

优化问题中的对偶性理论

本文讲的是优化问题中与对偶问题、对偶性理论相关的内容，包括[对偶问题](#)的最优解、[弱对偶性](#)、[强对偶性](#)、[共轭函数](#)、以及[KKT条件](#)等。

一、对偶函数

拉格朗日量（Lagrangian）

考虑一个标准形式的优化问题：

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

称之为原问题，记其最优解为 p^* ，最优解满足 $f_0(p^*) = \min\{f_0(x) \mid x \in \mathcal{D}\}$ 。

其中 $\emptyset \neq \mathcal{D} \subset \mathbb{R}^n$ 是可行域。

定义上述问题的拉格朗日量（Lagrangian）如下：

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

其中向量 λ 和 ν 称为该问题的 **对偶变量**（dual variables）或者 **拉格朗日乘子**（Lagrange multiplier vectors）。

对偶函数

同时定义 **拉格朗日对偶函数**（Lagrange dual function）如下：

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right)$$

有以下几点相关内容需要注意：

1、对偶函数的凸性

- 无论原问题中的对偶函数是否是凸函数，其对偶函数**总是凹的**。这是因为可数个仿射函数的逐点最小值构成的函数是凹的，即如下定理：
- 定理1** 如果 f 关于 (x, y) 是凸函数， \mathcal{C} 是非空凸集，则函数

$$g(x) = \inf_{y \in \mathcal{C}} f(x, y)$$

关于 x 是凹函数。其中 g 的定义域是 f 的定义域到 x 方向的投影，即：

$$\text{dom } g = \{x \mid (x, y) \in \text{dom } f \text{ for some } y \in \mathcal{C}\}$$

而对偶函数 $g(\lambda, \nu)$ 恰好是一族关于 (λ, ν) 的仿射函数的逐点最小，因此它是凹的，这与 $f_i(x)$ 与 $h_i(x)$ 的凸性无关。

2、拉格朗日乘子法

拉格朗日乘子法本是为了求多元函数的条件极值的：即 U 为 \mathbb{R}^n 中开集， $f: U \rightarrow \mathbb{R}$ 为 U 上多元函数， $\Phi: U \rightarrow \mathbb{R}^m$ ($m < n$) 为 C^1 映射，令 $A = \{x \in U: \Phi(x) = 0\}$ ，则 f 在 A 上的局部极值称为条件极值，方程 $\Phi(x) = 0$ 称为约束条件。

关于求条件极值的方法，有如下定理，可用隐函数定理证明，过程从略。

- **定理2** (Lagrange乘数法) 设 $f \in C^1(U)$ ， $x^0 \in A$ 为 f 的条件极值点。如果梯度 $\nabla \Phi(x^0)$ 的秩为 m ，则存在 $\lambda \in \mathbb{R}^m$ ，使得

$$\nabla f(x^0) - \lambda \cdot \nabla \Phi(x^0) = 0$$

注：由定理可知，若 x^0 为条件极值点，则 (x^0, λ) 为辅助函数

$$F(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i \cdot \phi_i(x)$$

的驻点。

3、形象解释

参考 MathWorld 上对 [Lagrangian Multiplier](#) 的形象解释。

最优值的下界估计

回忆原问题中的最优解 p^* ，引入对偶函数可以导出 p^* 的下界。

- **性质1** 对于任意的 $\lambda \succeq 0$ 和任意的 ν ，都有 $g(\lambda, \nu) \leq p^*$

这个证明是很容易的，设 \tilde{x} 是原问题的一个可行解，则有

$$\sum_{i=1}^m \lambda_i f_i(\tilde{x}) + \sum_{i=1}^p \nu_i h_i(\tilde{x}) \leq 0$$

从而，

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

因此，

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

由于 $g(\lambda, \nu) \leq f_0(\tilde{x})$ 对任意 \tilde{x} 皆成立，故 $g(\lambda, \nu) \leq p^*$ 。

有以下几点需要注意：

1、对偶可行（dual feasible）的概念

当 $g(\lambda, \nu) = -\infty$ 时，上述不等式给出了一个平凡的结论。因此约定对偶可行域：

$$\text{dom } g = \{(\lambda, \nu) \mid \lambda \succeq 0 \text{ and } g(\lambda, \nu) > -\infty\}$$

注意到上式用了记号 $\lambda \succeq 0$ ，因为 λ 是个向量，它等价于 $\lambda_i \geq 0, \quad i = 1, \dots, m$ 。

