后面都假设 p^* 为有限数。

对偶性的强弱（续）

显然，强对偶性是很好的性质。如果成立，则可以通过求解对偶问题来求解原问题的最优值。遗憾的是，**一般情况下强对偶性并不成立**。

但是，对于凸优化问题，在一定（不是特别强）的条件下，强对偶性是成立的，这也说明了凸优化问题的优势。

考虑一般形式的凸优化问题

$$
\begin{aligned}
\min \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \leqslant 0, \quad i = 1, \cdots, m \quad (f_0, \cdots, f_m \text{ 凸}) \\
& Ax = b.
\end{aligned}
$$

针对上述问题，文献中已经给出了很多强对偶性成立的条件，这些条件往往被称为约束规范条件 (**Constraint Qualification, CQ**)。其中一个相对简单且被广泛应用的条件是 Slater’s condition。

集合的相对内点集合

介绍 Slater's condition 之前，首先介绍集合 $\mathcal{D}$ 的相对内点集合 relint $\mathcal{D}$。

给定集合 $\mathcal{D}$，回顾其仿射包的定义：

$$\operatorname{aff}\mathcal{D} := \left\{ x \;\middle|\; x = \sum_{i=1}^{k}\theta_i x_i,\ \forall\, x_1, x_2, \cdots, x_k \in \mathcal{D},\ \sum_{i=1}^{k}\theta_i = 1 \right\}.$$

回顾其内点集合的定义：（$B(x,r)$ 表示以 x 为中心，半径为 r 的范数球）

$$\operatorname{int}\mathcal{D} := \left\{ x \in \mathcal{D} \mid \exists\, r > 0,\ \text{ s.t. } B(x,r) \subseteq \mathcal{D} \right\}.$$

思考：内点集合是否总是存在？

答：不一定存在！例：$\{(x_1, x_2)\ :\ 0 \le x_1 \le 1,\ x_2 = 0\}$

集合的相对内点集合（续）

定义 1 (相对内点 (relative interior) 集合)

集合 $\mathcal{D}$ 的相对内点集合定义为

$$\operatorname{relint}\mathcal{D} := \left\{ x \in \mathcal{D} \mid \exists\, r > 0,\ \text{ s.t. }\ B(x,r) \cap \operatorname{aff}\mathcal{D} \subseteq \mathcal{D} \right\}.$$

解析：不严谨地说，相对内点可以看作内点的推广！

- **当** $\operatorname{aff}\mathcal{D} = \mathbb{R}^n$ **时**，相对内点等价于内点；
- 当 $\mathcal{D}$ 本身的“维度”较低（例如，$\mathcal{D}$ 为 $\mathbb{R}^n$ 中的某个超平面）时，$\mathcal{D}$ 中不存在内点，因此需引入“相对”的概念，以适应更复杂的情况。

Slater's condition

Slater's condition

若存在一点 $x \in \operatorname{relint} \mathcal{D}$，使得凸优化问题的约束满足：

$$f_i(x) < 0, \quad i = 1, \cdots, m, \quad Ax = b.$$

则称：此问题满足 Slater's condition。进一步，此问题的强对偶性成立。

- 满足上述条件的点也称为**严格可行点**，因为不等式约束严格成立
- Slater's condition 要求 $\operatorname{relint} \mathcal{D}$ 中存在使得不等式约束严格成立的点

当不等式约束函数中有部分是仿射的时，Slater's condition 可弱化如下：

弱化的 Slater's condition

不失一般性，设前 k 个不等式约束函数 $f_1, \cdots, f_k$ 是仿射的，若存在一点 $x \in \operatorname{relint} \mathcal{D}$，使得凸优化问题的约束满足：

$$\underline{f_i(x) \le 0}, \quad i = 1, \cdots, k, \quad f_i(x) < 0, \quad i = k+1, \cdots, m, \quad Ax = b.$$

则此问题的强对偶性成立。由此可看出，仿射不等式不需要严格成立。

例：线性约束的最小二乘问题

线性约束的最小二乘问题

$$\begin{aligned} \min \quad & x^\top x \\ \text{s.t.} \quad & Ax = b \end{aligned} \qquad \text{(Primal)}$$

已推导其 Lagrange 对偶问题为：

$$\begin{aligned} & \max_{\nu} \quad g(\nu) = -\frac{1}{4}\nu^\top AA^\top \nu - b^\top \nu \\ \Longleftrightarrow \quad & \min_{\nu} \quad \frac{1}{4}\nu^\top AA^\top \nu + b^\top \nu \end{aligned} \qquad \text{(Dual)}$$

于是，对于该问题，只要原问题可行，即存在 x 满足 $Ax = b$，则必有强对偶性成立。

例：线性规划

考虑线性规划问题

$$\begin{aligned}\min_{x} \quad & c^\top x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0\end{aligned} \qquad \text{(Primal)}$$

已推导其 Lagrange 对偶问题：

$$\max_{\lambda, \nu} \quad -b^\top \nu \quad \text{s.t.} \quad A^\top \nu - \lambda + c = 0, \quad \lambda \geq 0. \qquad \text{(Dual)}$$

根据弱化的 Slater's condition，对于线性规划问题：

- 只要原问题(Primal)可行，则强对偶性成立；
- 由于对偶也是线性规划，因此如果对偶(Dual)可行，则强对偶性成立；
- 只有一种情况下强对偶性不成立： (Primal)和(Dual)均不可行。

★ 特别强调 ★：针对线性规划问题，这里的强对偶性（对偶间隙为零，包含 $p^* = d^* = -\infty$ 的情况）与运筹学教材中的强对偶定理有所不同!

例：最大熵问题

最大熵问题

$$
\begin{aligned}
\min \quad & f_0(x) = \sum_{i=1}^{n} x_i \log x_i \qquad (\mathrm{dom}\, f_0 = \mathbb{R}^n_{++}) \\
\text{s.t.} \quad & Ax \le b \\
& \mathbf{1}^\top x = 1
\end{aligned}
$$

根据之前推导的 Lagrange 对偶函数，直接写出对偶问题：

$$
\begin{aligned}
\max \quad & -b^\top \lambda - \nu - e^{-\nu-1} \sum_{i=1}^{n} e^{-a_i^\top \lambda} \\
\text{s.t.} \quad & \lambda \ge 0
\end{aligned}
$$

于是，由弱化的 Slater's condition，若 $\exists x > 0$ 使得 $Ax \le b$, $\mathbf{1}^\top x = 1$，则强对偶性成立。

例：QCQP 问题

二次约束二次规划问题(QCQP)

$$
\begin{aligned}
\min_{x} \quad & \frac{1}{2} x^\top P_0 x + q_0^\top x + r_0, & {\color{red} P_0 \in S_{++}^n}, \\
\text{s.t.} \quad & \frac{1}{2} x^\top P_i x + q_i^\top x + r_i \leqslant 0, & P_i \in S_+^n, \quad i = 1, 2, \cdots, m,
\end{aligned}
\qquad \text{(Primal)}
$$

引入 Lagrange 乘子向量 λ，写出其 Lagrange 函数为

$$
\begin{aligned}
L(x, \lambda) &= \frac{1}{2} x^\top P_0 x + q_0^\top x + r_0 + \sum_{i=1}^m \lambda_i \left(\frac{1}{2} x^\top P_i x + q_i^\top x + r_i \right) \\
&= \frac{1}{2} x^\top \underbrace{\left(P_0 + \sum_{i=1}^m \lambda_i P_i \right)}_{\triangleq P(\lambda)} x + \underbrace{\left(q_0 + \sum_{i=1}^m \lambda_i q_i \right)}_{\triangleq q(\lambda)}{}^\top x + \underbrace{\left(r_0 + \sum_{i=1}^m \lambda_i r_i \right)}_{\triangleq r(\lambda)} \\
&= \frac{1}{2} x^\top P(\lambda)\, x + q(\lambda)^\top x + r(\lambda)
\end{aligned}
$$

