线性规划(Linear Programming)

对偶(Duality)

我们已经注意到，问题的表述形式会严重影响问题的求解方法。有时候，一个问题存在对称的另一种“相反”方式的描述，这种表示在数学上被称为对偶表示（这里借用这个名称，真正的对偶在数学上都是有严格定义的）。有时将一个问题转成对偶形式，能更方便地求解。

以下考虑问题(12.1)的一个特例：$\mathcal{E}=\phi$，且目标函数f和$c_i(x)$都是凸的，$i\in\mathcal{I}$。也即(12.1)现在可以写成

$$\min f(x), \quad \text{s. t. } c(x)\geq 0, \tag{12.81}$$

这里

$$c(x):=[c_1(x),c_2(x),\cdots,c_m(x)]^T, \mathcal{I}=\{1,2,\cdots,m\}.$$

对应的Lagrange乘子函数为：

$$\mathcal{L}(x,\lambda)=f(x)-\lambda^T c(x).$$

现定义对偶目标函数$q:\mathbb{R}^n\mapsto\mathbb{R}$，

$$q(\lambda):=\inf_x \mathcal{L}(x,\lambda), \tag{12.82}$$

由于对某些λ，$q(\lambda)$可以取到$-\infty$，因此我们限制q的定义域为：

$$\mathcal{D}(q):=\{\lambda\,|q(\lambda)>-\infty\}. \tag{12.83}$$

（原问题我们一般是先确立x，然后判定λ。而在对偶情形下，我们先确定λ，然后再判定x。不论是原问题还是对偶，二者通过$\mathcal{L}(x,\lambda)$建立对应。）

注意当f和$c_i(x)$是凸的且$\lambda\geq 0$时，$\mathcal{L}(\cdot,\lambda)$也是凸的。因此任何局部极值都是其全局极值。定义问题(12.81)的对偶问题如下：

$$\max_{\lambda\in\mathbb{R}^n} q(\lambda), \quad \text{s. t. } \lambda\geq 0. \tag{12.84}$$

例 12.10

$$\min_x 0.5(x_1^2+x_2^2), \quad \text{s. t. } x_1-1\geq 0. \tag{12.85}$$

该问题的Lagrange乘子函数为

$$\mathcal{L}(x,\lambda)=0.5(x_1^2+x_2^2)-\lambda_1(x_1-1),$$

对给定的λ_1， 有$\mathcal{L}(x,\lambda_1)$是凸的，故其稳定点：

$$\left\{\begin{array}{rcl}\dfrac{\partial\mathcal{L}(x,\lambda_1)}{\partial x_1} &=& x_1-\lambda=0\\ \\ \dfrac{\partial\mathcal{L}(x,\lambda_1)}{\partial x_2} &=& x_2=0\end{array}\right. \Rightarrow \left\{\begin{array}{rcl}x_1 &=& \lambda_1\\ x_2 &=& 0\end{array}\right.$$

就是其全局极小值点，故

$$
\begin{aligned}
q(\lambda_1) &= \inf_x \mathcal{L}(x, \lambda_1) = \mathcal{L}(\lambda_1, 0, \lambda_1) \\
&= 0.5(x_1^2 + x_2^2) - \lambda_1(x_1 - 1)\big|_{x_1=\lambda_1, x_2=0} \\
&= 0.5(\lambda_1^2) - \lambda_1(\lambda_1 - 1) \\
&= -0.5\lambda_1^2 + \lambda_1.
\end{aligned}
$$

因此对偶问题为：

$$
\max_{\lambda_1 \geq 0} -0.5\lambda_1^2 + \lambda_1, \tag{12.86}
$$

解为$\lambda_1^* = 1$，代回有$x^* = (1, 0)^T$。（对偶问题和原问题等解。）

定理 12.10 q是凹(concave)的，$\mathcal{D}$是凸的。

证明 $\forall \lambda^0, \lambda^1 \in \mathbb{R}^n$，$x \in \mathbb{R}^n$，$\alpha \in [0, 1]$，有

$$
\mathcal{L}(x, (1-\alpha)\lambda^0 + \alpha\lambda^1) = (1-\alpha)\mathcal{L}(x, \lambda^0) + \alpha\mathcal{L}(x, \lambda^1),
$$

两边同时取$\inf$，并用不等式

$$
\inf(a + b) \geq \inf a + \inf b,
$$

有

$$
q((1-\alpha)\lambda^0 + \alpha\lambda^1) \geq (1-\alpha)q(\lambda^0) + \alpha q(\lambda^1),
$$