2、弱对偶性

由于 $g(\lambda, \nu) \leq p^*$ 对任意的 $\lambda \succeq 0$ 和 ν 都成立，故

$$\max_{\lambda \succeq 0, \nu} g(\lambda, \nu) \leq p^*$$

亦成立，这也是下面要提到的弱对偶性的概念。

二、对偶问题

对偶问题

在上一节中，讲到了对原问题最优值的下界估计，其中重要的一个不等式是 $g(\lambda, \nu) \leq p^*$ ，其中 $g(\lambda, \nu)$ 是原问题的（拉格朗日）对偶函数。

那么问题来了，既然对偶函数是 p^* 的下界估计，那么可以取到的**最优**下界是什么？

这就导出了另一个优化问题：

•

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

称之为原问题（primal problem）的**对偶问题**（Lagrange dual problem）。

此时，上一节提到的**对偶可行**（dual feasible）的概念也好理解了，即 (λ, ν) 对该对偶问题是可行的。同时，把该问题的最优解 (λ^*, ν^*) 称为**对偶最优解**（dual optimal）或最优拉格朗日乘子（optimal Lagrange multipliers）。

值得注意的是，原问题的对偶问题**总是**凸优化问题，与原问题是否是凸优化问题无关。

弱对偶性

我们记上述对偶问题的最优解为 d^* ，由性质1，便很容易推出：

$$d^* \leq p^*$$

这便是**弱对偶性**（weak duality），即对偶问题的最优解不超过原问题的最优解。

形象解释——跷跷板

对于约束条件 $h_i(x)$ 对应的乘子 ν_i ，其作用相当于把跷跷板任意旋转，支点是使 $h_i(\tilde{x}) = 0$ 的那个 \tilde{x} ，其值对应于原问题的最优解 p^* 。（将跷跷板置入坐标系，横轴代表 x ，纵轴代表函数值，改变 ν 的值相当于“旋转”跷跷板。）这种情况下跷跷板的高度对应于函数 $L(\lambda, \nu)$ ，其最低点则是 $g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu)$ 。而根据常识，跷跷板的最低点是不会超过它的支点的。即使对于一个任意形状的跷跷板，这个性质也依旧成立。

类比一下，这就是说，**在任何情况下，弱对偶性总是成立的。**

对于约束条件 $f_i(x)$ 对应的乘子 λ_i ，其作用相当于把跷跷板往下拉，支点是使 $f_i(\tilde{x}) = 0$ 的那个 \tilde{x} ，最优解（最小值点 p^* ）可能位于支点处，也可能处于支点的下方，更可能是下无界的（ $p^* = -\infty$ ）。这种情况下，由于跷跷板的单向运动，最小值会变得更低，因此以上性质依旧成立。

强对偶性

如果原问题与对偶问题的最优解相等，即等式 $d^* = p^*$ 成立，则称它满足**强对偶性**。

强对偶性一般而言是不成立的。但若原问题是凸优化问题，即在形式

•

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & Ax = b \end{array}$$

中 f_0, \dots, f_m 都是凸函数。则强对偶性 **通常** 是成立的。

之所以说通常，是因为可能存在两个问题都无可行解的状况，这时强对偶性**不成立**。所以有必要建立“使其有解”的约束条件，其中之一便是 **Slater's condition**：

- **Slater's condition** 存在一点 $x \in \text{relint } \mathcal{D}$ （ \mathcal{D} 的相对内点集）满足

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

这样的点也可称之为 **严格可行**的。

Slater's Theorem则给出了强对偶性成立的一个充分条件：

- **定理3 Slater's Theorem** 若 Slater's condition 成立且原问题是凸优化问题，则强对偶性成立。

当有部分约束条件是仿射函数时，Slater's condition 可以被弱化。假定前 k 个约束条件是仿射函数，那么可以弱化为：

- **Refined Slater's condition** 若 f_1, \dots, f_k 是仿射函数，则存在一点 $x \in \text{relint } \mathcal{D}$ （ \mathcal{D} 的相对内点集）满足

$$f_i(x) \leq 0, \quad i = 1, \dots, k, \quad f_i(x) < 0, \quad i = k + 1, \dots, m, \quad Ax = b$$

此外，满足 Slater's condition（或 Refined Slater's condition）不仅意味着（凸优化问题）强对偶性的成立，而且也表示当 $d^* > -\infty$ 时，存在一组对偶变量 (λ^*, ν^*) 满足 $g(\lambda^*, \nu^*) = d^* = p^*$ ，即此时对偶最优值是可取到的。