例：QCQP 问题（续）

不难看到，其 Lagrange 对偶函数，即 $\inf_x \{L(x,\lambda)\}$，的具体形式依赖于 λ。

具体来说，当 $\lambda \geq 0$ 时，有 $P(\lambda) \succ 0$，因此 $P(\lambda)$ 可逆；而其他情况下，则无法保证。于是，有

$$g(\lambda) = \inf_x \{L(x,\lambda)\} = \begin{cases} -(1/2)\, q(\lambda)^\top P(\lambda)^{-1} q(\lambda) + r(\lambda), & \lambda \geq 0, \\ \text{不易推得}, & \text{otherwise.} \end{cases}$$

进一步，可写出对偶问题

$$\max_{\lambda \geq 0} \quad -(1/2)\, q(\lambda)^\top P(\lambda)^{-1} q(\lambda) + r(\lambda). \qquad \text{(Dual)}$$

最后，由 Slater's condition，若 $\exists x$ 使得所有二次不等式约束严格成立，即

$$(1/2)\, x^\top P_i\, x + q_i^\top x + r_i < 0, \quad i = 1, 2, \cdots, m,$$

则 QCQP 的强对偶性成立。

Slater's condition 是凸问题强对偶性成立的充分条件，但不是必要条件！

例：

$$\begin{aligned} \min \quad & x \\ \text{s.t.} \quad & x \le 0, \quad -x \le 0. \end{aligned} \qquad \textbf{(Primal)}$$

(i). 由于(Primal)的可行点只有 $x = 0$，故可行域不存在相对内部，因此原问题显然不满足 Slater's condition。易知，(Primal)的最优值 $p^* = 0$。

(ii). 进一步，考察它的 Lagrange 对偶函数：

$$\Rightarrow \quad L(x, \lambda_1, \lambda_2) = x + \lambda_1 x - \lambda_2 x$$

$$\Rightarrow \quad g(\lambda_1, \lambda_2) = \inf_x L(x, \lambda_1, \lambda_2) = \begin{cases} 0, & 1 + \lambda_1 - \lambda_2 = 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

得到对偶问题：

$$\begin{aligned} \max \quad & 0 \\ \text{s.t.} \quad & 1 + \lambda_1 - \lambda_2 = 0, \quad \lambda_1,\ \lambda_2 \ge 0. \end{aligned} \qquad \textbf{(Dual)}$$

显然，(Dual)的最优值 $d^* = p^* = 0$，此时强对偶性成立。

Slater's condition 与强对偶性（续）

扩展知识：非凸问题仍然可能有强对偶性。

例：置信域问题 (trust region problem).[1] 在单位球内极小化一个二次函数

$$\begin{aligned} \min \quad & x^\top A x + 2b^\top x \quad (A \in S^n,\ b \in \mathbb{R}^n) \\ \text{s.t.} \quad & x^\top x \le 1. \end{aligned}$$

若 $A \not\succeq 0$，则目标函数 $f_0 = x^\top A x$ 是非凸的。但是，该问题对偶间隙为 0。

[1] 参见 **Section 5**, Ronald J. Stern and Henry Wolkowicz. Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations. *SIAM Journal on Optimization*, 5(2): 286–313, 1995.

§ Lagrange 函数与对偶函数

§ Lagrange 对偶问题

§ Lagrange 对偶的几种解释

§ 最优性条件

§ 敏感性分析（Perturbation Analysis）

§ 利用对偶理解“启发式”方法

几何解释

我们首先从函数值集合的角度去理解 Lagrange 对偶。

考虑仅有一个不等式约束的优化问题（假设目标函数值在可行域上有下界）

$$\begin{aligned} \inf_{x} \quad & f_0(x) \\ \text{s.t.} \quad & f_1(x) \le 0. \end{aligned}$$

记问题的定义域为 $\mathcal{D} := \operatorname{dom} f_0 \cap \operatorname{dom} f_1$.

首先，定义如下符号：

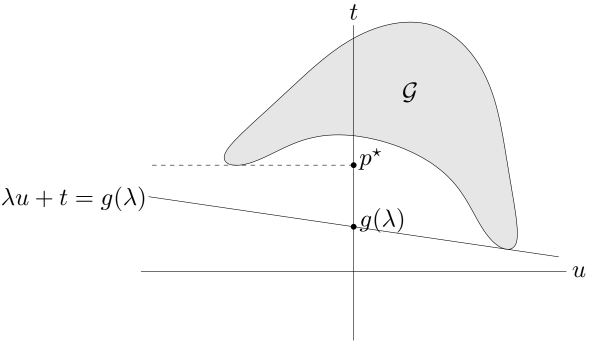
- 集合 $\mathcal{G} := \{\, (f_1(x),\, f_0(x)) \mid x \in \mathcal{D}\}$
- 最优值 $p^* := \inf_{t} \{t \mid (u,t) \in \mathcal{G},\ u \le 0\} \iff \inf_{x\in\mathcal{D}} \{f_0 \mid f_1 \le 0\}$
- 对偶函数

$$g(\lambda) := \inf_{t,u} \{t + \lambda u \mid (u,t) \in \mathcal{G}\} \iff \inf_{x\in\mathcal{D}} \underbrace{\{f_0(x) + \lambda f_1(x)\}}_{L(x,\lambda)}$$

几何解释（续）

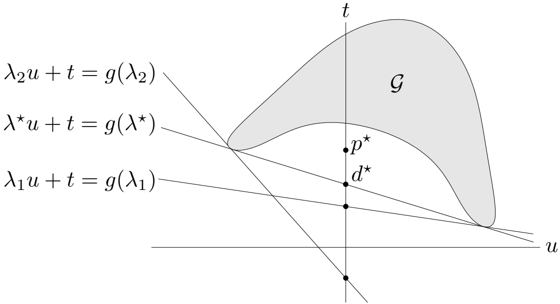
然后，作出变量 $(u,t) \in \mathcal{G}$ 的图像：

- 集合 $\mathcal{G}$ 对应图中的**阴影部分**；
- 最优值 p^* 对应在集合 $\mathcal{G}$ 中，当 $u \le 0$ 时，t 能取到的最小值；
- 对偶函数 $g(\lambda)$ 对应一条以 u 为自变量，t 为因变量，斜率为 $-\lambda$，必过集合 $\mathcal{G}$ 中一点，且在 t 方向上的“截距”最小的直线。



下图展示了 λ 分别取值 λ_1，λ_2 和 λ^* 时，对偶函数 $g(\lambda)$ 对应的直线：

- 对任意 λ，$g(\lambda)$ 对应的直线都与集合 $\mathcal{G}$ 相切（保证至少经过集合 $\mathcal{G}$ 中一点），且使得 t 方向截距最小；
- 对偶问题 $\sup\{g(\lambda) \mid \lambda \geq 0\}$：在斜率 $-\lambda \leq 0$ 的情况下找到上述“最小截距”中的“最大值，即原问题最优值的最大下界”。易分析，在图中，对偶最优解为 λ^*，则 $g(\lambda^*)$ 在 t 轴上的截距为 d^*；
- 在本例中，易观察到 $d^* < p^*$，即弱对偶性成立，但强对偶性不成立。



