运筹学与优化方法

晁国清
计算机科学与技术学院

课程回顾

➢共轭方向法

➢共轭方向：考虑正定矩阵Q及非零向量d^i, d^j, 如果

$$\left(d^i\right)^T Q d^j = 0$$

则称d^i, d^j关于矩阵Q共轭

Step 0：给定初始点x^0, 记$d^0 = -\nabla f(x^0), \epsilon > 0, k = 0$;

Step 1：判断$\|\nabla f(x^k)\| \le \epsilon$是否成立；是，则终止，否则转Step 2;

Step 2:计算步长$\alpha_k = -\frac{\nabla f(x^k)^T d^k}{(d^k)^T Q d^k}$;

Step 3:令$x^{k+1} = x^k + \alpha_k d^k$, 并计算方向

$$d^{k+1} = -\nabla f\left(x^{k+1}\right) + \beta_k d^k, \qquad \beta_k = \frac{\nabla f\left(x^{k+1}\right) Q d^k}{(d^k)^T Q d^k}$$

令$k = k + 1$,转Step 1.

➢共轭梯度法

Step 0:给定初始点x^0, 记$d^0 = -\nabla f(x^0), \epsilon > 0, k = 0$;

Step 1:判断$\|\nabla f(x^k)\| \le \epsilon$是否成立；是就终止，否则转Step 2;

Step 2:利用线性搜索计算步长α_k;

Step 3:令$x^{k+1} = x^k + \alpha_k d^k$, 并计算方向

$$d^{k+1} = -\nabla f\left(x^{k+1}\right) + \beta_k d^k$$

其中

$$\beta_k = \frac{\nabla f\left(x^{k+1}\right)^T \left(\nabla f\left(x^{k+1}\right) - \nabla f\left(x^k\right)\right)}{\nabla f(x^k)^T \nabla f(x^k)} (PRP)$$

或者

$$\beta_k = \frac{\nabla f\left(x^{k+1}\right)^T \nabla f\left(x^{k+1}\right)}{\nabla f(x^k)^T \nabla f(x^k)} \ (FR)$$

令$k = k + 1$, 转Step 1.

课程内容

- 线性规划的下界与对偶
- 最大流和最小割
- LP对偶的另一种理解方式
- 拉格朗日对偶函数
- 拉格朗日对偶问题
- 软对偶与强对偶

线性规划的下界与对偶

- 下界：在凸优化问题中，我们想要找到最优值的下界，$B \leq \min_{x} f(x)$
- 线性规划的例子：

$$\min_{x,y} x + y$$
$$s.t.\, x + y \geq 2$$
$$x, y \geq 0$$

下界是什么？

非常简单，取B =2

再来一个例子：

$$\min_{x,y} x + 3y$$
$$s.t.\, x + y \geq 2$$
$$x, y \geq 0$$

下界是什么？

$$x + y \geq 2$$
$$+2y \geq 0$$
$$= x + 3y \geq 2$$

下界B =2

哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

线性规划的下界与对偶

➢泛化一下：

下界是什么？

主线性规划（primal LP）

$$\min_{x,y} px + qy$$
$$s.t.\, x + y \geq 2$$
$$x, y \geq 0$$

$$a + b = p$$
$$a + c = q$$
$$a, b, c \geq 0$$

对于满足上述约束的a, b, c 下界$B = 2a$

找出下界，通过最大化下界，可以得到对偶线性规划(dual LP)

$$\max_{a,b,c} 2a$$
$$s.t.\, a + b = p$$
$$a + c = q$$
$$a, b, c \geq 0$$

注意点：
对偶变量的数目是原问题约束的数目

线性规划的下界与对偶

➢ 再举一个例子：

Primal LP:

$$\min_{x,y}\ px+qy$$
$$s.t.\ \ x\geq 0$$
$$y\leq 1$$
$$3x+y=2$$

Dual LP:

$$\max_{a,b,c} 2c-b$$
$$s.t.\ a+3c=p$$
$$-b+c=q$$
$$a,b\geq 0$$

一般形式线性规划的对偶：

给定$c\in R^n, A\in R^{m\times n}, b\in R^m, G\in R^{r\times n}, h\in R^r$

Primal LP:

$$\min_{x}\ c^Tx$$
$$s.t.\ Ax=b$$
$$Gx\leq h$$

Dual LP:

$$\max_{u,v}\ -b^Tu-h^Tv$$
$$s.t.\ -A^Tu-G^Tv=c$$
$$v\geq 0$$

解释：对于任意的u和$v\geq 0$，以及主规划问题的可行点x

$$u^T(Ax-b)+v^T(Gx-h)\leq 0$$
$$\Leftrightarrow (-A^Tu-G^Tv)^Tx\geq -b^Tu-h^Tv$$

当$c=-A^Tu-G^Tv$，可以得到原问题最优值的一个界

LP对偶的另一种理解方式

Primal LP:

$$\min_{x} c^T x$$
$$s.t.\ Ax = b$$
$$Gx \le h$$

Dual LP:

$$\max_{u,v} -b^T u - h^T v$$
$$s.t.\ -A^T u - G^T v = c$$
$$v \ge 0$$

- 另外一种理解方式：对于任意的$u, v \ge 0$以及主问题可行点x，有

$$c^T x \ge c^T x + u^T(Ax - b) + v^T(Gx - h) = L(x, u, v)$$

如果C表示主问题可行集，f^*表示主问题最优值，那么对于任意的$u, v \ge 0$，有

$$f^* \ge \min_{x \in C} L(x, u, v) \ge \min_{x} L(x, u, v) = g(u, v)$$

即$g(u, v)$对于任意$u, v \ge 0$是f^*的一个下界

注意：$g(u,v) = \begin{cases} -b^T u - h^T v, & if\ c = -A^T u - G^T v \\ -\infty & otherwise \end{cases}$