原问题与对偶问题间的联系

我们可以把原问题与对偶问题的对应关系作一个总结：

1. 原问题和对偶问题都是可行的，弱对偶性成立，强对偶性不一定成立。
2. 原问题和对偶问题都不可行，此时弱对偶性依旧成立，但强对偶性不成立。
3. 当原问题是下无界时，即 $p^* = -\infty$ ，由弱对偶性， $d^* = -\infty$ ，即对偶问题不可行。
4. 反之，当对偶问题上无界时， $d^* = +\infty$ ，必有 $p^* = +\infty$ ，即原问题不可行。

可以把以上规则整理成如下表格：

对偶问题 \ 原问题	可行	下无界 ($p^* = -\infty$)	不可行($p^* = +\infty$)
可行	√	*	*
上无界 ($d^* = +\infty$)	*	*	√
不可行 ($d^* = -\infty$)	*	√	√

如上表格可知，实际上只有4种情形可能发生。

三、补充内容：对偶问题的求法

给定一个原问题，如何求出它对偶问题的形式呢？这一节内容中，总结了几种求对偶问题的方法。

1、由定义求驻点

这一种方法比较直接，其思路是由 $g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu)$ 的定义，直接求最值。这种方法可用于 $L(x, \lambda, \nu)$ 的表达式比较简单的情形。举例如下：

考虑如下问题：

•

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

则该问题的拉格朗日量是 $L(x, \nu) = x^T x + \nu^T (Ax - b)$ ，对偶函数是 $g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu)$ 。因为 $L(x, \nu)$ 是关于 x 的二次凸函数，对 x 求导：

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0$$

因此推出 $x = -(1/2)A^T \nu$ ，代入对偶函数表达式，即得：

$$g(\nu) = L(-(1/2)A^T \nu, \nu) = -(1/4)\nu^T A A^T \nu - b^T \nu$$

2、隐式求解约束条件

这种方法的思路是先求出拉格朗日量 $L(x, \lambda, \nu)$ ，然后求出对偶可行域：即要使

$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) > -\infty$ 成立， (λ, ν) 所需满足约束条件。这些约束条件加上 $\lambda \succeq 0$ 便构成了对偶问题约束条件。

考虑标准形式的线性规划（Linear Programming）问题：

•

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

这时拉格朗日量是：

$$L(x, \nu) = c^T x - \sum_{i=1}^m \lambda_i x_i + \nu^T (Ax - b) = -b^T \nu + (A^T \nu + c - \lambda)^T x$$

注意到当 $A^T \nu + c - \lambda \neq 0$ 时， $g(\lambda, \nu) = -\infty$ ，因此：

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

于是，该问题的对偶问题可表示成如下形式：

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

3、共轭函数法

共轭函数法给出了从（线性约束条件下的）原问题到对偶问题的一般化形式，同时也揭示了对偶函数与共轭函数间的关系。

先给出共轭函数的定义：

- 一个函数 $f: R^n \rightarrow R$ 的共轭函数 f^* 定义如下：

$$f^*(y) = \sup_{x \in \text{dom } f} (y^T x - f(x))$$

接着考虑如下有线性等式和线性不等式约束的优化问题：

•

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

于是，由对偶函数及共轭函数的定义可得：

$$\begin{aligned}g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} (f_0(x) + \lambda^T(Ax - b) + \nu^T(Cx - d)) \\ &= -b^T \lambda - d^T \nu + \inf_{x \in \mathcal{D}} (f_0(x) + (A^T \lambda + C^T \nu)^T x) \\ &= -b^T \lambda - d^T \nu - f_0^*(-A^T \lambda - C^T \nu)\end{aligned}$$

且对偶函数 g 的定义域由共轭函数 f^* 的定义域所确定，即：

$$\text{dom } g = \{(\lambda, \nu) \mid -A^T \lambda - C^T \nu \in \text{dom } f_0^*\}$$