经济学解释

Lagrange 对偶也有着有趣的经济学解释。考虑这样一个场景：给定若干原材料，制定生产方案，卖产品，希望获得最大利润。

数学上，可以考虑如下含有多个不等式约束的优化问题：

$$
\begin{aligned}
\min \quad & -f_0(x) \\
\text{s.t.} \quad & x_i \le u_i, \quad i = 1, \cdots, m,
\end{aligned}
$$

其中

- x 表示产量；
- $f_0(x)$ 表示利润；（注：最小化负利润 $\Leftrightarrow$ 最大化利润）
- $x_i \le u_i$ 表示对第 i 种原材料用量的限制。

记求解该问题得到的（只卖产品的）最优生产方案为 x^*，最优值为 p^*。

经济学解释（续）

现假设原材料能够自由买卖，设第 i 种原材料的价格为 $\lambda_i \geq 0$。此时，我们可考虑更灵活的生产方案，既可以卖产品，也可以买卖原材料。

于是，对于一个生产方案 x：

- 若 $x_i \leq u_i$，表示第 i 种原材料库存有剩余，可以出售，以便增加利润 $\lambda_i(u_i - x_i)$。此时，总共的负利润可以表示为

$$-\left(f_0(x) + \textstyle\sum_{i=1}^m \lambda_i(u_i - x_i)\right)$$

- 若 $x_i > u_i$，表示第 i 种原材料要额外购入，以满足生产需求，但会产生额外原材料花销 $\lambda_i(x_i - u_i)$。此时，总共的负利润可以表示为

$$-\left(f_0(x) - \textstyle\sum_{i=1}^m \lambda_i(x_i - u_i)\right)$$

综上，在原材料能够自由买卖市场下，总共的负利润为 (Lagrange 函数)

$$L(x, \lambda) := -f_0(x) + \textstyle\sum_{i=1}^m \lambda_i(x_i - u_i)$$

于是，**对偶函数**表示给定原材料价格时，最优生产方案的最小负利润：

$$g(\lambda) \triangleq \inf_{x\in\mathcal{D}} \left\{ -f_0(x) + \sum_{i=1}^{m} \lambda_i (x_i - u_i) \right\}.$$

注意：无论原材料价格 λ 如何变化，由于可以自由买卖，生产方案更多样化，因此获利只可能更多，负利润只可能更小，即 $g(\lambda) \le p^*$。

进一步，若强对偶性成立，即 $d^* = g(\lambda^*) = p^*$，则表明：

当原材料定价为 λ^*，生产方案设定为 x^* 时，通过自由买卖购入额外的原材料扩大生产，或卖出原材料获得利润，都不能获得更高的收益，此时资源已调配至最优。

在经济学领域，λ^* 也被称为**影子价格**。

Lagrange 对偶也能从多目标规划的角度来理解：

考虑带有多个不等式约束的优化问题

$$\begin{aligned} \min_{x} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \cdots, m. \end{aligned}$$

写出 Lagrange 对偶函数：$L(x,\lambda) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x)$.

当 $\lambda \geq 0$ 时，$\inf_x L(x,\lambda) = \inf_x \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right\}$ 对应于求解多目标优化问题

$$\min_{x} \left[f_0(x), f_1(x), \cdots, f_m(x) \right]$$

的一个帕累托最优解（对应权重 λ）。

鞍点解释

Lagrange 对偶也能从鞍点角度来理解。首先，我们引入一个直观的不等式.

命题 1 (max-min 不等式)

对于任意函数 $f(w,z) : S_w \times S_z \to \mathbb{R}$，以下 max-min 不等式必成立：

$$\sup_{z \in S_z} \inf_{w \in S_w} f(w,z) \le \inf_{w \in S_w} \sup_{z \in S_z} f(w,z).$$

思考：max-min 不等式中的等式何时成立？

定义 2 (鞍点 (saddle point))

对于函数 $f(w,z) : S_w \times S_z \to \mathbb{R}$，若存在 $\exists (\tilde{w}, \tilde{z}) \in S_w \times S_z$ 使得

$$f(\tilde{w}, z) \le f(\tilde{w}, \tilde{z}) \le f(w, \tilde{z}), \qquad \forall\, w \in S_w, \forall\, z \in S_z,$$

则称 $(\tilde{w}, \tilde{z})$ 为函数 $f(w,z)$ 的一个鞍点 (saddle point)。

鞍点解释（续）

由鞍点的定义，不难看出，若 $(\tilde{w}, \tilde{z})$ 是一个鞍点，则

- $\tilde{w}$ 在 S_w 上极小化 $f(w, \tilde{z})$，即 $f(\tilde{w}, \tilde{z}) = \inf_{w \in S_w} f(w, \tilde{z})$
- $\tilde{z}$ 在 S_z 上极大化 $f(\tilde{w}, z)$，即 $f(\tilde{w}, \tilde{z}) = \sup_{z \in S_z} f(\tilde{w}, z)$

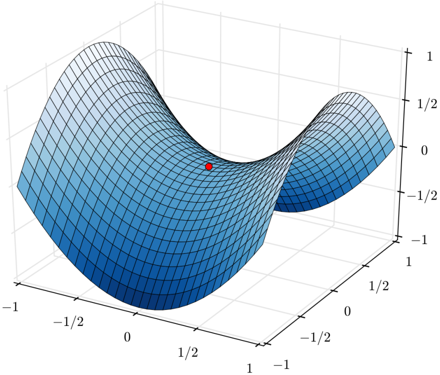


Fig. 鞍点示意图

鞍点解释（续）

命题 2

若 $(\tilde{w}, \tilde{z})$ 是 $f(w, z)$ 的一个鞍点，则 max-min 不等式在该鞍点处等式成立，且等于 $f(\tilde{w}, \tilde{z})$。

证明：因为 $(\tilde{w}, \tilde{z})$ 是 $f(w, z)$ 的一个鞍点，根据定义有

$$f(\tilde{w}, z) \le f(\tilde{w}, \tilde{z}) \le f(w, \tilde{z}), \qquad \forall\, w \in S_w,\ \forall\, z \in S_z.$$

于是有以下两个结论：

1. $\sup\limits_{z\in S_z} \inf\limits_{w\in S_w} f(w,z) \ge \inf\limits_{w\in S_w} f(w,\tilde{z}) \ge f(\tilde{w},\tilde{z})$
2. $\inf\limits_{w\in S_w} \sup\limits_{z\in S_z} f(w,z) \le \sup\limits_{z\in S_z} f(\tilde{w},z) \le f(\tilde{w},\tilde{z})$

综合 1 和 2，有 $\sup\limits_{z\in S_z} \inf\limits_{w\in S_w} f(w,z) \ge \inf\limits_{w\in S_w} \sup\limits_{z\in S_z} f(w,z).$

另一方面，由 max-min 不等式，可知

$$\sup_{z\in S_z} \inf_{w\in S_w} f(w,z) \le \inf_{w\in S_w} \sup_{z\in S_z} f(w,z).$$

综上，可得 $\sup\limits_{z\in S_z} \inf\limits_{w\in S_w} f(w,z) = \inf\limits_{w\in S_w} \sup\limits_{z\in S_z} f(w,z) = f(\tilde{w},\tilde{z})$，证毕。

鞍点解释（续）

考虑带有多个不等式约束的优化问题（假设目标函数值在可行域上有下界）

$$\begin{aligned} \min_{x} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, \quad i = 1, \cdots, m. \end{aligned} \qquad \text{(Primal)}$$

