Lagrange函数及其对偶函数

对于标准形式的优化问题：

$$
\begin{aligned}
& minimize\ f_0(x) \\
& subject\ to\ f_i(x) \leq 0, i=1,2,\ldots,m \\
& \qquad\qquad h_i(x)=0, i=1,2,\ldots,p
\end{aligned}
$$

拉格朗日函数为：

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

向量λ和v是对偶变量或者是这个式子的Lagrange乘子

Lagrange对偶函数是Lagrange函数关于x取得的最小值：

$$
\begin{aligned}
g(\lambda,v) &= \inf_{x\in D} L(x,\lambda,v) \\
&= \inf_{x\in D}(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}v_i h_i(x))
\end{aligned}
$$

对偶函数是一族关于(λ, v)的放射函数的逐点下确界，所以即使原问题不是凸的，对偶函数也是凹的

对偶函数构成了原问题最优值p^*的下界：即对任意$\lambda \geq 0$和v，下式成立：

$$
g(\lambda,v)\leq p^*
$$

证明：
设$\hat{x}$是原问题的一个可行点，即$f_i(\hat{x})\leq 0, h_i(x)=0$，则有：

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

代入Lagrange函数有：

$$
\begin{aligned}
L(\hat{x},\lambda,v) &= f_0(\hat{x})+\sum_{i=1}^{m}\lambda_i f_i(\hat{x})+\sum_{i=1}^{p}v_i h_i(\hat{x}) \\
&\leq f_0(\hat{x})
\end{aligned}
$$

因此：

$$
g(\lambda,v)\leq L(\hat{x},\lambda,v)\leq f_0(\hat{x})
$$

考虑标准形式的线性规划问题：

$$
\begin{aligned}
& minimize\ c^T x \\
& subject\ to\ Ax=b \\
& \qquad\qquad x\geq 0
\end{aligned}
$$

其拉格朗日函数为：

$$
\begin{aligned}
L(x,\lambda,v) &= c^T x-\sum_{i=1}^{n}\lambda_i x_i+v^T(Ax-b) \\
&= -bv^T+(c+A^T v-\lambda)^T x
\end{aligned}
$$

对偶函数为：

$$
\begin{aligned}
g(\lambda,v) &= \inf_{x} L(x,\lambda,v) \\
&= -b^T v+\inf_{x}(c+A^T v-\lambda)^T x
\end{aligned}
$$

线性函数只有恒为0时才有下界，因此：

$$
g(\lambda,v)=\begin{cases}-b^T v & c+A^T v-\lambda=0 \\ -\infty & 其它情况\end{cases}
$$

Lagrange对偶问题

Lagrange对偶函数给出了优化问题最优值p^*的一个下界，一个自然的问题是：从Lagrange函数能够得到的最好下界是什么？可以将这个问题表述为优化问题：

$$
\begin{aligned}
& maxmize\ g(\lambda,v) \\
& subject\ to\ \lambda\geq 0
\end{aligned}
$$

Lagrange对偶问题是一个凸优化问题，这是因为极大化的目标函数是凹函数，且约束集合是凸集

很多情况下，我们可以识别出函数g的定义域的仿射包，并将其表示为一系列线性等式约束，这样处理之后可以得到一个等价的问题

对于标准形式的线性规划，它的对偶问题是在满足约束$\lambda\geq 0$的条件下极大化对偶函数g。当且仅当$c+A^T v-\lambda=0$时对偶函数有界。因此，可得一个等价问题：

$$
\begin{aligned}
& maxmize\ \ -b^T v \\
& subject\ to\ c+A^T v-\lambda=0 \\
& \qquad\qquad \lambda\geq 0
\end{aligned}
$$

这个问题可进一步表述为：

$$
\begin{aligned}
& maxmize\ \ -b^T v \\
& subject\ to\ A^T v+c\geq 0
\end{aligned}
$$

KKT最优性条件

原问题最优值用p^*表示，Lagrange对偶问题的最优值用d^*表示，则有：

$$d^* \le p^*$$

即使原问题不是凸问题，这个不等式也成立，这个性质称为弱对偶性。当原问题很难求解时，弱对偶不等式可以给原问题最优值一个下界

如果等式$d^* = p^*$成立，即最优对偶间隙是零，那么强对偶性成立。一般情况下，强对偶性是不成立的。但如果原问题是凸问题，强对偶性一般成立

假设强对偶性成立，令x^*是原问题的最优解，(λ^*, v^*)是对偶问题的最优解，因此：

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

因此有：

$$\lambda_i^* f_i(x) = 0$$

这个条件称为**「互不松弛性」**

非凸问题的KKT条件

假设强对偶性成立，令x^*是原问题的最优解，(λ^*, v^*)是对偶问题的最优解。因为$L(x, \lambda^*, v^*)$处取得最小值，因此，函数在x^*处的导数必须为零，即：

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

由此可得：

$$\begin{aligned} f_i(x^*) &\le 0 \\ h_i(x^*) &= 0 \\ \lambda_i^* &\ge 0 \\ \lambda_i^* f_i(x^*) &= 0 \\ \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} v_i^* \nabla h_i(x^*) &= 0 \end{aligned}$$

上述式子就是**「KKT条件」**。如果强对偶性成立，那么任意一对原问题和对偶问题的最优解必须满足KKT条件

凸问题的KKT条件

当原问题是凸问题时，满足KKT条件的点也是原、对偶问题的最优解。换言之，$\hat{x}, \hat{\lambda}, \hat{v}$是满足KKT条件的点，那么$\hat{x}$和$(\hat{\lambda}, \hat{v})$分别是原问题和对偶问题的最优解，对偶间隙是零

附录

「凸集的定义」：如果集合C中任意两点间的线段仍然在C中，即对于任意$x_1, x_2 \in C$和$0 \le \theta \le 1$都有：

$$\theta x_1 + (1-\theta) x_2 \in C$$

那么集合C是凸集

「凸函数的定义」：函数$f: R^n \rightarrow R$的定义域是$dom\ \ f$，如果$dom\ \ f$是凸集，且对于任意$x, y \in dom\ \ f$和任意$0 \le \theta \le 1$，有：

$$f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y)$$

从几何意义上来看，上述不等式意味着点$(x, f(x))$和$(y, f(y))$之间的线段在函数f的上方

如果函数$-f$是凸的，那么函数f是凹的

「凸函数与逐点上确界」：几乎所有的凸函数都可以表示成一组仿射函数的逐点上确界

「凸优化问题定义」：凸优化问题是形如：

$$\begin{aligned} &minimize\ \ f_0(x) \\ &subject\ to\ \ f_i(x) \le 0 \\ &\qquad\qquad\ \ a_i^T x = b_i \end{aligned}$$

的问题，凸优化问题有三个附加的要求：

- 目标函数必须是凸的
- 不等式约束函数必须是凸的
- 等式约束函数必须是仿射的