于是，求对偶函数的问题便转换成了求共轭函数的问题。

下面列举一些常用目标函数的共轭函数：

- **线性函数** $f(x) = ax + b$ 的共轭函数是 $f^*(a) = -b$ ，且 f^* 的定义域是单点 $\{a\}$ 。
- **负对数** $f(x) = -\log x$ 的共轭函数是 $f^*(y) = -\log(-y) - 1$ ，定义在 $y < 0$ 。
- **指数函数** $f(x) = e^x$ 的共轭函数是 $f^*(y) = y \log y - y$ ，定义在 $y \geq 0$ 。
- **负熵函数** $f(x) = x \log x$ 的共轭函数是 $f^*(y) = e^{y-1}$ ，定义域为 R 。
- **Hinge Loss** $f(x) = \max\{0, 1 - x\}$ 的共轭函数是 $f^*(y) = y$ ，定义域是 $-1 \leq y \leq 0$ 。
- **任意范数** $f(x) = \|x\|$ 的共轭函数是 $f^*(y) = \begin{cases} 0 & \|y\|_* \leq 1 \\ \infty & \text{otherwise} \end{cases}$

其中 $\|y\|_*$ 是原范数 $\|x\|$ 的对偶范数。

- **范数平方** $f(x) = (1/2)\|x\|^2$ 的共轭函数是 $f^*(y) = (1/2)\|y\|_*^2$ ，其中 $\|y\|_*$ 是原范数 $\|x\|$ 的对偶范数。

用这种方法可以很快捷地推出上面两种方法中提到的例子，可自行检验之。

- **二次型** $f(x) = (1/2)x^T Q x$ 的共轭函数是 $f^*(y) = (1/2)y^T Q^{-1} y$ 。

TODO：需要推导 $\|Ax - b\|$ 的共轭函数。

四、最优条件（Optimality conditions）

这一节中，我们关注最优解（包括原问题和对偶问题的）所需要满足的条件。并且除非特殊提及，均不假定这些问题是凸的。

松弛互补（Complementary slackness）

假设强对偶性成立，则原问题与对偶问题的解相等。记 x^* 和 (λ^*, ν^*) 分别表示原问题与对偶问题的最优解，那么，

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

故上述不等号皆可取等，这其中蕴含着两个重要的结论，一是“ x^* 最小化了 $L(x, \lambda^*, \nu^*)$ 的值”，这个结论会在下一小节中用到，另一个重要结论是：

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

又因为 $\lambda_i^* \geq 0$, $f_i(x^*) \leq 0$, 所以,

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

这个条件被称为**互补松弛条件** (Complementary slackness)。它对原问题的 任意的最优解 x^* 和对偶问题最优解 d^* 皆成立。

这个条件实际上描述的是：

1. 如果对偶变量的某个分量非零 $\lambda_i^* > 0$, 则该分量对应的约束条件严格取等号, 即 $f_i(x^*) = 0$ 。
2. 如果某约束条件不严格取等号, 即 $f_i(x^*) < 0$, 则对应的对偶变量分量一定为零, 即 $\lambda_i^* = 0$ 。

- **注：**这个条件也可以理解为某种相态的变化。

以水凝固成冰为例，考虑水的温度和冰占冰水混合物的比例两个变量。当水温逐渐下降时，只要温度高于凝固点(温度 $>0^\circ\text{C}$)，冰的比例始终是(比例 $=0\%$)；反之，当出现冰水混合物(比例 $>0\%$)，其温度一定是(温度 $=0^\circ\text{C}$)。这种性质与上述互补松弛条件是很相似的。

KKT条件

在继续之前，需要假定函数 $f_0, \dots, f_m, h_1, \dots, h_p$ 是一阶可微的，因此定义域是开集。并且对问题的凸性仍无任何假定。

必要性（对任意问题）

还是记 x^* 和 (λ^*, ν^*) 分别表示原问题与对偶问题的最优解。则 x^* 最小化了 $L(x, \lambda^*, \nu^*)$ （这是上一小节的一个结论），从而梯度为 0，即有：

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

因此，我们有，

- **KKT条件**

$$\begin{aligned}
f_i(x^*) &\leq 0, & i &= 1, \dots, m \\
h_i(x^*) &= 0, & i &= 1, \dots, p \\
\lambda_i^* &\geq 0, & i &= 1, \dots, m \\
\lambda_i^* f_i(x^*) &= 0, & i &= 1, \dots, m \\
\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(x^*) &= 0
\end{aligned}$$

这便是**KKT条件**（*Karush-Kuhn-Tucker condition*）。对任何优化问题而言，最优解一定满足KKT条件。

- **KKT条件必要性** 对于任何优化问题，只要 x^* 和 (λ^*, ν^*) 分别是原问题与对偶问题的最优解，则 (x^*, λ^*, ν^*) 一定满足KKT条件。即KKT条件是使一组解成为最优解的**必要条件**。