写出其 Lagrange 函数。然后，关于 λ 极大化 $L(x, \lambda)$，可得：

$$\begin{aligned} \sup_{\lambda \ge 0} L(x, \lambda) &= \sup_{\lambda \ge 0} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right\} \\ &= \begin{cases} f_0(x), & f_i(x) \le 0, \quad i = 1, \cdots, m, \\ +\infty, & \text{otherwise.} \end{cases} \end{aligned} \qquad (\star)$$

借助上一页的表达式，我们可将 (Primal) 的最优值表述为：

$$p^* = \inf_{x\in\mathcal{D}} \left\{ f_0(x) \mid f_i(x) \le 0,\ i = 1, \cdots, m \right\} \overset{(\bigstar)}{=} \inf_{x\in\mathcal{D}} \sup_{\lambda\ge 0} L(x,\lambda).$$

回忆对偶问题(Dual)及对偶最优值：

$$d^* = \sup_{\lambda\ge 0} \inf_{x\in\mathcal{D}} L(x,\lambda).$$

于是，根据 max-min 不等式，有

$$d^* = \sup_{\lambda\ge 0} \inf_{x\in\mathcal{D}} L(x,\lambda) \ \le\ \inf_{x\in\mathcal{D}} \sup_{\lambda\ge 0} L(x,\lambda) = p^*.$$

重要结论：

1. 由 max-min 不等式直接可得 $d^* \le p^*$ （**弱对偶性**）；
2. 当 $L(x,\lambda)$ 有鞍点时，**强对偶性成立**，且鞍点是原对偶最优解。

定理 2 (鞍点定理)

$(\tilde{x}, \tilde{\lambda})$ 是 Lagrange 函数 $L(x, \lambda)$ 的一个鞍点 $\iff$ 强对偶性成立且 $(\tilde{x}, \tilde{\lambda})$ 为一对原对偶最优解

证明： 先证 “$\Rightarrow$”。

若 $(\tilde{x}, \tilde{\lambda})$ 为 $L(x, \lambda)$ 的一个鞍点，则强对偶性成立。（**命题2**的结论）

根据上述结论和鞍点性质，有

$$\text{(a)} \quad L(\tilde{x}, \lambda) \le L(\tilde{x}, \tilde{\lambda}) \le L(x, \tilde{\lambda}), \quad \forall \lambda \ge 0, \forall x \in \mathcal{D},$$

$$\text{(b)} \quad \sup_{\lambda \ge 0} \inf_{x \in \mathcal{D}} L(x, \lambda) = L(\tilde{x}, \tilde{\lambda}) = \inf_{x \in \mathcal{D}} \sup_{\lambda \ge 0} L(x, \lambda).$$

接下来，我们证明 $\tilde{x}$ 和 $\tilde{\lambda}$ 分别是原对偶问题的最优解。

（接上一页）首先，可以分析

$$g(\tilde{\lambda}) = \inf_{x \in \mathcal{D}} L(x, \tilde{\lambda}) \overset{\text{(a)}}{=} L(\tilde{x}, \tilde{\lambda}) \overset{\text{(b)}}{=} \sup_{\lambda \geq 0} \inf_{x \in \mathcal{D}} L(x, \lambda) = \sup_{\lambda \geq 0} g(\lambda),$$

表明 $\tilde{\lambda}$ 是对偶问题的一个最优解。

另一方面，由前面式 ($\star$) 的推导可知

$$f_0(\tilde{x}) \leq \sup_{\lambda \geq 0} \underbrace{\left\{ f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) \right\}}_{L(\tilde{x}, \lambda)} \overset{\text{(a)}}{=} L(\tilde{x}, \tilde{\lambda})$$

可知 $\sup_{\lambda \geq 0} L(\tilde{x}, \lambda)$ 有意义，因此 $\tilde{x}$ 是原问题的一个可行解，即 $f_i(\tilde{x}) \leq 0$, $i = 1, \cdots, m$。进一步，由

$$f_0(\tilde{x}) \leq L(\tilde{x}, \tilde{\lambda}) \overset{\text{(b)}}{=} \inf_{x \in \mathcal{D}} \sup_{\lambda \geq 0} L(x, \lambda) \overset{(\star)}{=} \inf_{x \in \mathcal{D}} \{ f_0(x) \mid f_i(x) \leq 0, \ i = 1, \cdots, m \},$$

可知 $\tilde{x}$ 是原问题的一个最优解。综上，“$\Rightarrow$” 得证。

下面证明："$\Leftarrow$"

1. 若已知 $(\tilde{x}, \tilde{\lambda})$ 为一对原对偶最优解，则原对偶必可行，即

$$f_i(\tilde{x}) \leq 0, \quad i = 1, \cdots, m, \quad \tilde{\lambda} \geq 0.$$

2. 又因为强对偶性成立，因此

$$f_0(\tilde{x}) = g(\tilde{\lambda}) = \underbrace{\inf_{x \in \mathcal{D}} \left\{ f_0(x) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) \right\}}_{\text{Lagrange 对偶函数}} \leq f_0(\tilde{x}) + \underbrace{\sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x})}_{\leq 0} \leq f_0(\tilde{x}),$$

故上式中的等号均成立。

鞍点解释（续）

由上式等号都成立，有

$$f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) = \inf_{x\in\mathcal{D}} \left\{ f_0(x) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) \right\} \Leftrightarrow L(\tilde{x}, \tilde{\lambda}) = \inf_{x\in\mathcal{D}} L(x, \tilde{\lambda}),$$

即 $L(\tilde{x}, \tilde{\lambda}) \le L(x, \tilde{\lambda}),\ \forall\, x \in \mathcal{D}$。

另一方面，由前面式 ($\star$) 的表达式可知

$$\begin{aligned} f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) = f_0(\tilde{x}) &= \sup_{\lambda\ge 0} \left\{ f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) \right\} \\ \Leftrightarrow \quad L(\tilde{x}, \tilde{\lambda}) &= \sup_{\lambda\ge 0} L(\tilde{x}, \lambda), \end{aligned}$$

即 $L(\tilde{x}, \tilde{\lambda}) \ge L(\tilde{x}, \lambda),\ \forall\, \lambda \ge 0$。

综上可知，$(\tilde{x}, \tilde{\lambda})$ 是 Lagrange 函数 $L(x, \lambda)$ 的一个鞍点，证毕。

- § Lagrange 函数与对偶函数
- § Lagrange 对偶问题
- § Lagrange 对偶的几种解释
- § 最优性条件
- § 敏感性分析（Perturbation Analysis）
- § 利用对偶理解“启发式”方法

最优性条件：研究最优解所要满足的条件。

考虑一般优化问题（可能非凸）

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leqslant 0, \quad i = 1, \cdots, m, \\ & h_j(x) = 0, \quad j = 1, \cdots, p. \end{aligned} \qquad \textbf{(Primal)}$$

假设：（考虑简单情形）

- 问题的定义域为 $\mathbb{R}^n$，即 $\left(\bigcap_{i=0}^m \operatorname{dom} f_i\right) \cap \left(\bigcap_{j=1}^p \operatorname{dom} h_j\right) = \mathbb{R}^n$；
- 函数 f_i, $i = 0, 1, \cdots, m$ 和 h_j, $j = 1, 2, \cdots, p$ 均连续可微；
- 问题的最优解 x^* 和 最优值 p^* 存在，且强对偶成立。