即q是凹的（和凸不等式反向）。而对$\lambda^0, \lambda^1 \in \mathcal{D}$，则上式右端$> -\infty$，而上式左端亦$> -\infty$，于是

$$
(1-\alpha)\lambda^0 + \alpha\lambda^1 \in \mathcal{D},
$$

即$\mathcal{D}$是凸的。

从上述证明可以看到，对偶问题的目标函数，事实上是原问题目标函数的下界。这个事实总结为以下定理：

定理 12.11(弱对偶) 任取(12.81)的可行点$\bar{x}$，以及$\bar{\lambda} \geq 0$，有$q(\bar{\lambda}) \leq f(\bar{x})$。

证明：略。

在(12.81)的形式下，KKT条件有如下形式：

$$
\begin{aligned}
\nabla f(\bar{x}) - \nabla c(\bar{x})\bar{\lambda} &= 0, && (12.87\text{a}) \\
c(\bar{x}) &\geq 0, && (12.87\text{b}) \\
\bar{\lambda} &\geq 0, && (12.87\text{c}) \\
\bar{\lambda}_i c_i(\bar{x}) &= 0, i = 1, 2, \cdots, m. && (12.87\text{d})
\end{aligned}
$$

其中$\nabla c(x)$表示$n \times m$矩阵

$$
\nabla c(x) = \left[\nabla c_1(x), \nabla c_2(x), \cdots, \nabla c_m(x)\right].
$$

定理12.12 若$\bar{x}$是问题(12.18)的解，且f和$-c_i$均凸并在$\bar{x}$点连续可微，$i = 1, 2, \cdots, m$，则对任何满足KKT条件的$(\bar{x}, \bar{\lambda})$，$\bar{\lambda}$是(12.84)的解。

定理12.13 假设f和$-c_i$，$i = 1, 2, \cdots, m$是凸的且连续可微。若$\bar{x}$是(12.81)的解且LICQ成立。假设$\hat{\lambda}$是(12.84)的解，且$\inf_x \mathcal{L}(x, \lambda)$在$\hat{x}$取到。若$\mathcal{L}(\cdot, \hat{\lambda})$是严格凸的，则$\bar{x} = \hat{x}$，$f(\bar{x}) = \mathcal{L}(\hat{x}, \hat{\lambda})$。

以上两个定理证明略，它们说明可以在必要的时候用对偶问题来代替原问题求解。而在计算上，我们讲对偶问题整理成更方便的形式，称为Wolfe对偶：

$$
\max_{x, \lambda} \mathcal{L}(x, \lambda), \quad \text{s. t. } \nabla_x \mathcal{L}(x, \lambda) = 0, \lambda \geq 0. \tag{12.88}
$$

这种转换只需要机械地计算。

定理 12.14 假设f和$-c_i$，$i=1,2,\cdots,m$是凸的并连续可微。若$(\bar{x},\bar{\lambda})$是(12.81)的解，且有LICQ成立。则$(\bar{x},\bar{\lambda})$也是(12.88)的解。

例 12.11（线性规划，Linear Programming）

$$\min c^T x, \text{s. t. } Ax-b\geq 0. \tag{12.89}$$

对偶目标为：

$$q(\lambda)=\inf_x\left[c^T x-\lambda^T(Ax-b)\right]=\inf_x\left[(c-A^T\lambda)^T x+b^T\lambda\right],$$

若$c-A^T\lambda\neq 0$，则$q(\lambda)=-\infty$（无界）。故由定义，$q(\lambda)$需满足$A^T\lambda=c$。而此时$q(\lambda)=b^T\lambda$，对偶问题为：

$$\max_\lambda b^T\lambda, \quad \text{s. t. } A^T\lambda=c,\lambda\geq 0. \tag{12.90}$$

直接计算可得，对应的Wolfe对偶为

$$\max_{\lambda,x} c^T x-\lambda^T(Ax-b), \quad \text{s. t.} A^T\lambda=c,\lambda\geq 0.$$

和(12.90)是等价的。对于A的某些形式，(12.90)比(12.89)更易求解。

例 12.12（凸二次规划，Convex Quadratic Programming）

$$\min \frac{1}{2}x^T Gx+c^T x, \quad \text{s. t. } Ax-b\geq 0, \tag{12.91}$$

其中G正定。对偶目标函数为

$$q(\lambda)=\inf_x\mathcal{L}(x,\lambda)=\inf_x\frac{1}{2}x^T Gx+c^T x-\lambda^T(Ax-b), \tag{12.92}$$