充分性（对凸优化问题）

既然任意的最优解都满足KKT条件，那么自然要问，满足KKT条件的解是否就是问题的最优解呢？

实际上，当问题是凸优化问题时，KKT条件便是充分的了。简单证明如下：

设 $(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ 满足KKT条件。因为 $\lambda_i \geq 0$ ，所以 $L(x, \tilde{\lambda}, \tilde{\nu})$ 是关于 x 的凸函数。由KKT最后一个条件知， \tilde{x} 最小化了 $L(x, \tilde{\lambda}, \tilde{\nu})$ 的值。因此，

$$\begin{aligned}
g(\tilde{\lambda}, \tilde{\nu}) &= L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \\
&= f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i h_i(\tilde{x}) \\
&= f_0(\tilde{x})
\end{aligned}$$

因此对偶差（duality gap）为0，解分别是原问题最优和对偶最优的。

小结一下，这就是说，对于任何目标函数和约束函数都一阶可微的凸优化问题，任何满足KKT条件的 x^* 和 (λ^*, ν^*) ，其对偶差为0。此时若 Slater's condition 满足，则说明这个对偶最优值可以被取到，因而它就是问题的最优解了。

- **KKT条件充分性** 对于任何目标函数和约束函数都一阶可微的凸优化问题，且满足 Slater's condition，则 x 和 (λ, ν) 分别是原问题与对偶问题的最优解，当且仅当 (x, λ, ν) 满足KKT条件。即KKT条件是使一组解成为这类问题最优解的**充分（也是必要）条件**。

KKT条件的作用在于，某一类问题的KKT条件可以直接被（分析地）解出。而对大多数其他凸优化问题，便把任务由解优化问题转移到解KKT条件上，因为两者是等价的。

实例：注水算法

这里举一个例子来说明KKT条件在处理凸优化问题中的作用，考虑如下的凸优化问题：

-

$$\begin{aligned} & \text{minimize} && -\sum_{i=1}^n \omega_i \log(\alpha_i + x_i) \\ & \text{subject to} && x \succeq 0, \quad \mathbf{1}^T x = 1 \end{aligned}$$

其中 $\omega_i > 0, \alpha_i > 0$ 。当 $\omega_i = 1$ 时，这个问题的背景是信息论中根据不同的信道状况去分配不同的发送功率，来达到最大化传输效率。变量 x_i 表示分配给第 i 个信道的功率强弱，而 $\log(\alpha_i + x_i)$ 则表示该信道的容量或者传输速率，最后当然是总传输速率越大越好。

为了求出其对偶问题的形式，我们引入拉格朗日乘子 $\lambda \in R^n$ 对应于约束 $x \succeq 0$ ，引入 $\nu \in R$ 对应于约束 $\mathbf{1}^T x = 1$ ，由KKT条件，可以得到一系列关系式：

$$x^* \succeq 0, \quad \mathbf{1}^T x^* = 1, \quad \lambda^* \succeq 0, \quad \lambda_i^* x_i^* = 0$$

以及，

$$-\frac{\omega_i}{\alpha_i + x_i^*} - \lambda_i^* + \nu^* = 0$$

注意到上述最后一个等式， λ_i 在其中实际起到了 松弛变量 的作用，故可把它等价写成：

$$\alpha_i + x_i^* \geq \frac{\omega_i}{\nu^*}, \quad i = 1, \dots, n$$

这时 $\lambda_i = 0$ 便对应于取等号的情形。下面分情况讨论：

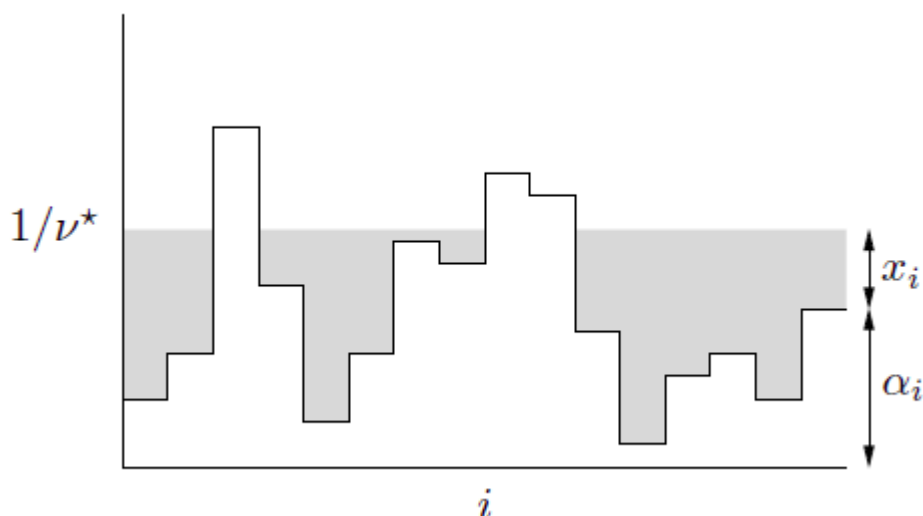
1. 当 $\frac{\omega_i}{\nu^*} \leq \alpha_i$ 时，一定有 $x_i^* = 0$ 。若不然 $x_i^* > 0$ ，则不等式严格成立，即 $\lambda_i^* > 0$ ，这违背了松弛互补条件。
2. 当 $\frac{\omega_i}{\nu^*} > \alpha_i$ 时，有 $x_i > 0$ ，由松弛互补条件， $\lambda_i = 0$ ，这意味着不等式等号取到，即推出 $x_i^* = \frac{\omega_i}{\nu^*} - \alpha_i$