写出其对偶函数 $g(\lambda, \nu) = \inf_x \left\{ f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^p \nu_j h_j(x) \right\}$

和对偶问题

$$\max_{\lambda, \nu} \quad g(\lambda, \nu), \quad \text{s.t.} \quad \lambda \geq 0. \qquad \textbf{(Dual)}$$

最优性条件：互补松弛

进一步，假设对偶问题的解也存在，记为 (λ^*, ν^*)。于是，有

$$\begin{aligned}
\underbrace{f_0(x^*) = g(\lambda^*, \nu^*)}_{\text{根据强对偶性假设}} &= \inf_x \left\{ f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{j=1}^p \nu_j^* h_j(x) \right\} \\
&\le f_0(x^*) + \sum_{i=1}^m \underbrace{\lambda_i^* f_i(x^*)}_{\le 0} + \sum_{j=1}^p \underbrace{\nu_j^* h_j(x^*)}_{=0} \\
&\le f_0(x^*),
\end{aligned} \tag{1}$$

其中

- 根据 $f_i(x^*) \le 0$ 和 $\lambda_i^* \ge 0$, $\forall i = 1, 2, \cdots, m$，有 $\lambda_i^* f_i(x^*) \le 0$；
- 根据 $h_j(x^*) = 0$, $\forall j = 1, 2, \cdots, p$，有 $\nu_j^* h_j(x^*) = 0$。

显然，上述式中等式成立，即

$$f_0(x^*) = f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{j=1}^p \nu_j^* h_j(x^*).$$

最优性条件：互补松弛（续）

于是，可以得到如下重要结论：

$$\sum_{i=1}^{m} \lambda_i^* f_i(x^*) = 0.$$

同时，由<u>**原对偶可行性**</u>可知，求和项每项都非正，因此

$$\lambda_i^* f_i(x^*) = 0, \quad \forall i = 1, \cdots, m,$$

$$\Longrightarrow \begin{cases} \lambda_i^* > 0, & \Longrightarrow \quad f_i(x^*) = 0, \\ f_i(x^*) < 0, & \Longrightarrow \quad \lambda_i^* = 0. \end{cases}$$

上述性质称为**互补松弛条件 (complementary slackness)**。

该条件表明：在最优点处，如果第 i 个约束没起作用（即 $f_i(x^*) < 0$ ），则相应的最优 Lagrange 乘子 $\lambda_i^* = 0$。

最优性条件：稳定性

进一步，由前面的式 (1) 可知：

$$\inf_x L(x,\lambda^*,\nu^*) = \inf_x \left\{ f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{j=1}^p \nu_j^* h_j(x) \right\}$$
$$= f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{j=1}^p \nu_j^* h_j(x^*) = L(x^*,\lambda^*,\nu^*).$$

上式表明 $L(x,\lambda^*,\nu^*)$ 关于 x 在 x^* 处取得极小值。**根据无约束优化问题的最优性条件可知：（非凸时仅为必要条件）**

$$\left.\frac{\partial L(x,\lambda^*,\nu^*)}{\partial x}\right|_{x=x^*} = 0,$$
$$\Rightarrow \quad \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{j=1}^p \nu_j^* \nabla h_j(x^*) = 0.$$

上述等式称为**稳定性条件 (stationarity)**。

KKT 条件

KKT 条件（最优解的必要条件）

对于连续可微且对偶间隙为 **0** 的优化问题，若原对偶最优解 (x^*, λ^*, ν^*) 存在，则必须满足以下条件：

$$
\begin{aligned}
f_i(x^*) &\le 0, \quad i = 1, \cdots, m, && \text{(primal feasibility)} \\
h_j(x^*) &= 0, \quad j = 1, \cdots, p, && \text{(primal feasibility)} \\
\lambda_i^* &\ge 0, \quad i = 1, \cdots, m, && \text{(dual feasibility)} \\
\lambda_i^* f_i(x^*) &= 0, \quad i = 1, \cdots, m, && \text{(complementary slackness)} \\
\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(x^*) &= 0, && \text{(stationarity)}
\end{aligned}
$$

以上条件合称为 **Karush-Kuhn-Tucker** (**KKT**) 条件。

思考：KKT 条件一般只是必要条件，那么对于凸优化问题如何？即如果 $(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ 满足 KKT 条件，是否能推导它是 (Primal) 和 (Dual) 的最优解？

定理 3 (KKT 条件是原对偶最优解的充要条件)

对于目标函数和约束函数均连续可微，且强对偶性成立的凸优化问题，有：

$$(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \text{ 为原对偶最优解} \quad \Longleftrightarrow \quad (\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \text{ 满足 } \textbf{KKT} \text{ 条件}.$$

证明：

- 对于 "$\Rightarrow$"（**必要性**），我们已经证明了最优解必满足KKT条件；
- 对于 "$\Leftarrow$"（**充分性**），由 $\tilde{x}$ 和 $(\tilde{\lambda}, \tilde{\nu})$ 可行，只需证明 $g(\tilde{\lambda}, \tilde{\nu}) = f_0(\tilde{x})$，即可进一步推出 $\tilde{x}$ 和 $(\tilde{\lambda}, \tilde{\nu})$ 分别是(Primal)和(Dual)的最优解。具体分析如下：

1. 由于目标函数和约束函数均是连续可微凸函数，因此 $L(x, \tilde{\lambda}, \tilde{\nu})$ 是关于 x 的连续可微凸函数。于是，可知

$$\underbrace{\left.\frac{\partial L(x, \tilde{\lambda}, \tilde{\nu})}{\partial x}\right|_{x=\tilde{x}} = 0}_{\text{stationarity}} \quad \Rightarrow \quad \tilde{x} = \arg\min_{x} \; L(x, \tilde{\lambda}, \tilde{\nu})$$

2. 根据 Lagrange 对偶函数的定义：

$$g(\tilde{\lambda}, \tilde{\nu}) = \inf_x \; L(x, \tilde{\lambda}, \tilde{\nu}) \overset{\text{在 } \tilde{x} \text{ 处最优}}{=} L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$$

$$= f_0(\tilde{x}) + \underbrace{\sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x})}_{=0,\ \text{由互补松弛}} + \underbrace{\sum_{j=1}^{p} \tilde{\nu}_j h_j(\tilde{x})}_{=0,\ \text{由原问题可行性}} = f_0(\tilde{x})$$

3. 根据弱对偶定理，有 $d^* \le p^*$。另一方面，可以分析

$$d^* = \sup_{\lambda \ge 0,\, \nu} g(\lambda, \nu) \ge g(\tilde{\lambda}, \tilde{\nu}) = f_0(\tilde{x}) \ge p^*.$$

故上式中等式成立，于是

$$d^* = \sup_{\lambda \ge 0,\, \nu} g(\lambda, \nu) = g(\tilde{\lambda}, \tilde{\nu}),$$

$$p^* = f_0(\tilde{x}).$$

因此，$(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ 为一对**原对偶最优解**，证毕。

KKT 条件

总结：

- 对于一般的**连续可微凸优化问题**，如果**强对偶性成立**，则 KKT 条件为原对偶最优解需要满足的**充要条件**
- 对于一般的**连续可微（非凸）优化问题**，如果**强对偶成立**，则 KKT 条件为原对偶最优解需要满足的**必要条件**
- 绝大部分优化算法都是在寻找满足 KKT 条件及其推广条件的点

KKT (Karush-Kuhn-Tucker) 条件的相关历史（非正式叙述）：

- Lagrange 最早研究了带有等式约束的优化问题的最优性条件
- Kuhn 和 Turker 在1951年给出了带有不等式约束的优化问题的最优性条件
- 之后，人们发现 Karush 早在他1939年未发表的硕士论文中已提出了同样的最优性条件

KKT 的应用：非负约束凸优化问题

考虑非负约束的连续可微凸优化问题：

$$\min_x \quad f_0(x) \quad \text{s.t.} \quad x \geq 0.$$

写出它的 KKT 条件：

$$\begin{cases} x^* \geq 0, \\ \lambda^* \geqslant 0, \\ \lambda_i^*(-x_i^*) = 0, \; i = 1, \cdots, n, \\ \nabla f_0(x^*) + (-\lambda^*) = 0. \end{cases} \iff \begin{cases} x^* \geqslant 0, \\ \nabla f_0\left(x^*\right) \geqslant 0, \\ x_i^*\left[\nabla f_0\left(x^*\right)\right]_i = 0, \; i = 1, \cdots, n. \end{cases}$$

显然，对了该问题，Slater's condition 成立，因此强对偶性成立，进而满足 KKT 条件的点即为问题的最优解。

请同学们同时回顾第一章课件P102**例2**.