由G对称正定，$\mathcal{L}(x,\lambda)$严格凸，故inf在$\nabla_x\mathcal{L}(x,\lambda)=0$，即

$$Gx+c-A^T\lambda\Rightarrow x=G^{-1}(A^T\lambda-c). \tag{12.93}$$

由上式，得

$$\begin{aligned}
q(\lambda)&=x^T Gx+c^T x-\lambda^T Ax+\lambda^T b-\frac{1}{2}x^T Gx\Big|_{Gx+c-A^T\lambda=0}\\
&=-\frac{1}{2}(A^T\lambda-c)^T G^{-1}GG^{-1}(A^T\lambda-c)+b^T\lambda\\
&=-\frac{1}{2}(A^T\lambda-c)^T G^{-1}(A^T\lambda-c)+b^T\lambda.
\end{aligned}$$

同样，我们可以给出Wolfe对偶：

$$\max_{\lambda,x} -\frac{1}{2}x^T Gx+\lambda^T b, \quad \text{s. t. } Gx+c-A^T\lambda=0,\lambda\geq 0. \tag{12.95}$$

这里事实上G只需半正定。

以上都只是引言，我们现在真正将注意力集中到线性规划上来。这是一类最简单，也是最重要的优化模型，同时也是一种最基本的建模思路。它的目标函数和约束，全部都是线性的。

例：如何找对象？

我们如何用一个简单的线性模型来表达这么复杂的一个主题呢？

$$\min f(x)=-c^T x, \quad \text{s. t. } Ax\geq b.$$

也即

$$
\begin{aligned}
\min \quad & f(x) = -\sum_{i=1}^{n} c_i x_i, \\
\text{s. t.} \quad & \begin{cases}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & \geq & b_1, \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & \geq & b_2, \\
 & \cdots & \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n & \geq & b_n.
\end{cases}
\end{aligned}
$$

我们可以用这样一张表格来表达我们的建模思路：

变量	**权重**	**目标成份**
x_1	c_1	性格
x_2	c_2	颜值
$\vdots$	$\vdots$	$\vdots$
x_{n-1}	c_{n-1}	学历
x_n	c_n	财产

我们马上体会到，这样的建模的优点：

- 简单粗暴直接。你只须确定一下c和b，立刻就把你的需求暴露无疑。比如：

  $$c = [0.5, 0.3, \cdots, 0.8, 100.0]^T$$

  而b就代表了你的接受下界；
- 方便求解。线性规划有可靠的理论和数值求解算法，能够确保得到正确的解（或者明确无解）；
- 能够表达复杂需求。

当然它的上述优点同时也可以认为是它的缺点。比如我们的例子问题，你真的打算尝试么？不过在大量的实际生产和研究领域，它都提供了一种可操作可供参考的建设性模型工具。

为了能进一步方便求解，我们一般也会对线性规划如何书写，提出规范要求，称做"标准形式"。注意，标准形式也是因书而异，因软件而异的。本书本课程约定的线性规划标准形式为：

$$\min_{x \in \mathbb{R}^n} f(x), \quad \text{s.t. } Ax = b, x \geq 0. \tag{13.1}$$

其中A是$m \times n$矩阵，m是等值约束个数。

首先讨论如何将一般形式

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

转换成标准形式。对于一般的不等值约束$Ax \leq b$，我们先增加松弛变量z，将其变为：

$$Ax \leq b \Rightarrow Ax + z = b, z \geq 0, \tag{13.2}$$

新增的变量z是满足$z \geq 0$的标准化要求的，但之前的x并没有，为此，对x做分裂：

$$x = x^+ - x^-, x^+ = \max(x, 0), x^- = \max(-x, 0),$$

而将原来目标函数写成：

$$
\begin{aligned}
f(x) &= \sum_{i=1}^{n} c_i x_i + 0 \cdot z \\
&= \sum_{i=1}^{n} c_i (x_i^+ - x_i^-) + 0 \cdot z \\
&= \sum_{i=1}^{n} c_i x_i^+ + \sum_{i=1}^{n} (-c_i) x_i^- + 0 \cdot z,
\end{aligned}
$$

重新整理一下，现令

$$
x = [x^+ \quad x^- \quad z]^T, c = [c \quad -c \quad 0]^T, A = [A \quad -A \quad I],
$$

则一般问题就转变成等解的标准形式：

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

例

$$
\begin{cases}
\min & x_1 - 2x_2 + 5x_3, \\
\text{s.t.} & x_1 - x_3 \geq 6, \\
& 2x_2 + x_3 \leq 2, \\
& x_1, x_2 \geq 0.
\end{cases}
$$