综上所述， $x_i^* = \max\{0, \frac{\omega_i}{\nu^*} - \alpha_i\}$ 。将其代入条件 $\mathbf{1}^T x = 1$ ，我们得到：

$$\sum_{i=1}^n \max\{0, \frac{\omega_i}{\nu^*} - \alpha_i\} = 1$$

等式左端是一个关于变量 $1/\nu^*$ 的单调增分段线性函数，在 α_i/ω_i 处有间断点。因此这个方程显然有唯一解。

当 $\omega_i = 1$ 时，解这个方程的方法一般称之为**注水算法**。设想有 n 个平台，每个平台的高度为 α_i ，将其注水至统一高度 $1/\nu^*$ ，则总共需要的水量即为 $\sum_{i=1}^n \max\{0, \frac{\omega_i}{\nu^*} - \alpha_i\}$ 。假设总水量是“单位1”，那么第 i 个平台上水的高度恰好就是 x_i^* 。如下图所示：



注水算法图示。每个平台的高度是 α_i ，整个容器的水位是 $1/\nu^*$ ，每个平台水的高度（以阴影显示）便是最优解 x_i^* 。

如果 $\omega_i \neq 1$ ，那么解法会稍有变化，我将其称为 *改进的注水算法*。这时每个 ω_i 实际上代表了第 i 个容器中液体的密度，而 x_i 代表了第 i 个容器中液体的质量，总质量是守恒的。而 x_i/ω_i 代表了液体的体积(高度)，让 α_i/ω_i 代表了平台的高度，这样第 i 个容器中液体的质量（=密度*体积）便为

$$\omega_i \cdot \max\left\{0, \frac{1}{\nu^*} - \frac{\alpha_i}{\omega_i}\right\} = \max\left\{0, \frac{\omega_i}{\nu^*} - \alpha_i\right\}$$

因此所有容器中液体的总质量为“单位1”。这表示将一定质量的液体分配到不同的容器中，而每个容器由于自身环境不同会导致液体密度出现变化，但总体的水位高度要保持持平。这时水位的高度对应于对偶问题的解 $1/\nu^*$ ，每个容器中液体的质量对应于原问题的解 x_i^* 。

这个例子阐明了对于某一类问题，可以直接求解KKT条件，从而解决原问题。同样地，对偶变量一般（在物理意义上）代表着某种“相变”的过程。例如上例中的 λ_i ，当它非零时，表示容器中没有水；而容器中有水时，它一定为零。它可以看做一个 *状态的指示量*。而另一个对偶变量 ν ，它来源于等式约束。它通常代表着一些 *隐含的属性*，或者是另一种看待问题的视角。像本例中的“统一水位”，或者经济学中涉及到的“影子价格”，都属于此类。

五、总结

本文是对优化问题中对偶性理论相关内容的整理。第一、二节中，从拉格朗日量出发，构造对偶函数，继而引入对偶问题和弱对偶性，然后讨论强对偶性的充分性条件，以及原问题与对偶问题间的联系。第三节补充了一些常见原问题求其对偶形式的方式，并且引入共轭函数的概念给出了一大类线性约束条件的问题的通用对偶形式。第四节重点关注最优解满足的条件，包括松弛互补条件、KKT条件等。最后举了一个实例来说明KKT条件在凸优化问题中的作用。

参考资料

Boyd S, Vandenberghe L. [Convex optimization](#). Cambridge university press, 2004. Chapter 5.