KKT 的应用：注水算法

注水(water filling)算法与问题背景

将总和为1的总功率的信号分配到 n 个信道传输，以获得最大通信速率：

$$
\begin{aligned}
\min_{x} \quad & -\sum_{i=1}^{n} \log(\alpha_i + x_i) && \Leftarrow \quad \text{等价于最小化负通信率} \\
\text{s.t.} \quad & x \geq 0 && \Leftarrow \quad \text{功率非负} \\
& \mathbf{1}^\top x = 1 && \Leftarrow \quad \text{功率总和为1}
\end{aligned}
$$

其中 $\alpha_i > 0$ 是已知的量，它代表第 i 个信道的状况。

该问题的 KKT 条件为

$$
\begin{aligned}
x^* &\geq 0 \\
\mathbf{1}^\top x^* &= 1 \\
\lambda^* &\geq 0 \\
\lambda_i^* x_i^* &= 0, \quad i = 1, \cdots, n \\
-\frac{1}{\alpha_i + x_i^*} - \lambda_i^* + \nu^* &= 0, \quad i = 1, \cdots, n
\end{aligned}
$$

KKT 的应用：注水算法（续）

根据 $-\dfrac{1}{\alpha_i + x_i^*} - \lambda_i^* + \nu^* = 0,\ \lambda^* \geq 0$ 和 $\lambda_i^* x_i^* = 0$ 有：

$$\nu^* \geq \frac{1}{\alpha_i + x_i^*}, \quad x_i^* \left(\nu^* - \frac{1}{\alpha_i + x_i^*}\right) = 0, \quad i = 1, \cdots, n$$

1. 若 $\nu^* < \dfrac{1}{\alpha_i}$，根据 $\nu^* \geq \dfrac{1}{\alpha_i + x_i^*}$，可得 $\dfrac{1}{\alpha_i} > \nu^* \geq \dfrac{1}{\alpha_i + x_i^*} \Rightarrow x_i^* > 0$;
于是，由互补条件可知 $\lambda_i^* = 0$，此时 $x_i^* = \dfrac{1}{\nu^*} - \alpha_i$;

2. 若 $\nu^* > \dfrac{1}{\alpha_i}$，则由 $x_i^* \geq 0$ 可知 $\nu^* > \dfrac{1}{\alpha_i + x_i^*}$。再由
$x_i^* \left(\nu^* - 1/(\alpha_i + x_i^*)\right) = 0$，可知 $x_i^* = 0$;

3. 若 $\nu^* = \dfrac{1}{\alpha_i}$，也可根据 $x_i^* \left(\nu^* - 1/(\alpha_i + x_i^*)\right) = 0$ 推出 $x_i^* = 0$.

KKT 的应用：注水算法（续）

综上，$x_i^* = \max\left\{0, \dfrac{1}{\nu^*} - \alpha_i\right\}$，其中 ν^* 满足 $\sum_{i=1}^{n} \max\left\{0, \dfrac{1}{\nu^*} - \alpha_i\right\} = 1$；

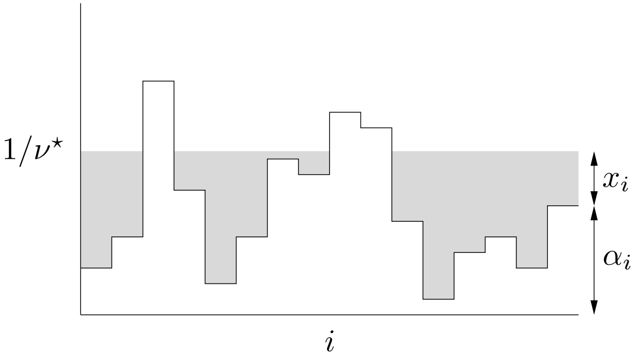


Fig. 求解该问题 KKT 点的算法有个很“形象”的名称：注水算法。假设我们有一口底部不平整的缸，α_i 是缸底第 i 片区域的高度。逐渐向缸中注水，记水位为 $\dfrac{1}{\nu}$。不断注水直至总水量为 1。此时，第i 个区域对应的水位深度即为 x_i^*！

- § Lagrange 函数与对偶函数
- § Lagrange 对偶问题
- § Lagrange 对偶的几种解释
- § 最优性条件
- § 敏感性分析（Perturbation Analysis）
- § 利用对偶理解“启发式”方法

在本节，我们希望研究约束条件的变化会如何影响原问题的最优值。

考虑一般优化问题

$$
\begin{aligned}
\min \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \le 0, \quad i = 1, \cdots, m, \\
& h_j(x) = 0, \quad j = 1, \cdots, p.
\end{aligned}
$$

该问题的定义域记为 $\mathcal{D} = \bigcap_{i=0}^{m} \operatorname{dom} f_i \cap \bigcap_{j=0}^{p} \operatorname{dom} h_j$.

考虑该优化问题的约束发生扰动：

一般优化问题的扰动问题

$$
\begin{aligned}
\min \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \le u_i, \quad i = 1, \cdots, m, \\
& h_j(x) = w_j, \quad j = 1, \cdots, p.
\end{aligned}
$$

记扰动问题的最优值为 $p^*(u, w)$。原问题的最优值自然对应为 $p^*(0, 0)$

性质1

若原问题为凸优化问题，则 $p^*(u,w)$ 是关于 (u,w) 的凸函数。

证明： 回顾凸优化问题的定义：

$$
\begin{array}{lll}
\min\limits_{x} & f_0(x) & f_0: \text{凸函数}; \\
\text{s.t.} & f_i(x) \le 0, \quad i = 1, \cdots, m, & f_i: \text{凸函数}; \\
& h_j(x) = 0, \quad j = 1, \cdots, p, & h_j(x) = a_j^\top x - b_j: \text{仿射函数}.
\end{array}
$$

根据带有扰动的一般优化问题的定义，将 $p^*(u,w)$ 记为

$$
p^*(u,w) = \begin{cases} \inf\limits_{x \in \mathcal{D}} \left\{ f_0(x) \,\middle|\, \begin{array}{ll} f_i(x) \le u_i, & i = 1, \cdots, m \\ h_j(x) = w_i, & j = 1, \cdots, p \end{array} \right\}, & \text{如果问题可行}; \\ +\infty, & \text{如果问题不可行}. \end{cases}
$$

对于函数 $p^*(u,w)$，易知它的定义域 $\operatorname{dom} p^* = \mathbb{R}^m \times \mathbb{R}^p$ 为凸集。