化为标准形式：

引入$x_4 \geq 0$，使

$$
x_1 - x_3 \geq 6 \Leftrightarrow x_1 - x_3 - x_4 = 6.
$$

引入$x_5 \geq 0$，使

$$
2x_2 + x_3 \leq 2 \Leftrightarrow 2x_2 + x_3 + x_5 = 2.
$$

引入$x_6, x_7 \geq 0$，并令

$$
x_3 = x_6 - x_7,
$$

则原问题化为：

$$
\begin{cases}
\min & x_1 - 2x_2 + 5x_6 - 5x_7, \\
\text{s.t.} & x_1 - x_4 - x_6 + x_7 = 6, \\
& 2x_2 + x_6 - x_7 + x_5 = 2, \\
& x_1, x_2, x_4, x_5, x_6, x_7 \geq 0.
\end{cases}
$$

我们对x的各分量做一个重排：

$$
x_4 \to x_3, x_5 \to x_4, x_6 \to x_5, x_7 \to x_6,
$$

其余分量不变，则原问题整理成标准形式：

$$
\begin{cases}
\min & x_1 - 2x_2 + 5x_5 - 5x_6, \\
\text{s.t.} & x_1 - x_3 - x_5 + x_6 = 6, \\
& 2x_2 + x_4 + x_5 - x_6 = 2, \\
& x \geq 0.
\end{cases}
\Leftrightarrow
\begin{cases}
\min & f(x) = c^T x, \\
\text{s.t.} & Ax = b, \\
& x \geq 0.
\end{cases}
$$

其中，

$$
\begin{gathered}
c = (1, -2, 0, 0, 5, -5)^T, \quad x = (x_1, x_2, \cdots, x_6)^T, \quad b = (6, 2)^T, \\
A = \begin{bmatrix} 1 & 0 & -1 & 0 & -1 & 1 \\ 0 & 2 & 0 & 1 & 1 & -1 \end{bmatrix}.
\end{gathered}
$$

我们可以写出(13.1)的Lagrange乘子函数，

$$\mathcal{L}(x,\lambda,s)=c^Tx-\lambda^T(Ax-b)-s^Tx. \tag{13.3}$$

这里λ是针对等值约束的m维Lagrange乘子向量，s则是针对不等值约束$x\geq 0$的n维Lagrange乘子。对线性规划应用定理12.1，就得到x^*是问题(13.1)的解的必要条件为：

$$\begin{aligned}
A^T\lambda+s&=c, &(13.4\text{a})\\
Ax&=b, &(13.4\text{b})\\
x&\geq 0, &(13.4\text{c})\\
s&\geq 0, &(13.4\text{d})\\
x_is_i&=0, i=1,2,\cdots,n. &(13.4\text{e})
\end{aligned}$$

令(x^*,λ^*,s^*)表示一个满足条件(13.4)的向量三元组，由(13.4a)，(13.4d)和(13.4e)有

$$c^Tx^*=(A^T\lambda^*+s^*)^Tx^*=(Ax^*)^T\lambda^*=b^T\lambda^*. \tag{13.5}$$

在线性规划情形，容易看出(13.4)也是有解的充分条件（问题也只有一阶），即满足(13.4)的x^*必是问题(13.1)的全局最优解。令$\bar{x}$是任何可行点，则

$$A\bar{x}=b,\quad \bar{x}\geq 0.$$

于是

$$c^T\bar{x}=(A\lambda^*+s^*)^T\bar{x}=b^T\lambda^*+\bar{x}^Ts^*\geq b^T\lambda^*=c^Tx^*. \tag{13.6}$$

也即不会有比x^*取值更小的可行点。并且由(13.6)取等号的条件我们还可以看到，若可行点$\bar{x}$是问题(13.1)的全局最优点，当且仅当

$$\bar{x}^Ts^*=0.$$

线性规划问题的对偶问题是非常简单干净的，问题(13.1)的对偶问题为：

$$\max b^T\lambda,\quad \text{s.t. } A^T\lambda\leq c. \tag{13.7}$$

现在变量变成了λ。相应地，(13.1)被称为原问题(primal)。这个问题也可以引入“对偶松弛变量”(dual slack variables)s，将其改写成类似的标准形式：

$$\max b^T\lambda,\quad \text{s.t.} A^T\lambda+s=c, s\geq 0. \tag{13.8}$$

在这里，(λ,s)一起成为新问题的变量，有时也被称为对偶变量。

对偶问题和原问题其实是同一个问题的不同建模方式，它们在Lagrange乘子函数和解的条件上达成了一致。这里我们不再过多讨论细节，有兴趣的同学自己看书第361页后剩下的部分。我们直接用一个定理来总结一下：

定理 13.1（强对偶）

1. 若(13.1)和(13.7)其中之一有有限解，则另一个也有有限解，且解和优化目标值相同；
2. 若(13.1)和(13.7)其中之一的解无界，则另一个必不可行（无解）。