考虑任意两组 $(u^{(1)}, w^{(1)})$, $(u^{(2)}, w^{(2)}) \in \operatorname{dom} p^*$，其对应扰动问题的可行域，分别记为

$$\mathcal{X}_1^* = \left\{ x \in \mathcal{D} \;\middle|\; \begin{array}{ll} f_i(x) \le u_i^{(1)}, & i = 1, \cdots, m \\ h_j(x) = w_i^{(1)}, & j = 1, \cdots, p \end{array} \right\}$$

和

$$\mathcal{X}_2^* = \left\{ x \in \mathcal{D} \;\middle|\; \begin{array}{ll} f_i(x) \le u_i^{(2)}, & i = 1, \cdots, m \\ h_j(x) = w_i^{(2)}, & j = 1, \cdots, p \end{array} \right\}$$

1. 如果 $\mathcal{X}_1^* = \emptyset$ 或 $\mathcal{X}_2^* = \emptyset$，则 $p^*(u^{(1)}, w^{(1)}) = +\infty$ 或 $p^*(u^{(2)}, w^{(2)}) = +\infty$。

因此，对 $\forall \theta \in [0,1]$ 有

$$\begin{aligned} &\theta p^*(u^{(1)}, w^{(1)}) + (1-\theta) p^*(u^{(2)}, w^{(2)}) \\ &\qquad \ge p^*\left(\theta u^{(1)} + (1-\theta) u^{(2)}, \theta w^{(1)} + (1-\theta) w^{(2)}\right) \end{aligned}$$

2. 如果 $\mathcal{X}_1^* \neq \emptyset$ 且 $\mathcal{X}_2^* \neq \emptyset$，则根据凸优化问题的性质，对 $\forall x_1 \in \mathcal{X}_1^*$，$\forall x_2 \in \mathcal{X}_2^*$ 和 $\forall \theta \in [0,1]$，有以下结论：

- $\theta x_1 + (1-\theta)x_2 \in \mathcal{D}$；（根据 $x_1, x_2 \in \mathcal{D}$ 且 $\mathcal{D}$ 为凸集）
- $f_i(\theta x_1 + (1-\theta)x_2) \leq \theta f_i(x_1) + (1-\theta) f_i(x_2) \leq \theta u_i^{(1)} + (1-\theta)u_i^{(2)}$；（根据 $f_i,\ i = 1, \cdots, m,$ 为凸函数）
- $h_j(\theta x_1 + (1-\theta)x_2) = \theta h_j(x_1) + (1-\theta) h_j(x_2) = \theta w_j^{(1)} + (1-\theta)w_j^{(2)}$；（根据 $h_j,\ j = 1, \cdots, p,$ 为仿射函数）

上述结论表明：凸组合 $\theta x_1 + (1-\theta)x_2$ 是扰动问题

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq \theta u_i^{(1)} + (1-\theta)u_i^{(2)}, \quad i = 1, \cdots, m, \\ & h_j(x) = \theta w_i^{(1)} + (1-\theta)w_i^{(2)}, \quad j = 1, \cdots, p, \end{aligned}$$

的一个**可行解**，因此对 $\forall x_1 \in \mathcal{X}_1^*, \forall x_2 \in \mathcal{X}_2^*$ 和 $\forall \theta \in [0,1]$ 必有

$$f_0(\theta x_1 + (1-\theta)x_2) \geq p^*(\theta u^{(1)} + (1-\theta)u^{(2)}, \theta w^{(1)} + (1-\theta)w^{(2)})$$

3. 进一步，由 f_0 的凸性和上述不等式可知，对 $\forall x_1 \in \mathcal{X}_1^*, \forall x_2 \in \mathcal{X}_2^*$，有

$$\begin{aligned}\theta f_0(x_1) + (1-\theta) f_0(x_2) &\geq f_0(\theta x_1 + (1-\theta) x_2) \\ &\geq p^*(\theta u^{(1)} + (1-\theta) u^{(2)}, \theta w^{(1)} + (1-\theta) w^{(2)}).\end{aligned}$$

于是，取不等式左侧 $\theta f_0(x_1) + (1-\theta) f_0(x_2)$ 的下确界，有

$$\begin{aligned}&\inf_{x_1 \in \mathcal{X}_1^*,\, x_2 \in \mathcal{X}_2^*} \{\theta f_0(x_1) + (1-\theta) f_0(x_2)\} \\ =& \inf_{x_1 \in \mathcal{X}_1^*} \theta f_0(x_1) + \inf_{x_2 \in \mathcal{X}_2^*} (1-\theta) f_0(x_2) \qquad (\text{根据变量 } x_1, x_2 \text{ 的可分性}) \\ =& \theta p^*(u^{(1)}, w^{(1)}) + (1-\theta) p^*(u^{(2)}, w^{(2)}).\end{aligned}$$

结合上述两式可得

$$\theta p^*(u^{(1)}, w^{(1)}) + (1-\theta) p^*(u^{(2)}, w^{(2)}) \geq p^*(\theta u^{(1)} + (1-\theta) u^{(2)}, \theta w^{(1)} + (1-\theta) w^{(2)}).$$

函数 $p^*(u, w)$ 的凸性由此得证!

性质2

若原问题为凸问题，对偶间隙为 0，设 (λ^*, ν^*) 为原问题对偶最优解，则

$$p^*(u,w) \geq p^*(0,0) - (\lambda^*)^\top u - (\nu^*)^\top w.$$

证明：设 $\tilde{x}$ 为扰动问题的任意一个可行解，即

$$\tilde{x} \in \mathcal{X}_{u,w} = \left\{ x \in \mathcal{D} \;\middle|\; \begin{array}{ll} f_i(\tilde{x}) \leq u_i & i = 1, \cdots, m \\ h_j(\tilde{x}) = w_j & j = 1, \cdots, p \end{array} \right\}$$

进一步，根据对偶间隙为 0，有

$$\begin{aligned}
p^*(0,0) &= g(\lambda^*, \nu^*) \\
&\leq f_0(\tilde{x}) + \textstyle\sum_{i=1}^m \lambda_i^* f_i(\tilde{x}) + \sum_{j=1}^p \nu_j^* h_j(\tilde{x}) && \text{根据} g(\lambda^*, \nu^*) = \inf_x L(x, \lambda^*, \nu^*) \\
&\leq f_0(\tilde{x}) + (\lambda^*)^\top u + (\nu^*)^\top w && \text{根据} f_i(\tilde{x}) \leq u_i, h_j(\tilde{x}) = w_j \\
\Rightarrow \quad & f_0(\tilde{x}) \geq p^*(0,0) - (\lambda^*)^\top u - (\nu^*)^\top w \\
\Rightarrow \quad & p^*(u,w) \geq p^*(0,0) - (\lambda^*)^\top u - (\nu^*)^\top w. && \text{根据 } p^*(u,w) = \inf_{\tilde{x} \in \mathcal{X}_{u,w}} f_0(x)
\end{aligned}$$

至此性质 2 得证。

敏感性分析（续）

根据**性质2**可以观察到：

- 若 λ_i^* 很大，如果加强不等式约束为 $f_i(x) \le u_i < 0$，则最优值增大。
- 若 ν_i^* 为较大的正值，如果 $w_i < 0$，则最优值增大。
- 若 ν_i^* 为较小的负值，如果 $w_i > 0$，则最优值增大。

性质3 (局部敏感性)

若原问题为凸，对偶间隙为 0 且 $p^*(u,w)$ 在 $(u,w)=(0,0)$ 点处连续可微，则原始问题的最优对偶变量 (λ^*,ν^*) 满足

$$\lambda_i^* = -\nabla_{u_i} p^*(0,0), \qquad \nu_i^* = -\nabla_{w_i} p^*(0,0).$$

性质 3 表明，在局部区域有

$$p^*(u,w) \approx p^*(0,0) + (\nabla p^*(0,0))^\top \begin{bmatrix} u \\ w \end{bmatrix} = p^*(0,0) - (\lambda^*)^\top u - (\nu^*)^\top w.$$

§ Lagrange 函数与对偶函数

§ Lagrange 对偶问题

§ Lagrange 对偶的几种解释

§ 最优性条件

§ 敏感性分析（Perturbation Analysis）

§ 利用对偶理解“启发式”方法

利用对偶理解“启发式”方法

0-1 整数线性规划问题（非凸）

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & Ax \le b, \\ & x_i \in \{0,1\}, \quad i = 1, \cdots, n, \end{aligned}$$

其中决策变量 x 的每一个分量 x_i 只能取值 0 或 1。

思路1： 一种容易想到的思路是对约束 $x_i \in \{0,1\}$ 进行松弛。

0-1 整数线性规划的松弛问题

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & Ax \le b, \\ & 0 \le x_i \le 1, \quad i = 1, \cdots, n. \end{aligned}$$

- 松弛过后的问题成为线性规划问题（凸）；
- 显然，松弛问题的最优值比原来要更小，但最优解不一定取值为 0 或 1。

思路2：为了避免改变可行域，我们考虑等价地改写约束 $x_i \in \{0,1\}$。

0-1 整数线性规划的等价问题

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & Ax \leq b, \\ & x_i\left(x_i-1\right)=0, \quad i=1,\cdots,n. \end{aligned}$$

显然，该等价问题的可行域和最优解都与原问题相同。

进一步，写出等价问题的 Lagrange 函数

$$\begin{aligned} L(x,\lambda,\nu) &= c^\top x + \lambda^\top(Ax-b) + \sum_{i=1}^{n} \nu_i(x_i^2 - x_i) \\ &= -\lambda^\top b + \sum_{i=1}^{n}\left[\nu_i x_i^2 + \left(\lambda^\top a_i + c_i - \nu_i\right)x_i\right], \end{aligned}$$

其中 a_i 为矩阵 A 的第 i 列。

利用对偶理解“启发式”方法

写出对偶函数：

$$
\begin{aligned}
g(\lambda,\nu) &= \inf_x \; L(x,\lambda,\nu) \\
&= \begin{cases} -\lambda^\top b - \dfrac{1}{4}\displaystyle\sum_{i=1}^n \dfrac{(c_i + a_i^\top \lambda - \nu_i)^2}{\nu_i}, & \nu_i \ge 0,\ \forall i = 1,2,\cdots,n, \\ -\infty, & \text{otherwise.} \end{cases}
\end{aligned}
$$

写出对偶问题：

$$
\begin{aligned}
\max_{\lambda,\nu} \quad & -\lambda^\top b - \frac{1}{4}\sum_{i=1}^n \frac{(c_i + a_i^\top \lambda - \nu_i)^2}{\nu_i} \\
\text{s.t.} \quad & \lambda \ge 0, \quad \nu \ge 0.
\end{aligned}
$$

利用对偶理解“启发式”方法

进一步，可以利用 $\max_{\lambda,\nu} g(\lambda,\nu) = \max_{\lambda}\max_{\nu} g(\lambda,\nu)$ 消去 ν，以简化计算。

事实上，对任意 λ，考虑关于 ν 的优化问题：

$$\max_{\nu\geq 0} \quad -\lambda^\top b - \frac{1}{4}\sum_{i=1}^{n}\frac{(c_i + a_i^\top\lambda - \nu_i)^2}{\nu_i}$$

$$\Leftrightarrow \quad -\lambda^\top b + \sum_{i=1}^{n}\max_{\nu_i\geq 0}\left\{-\frac{1}{4}\frac{(c_i + a_i^\top\lambda - \nu_i)^2}{\nu_i}\right\}$$

对于 $z_i^* := \max_{\nu_i\geq 0}\left\{-\frac{1}{4}\frac{(c_i + a_i^\top\lambda - \nu_i)^2}{\nu_i}\right\}$（一维问题），可分情况讨论：

$$z_i^* := \begin{cases} c_i + a_i^\top\lambda, & c_i + a_i^\top\lambda \leq 0, \quad (\nu_i^{*,\lambda} = -(c_i + a_i^\top\lambda)) \\ 0, & c_i + a_i^\top\lambda > 0, \quad (\nu_i^{*,\lambda} = c_i + a_i^\top\lambda) \end{cases}$$

$$= \min\left\{0,\ c_i + a_i^\top\lambda\right\}.$$

于是，可以将对偶问题等价地改写为

$$\max_{\lambda} \quad -\lambda^\top b + \sum_{i=1}^{n} \min\left\{0,\, c_i + a_i^\top \lambda\right\}, \quad \text{s.t.} \quad \lambda \geq 0.$$

引入新变量 ω，上述问题可进一步等价于

$$\begin{aligned} \max_{\lambda,\,\omega} \quad & -\lambda^\top b + \mathbf{1}^\top \omega \\ \text{s.t.} \quad & \lambda \geq 0, \quad \omega \leq 0, \\ & \omega_i \leq a_i^\top \lambda + c_i, \quad i = 1, \cdots, n, \end{aligned}$$

其中 $\mathbf{1}$ 为元素都为 1 的向量。

利用对偶理解“启发式”方法

回顾 0-1 整数线性规划的松弛问题：

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & Ax \le b \\ & 0 \le x_i \le 1, \quad i = 1, \cdots, n. \end{aligned}$$

它的拉格朗日函数为

$$\begin{aligned} L(x,u,v,t) &= c^\top x + u^\top (Ax - b) - v^\top x + t^\top (x - \mathbf{1}) \\ &= (c + A^\top u - v + t)^\top x - b^\top u - \mathbf{1}^\top t. \end{aligned}$$

对偶函数为

$$g(u,v,t) = \begin{cases} -b^\top u - \mathbf{1}^\top t, & c + A^\top u - v + t = 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

对偶问题为

$$\begin{aligned} \max_{u,v,t} \quad & -b^\top u - \mathbf{1}^\top t \\ \text{s.t.} \quad & c + A^\top u - v + t = 0 \\ & u \ge 0, \quad v \ge 0, \quad t \ge 0. \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} \max_{u,t} \quad & -b^\top u - \mathbf{1}^\top t \\ \text{s.t.} \quad & c + A^\top u + t \ge 0 \\ & u \ge 0, \quad t \ge 0. \end{aligned}$$

对比“松弛问题”的对偶和“等价问题”的对偶：

$$
\begin{aligned}
\max_{u,t}\quad & -b^\top u - \mathbf{1}^\top t \\
\text{s.t.}\quad & c + A^\top u + t \ge 0 \\
& u \ge 0,\quad t \ge 0.
\end{aligned}
\quad \Longleftrightarrow \quad
\begin{aligned}
\max_{\lambda,\omega}\quad & -\lambda^\top b + \mathbf{1}^\top \omega \\
\text{s.t.}\quad & \lambda \ge 0,\quad \omega \le 0, \\
& \omega_i \le a_i^\top \lambda + c_i,\quad i = 1,\cdots,n.
\end{aligned}
$$