可以通过在任意$u, v \ge 0$上最大化$g(u, v)$得到最靠近的一个界，也就得到了对偶线性规划

拉格朗日对偶函数

- 考虑一般的最小化问题

$$\min_{x} f(x)$$
$$s.t.\ h_i(x) \le 0, i = 1, \cdots, m$$
$$\ell_j(x) = 0, j = 1, \cdots, r$$

注意：这里的函数可以不是凸函数

定义拉格朗日为

$$L(x,u,v) = f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j \ell_j(x)$$

这里$u \in R^m, v \in R^r, u \ge 0\ (else\ \min_{x} L(x,u,v) = -\infty)$

重要的性质：对于每一个可行点x, 有$f(x) \ge L(x,u,v)$

原因：

$$L(x,u,v) = f(x) + \underbrace{\sum_{i=1}^{m} u_i h_i(x)}_{\le 0} + \underbrace{\sum_{j=1}^{r} v_j \ell_j(x)}_{= 0} \le f(x)$$

拉格朗日对偶函数

- 令C作为主可行集，f^*表示主最优值，在所有的x上最小化$L(x,u,v)$给出一个下界

$$f^* \geq \min_{x\in C} L(x,u,v) \geq \min_{x} L(x,u,v) = g(u,v)$$

我们称$g(u,v)$为拉格朗日对偶函数，它给出了对于任意$u \geq 0, v$在f^*的一个下界，相应的u,v称为对偶可行点

二次规划的例子

- 考虑二次规划

$$\min_{x} \frac{1}{2}x^T Q x + c^T x$$
$$s.t. Ax = b, x \geq 0$$

$Q \succ 0$

拉格朗日$L(x,u,v) = \frac{1}{2}x^T Q x + c^T x - u^T x + v^T(Ax - b)$

拉格朗日函数：

$g(u,v)$

$= \min_{x} L(x,u,v) = -\frac{1}{2}(c - u + A^T v)^T Q^{-1}(c - u$

$+ A^T v) - b^T v$

这里取$\nabla_x L = 0$, 解出$x^* = -Q^{-1}(c - u + A^T v)$然后带入拉格朗日即可得到上式。

对于任意的$u \geq 0$和v，这给出了主问题最优值的下界

二次规划的例子

➢相同的问题

$$\min_{x} \frac{1}{2} x^T Q x + c^T x$$
$$s.t. Ax = b, x \geq 0$$

现在$Q \succcurlyeq 0$，

拉格朗日$L(x,u,v) = \frac{1}{2} x^T Q x + c^T x - u^T x + v^T (Ax - b)$

拉格朗日对偶函数：

$$g(u,v) = \begin{cases} -\frac{1}{2}(c - u + A^T v)^T Q^+ (c - u + A^T v) - b^T v \\ \qquad\qquad\qquad if\ c - u + A^T v \perp null(Q) \\ -\infty, \qquad otherwise \end{cases}$$

这里Q^+表示Q的泛化逆。对于任意的$u \geq 0, v$以及$c - u + A^T v \perp null(Q)$. $g(u,v)$是f^*上的一个非平凡下界

拉格朗日对偶问题

- 给定主问题：

$$\begin{aligned} &\min_{x} f(x) \\ s.t.\quad &h_i(x) \le 0, i = 1, \cdots, m \\ &\ell_j(x) = 0, j = 1, \cdots, r \end{aligned}$$

对偶函数$g(u,v)$对于任意的$u \ge 0$和v，满足$f^* \ge g(u,v)$

通过在对偶可行变量u,v上最大化$g(u,v)$可产生拉格朗日对偶问题：

$$\begin{aligned} &\max_{u,v} g(u,v) \\ &s.t. u \ge 0 \end{aligned}$$

重要性质一弱对偶：如果对偶问题最优值记为g^*，则$f^* \ge g^*$

即便对于原问题为非凸的情况该性质依然具备

重要性质二：对偶问题是凸问题(凹最大化问题)

$$\begin{aligned} g(u,v) &= \min_{x}\left\{ f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j \ell_j(x) \right\} \\ &= -\max_{x} \underbrace{\left\{ -f(x) - \sum_{i=1}^{m} u_i h_i(x) - \sum_{j=1}^{r} v_j \ell_j(x) \right\}}_{\text{对于}(u,v)\text{是凸函数的逐点最大}} \end{aligned}$$

哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

强对偶

- 之前讲了弱对偶$f^* \geq g^*$当$f^* = g^*$成立的时候就称之为强对偶
- Slater 条件：如果主问题是一个凸问题并且存在至少一个严格可行点$x \in R^n$，即

$h_1(x) < 0, \cdots, h_m(x) < 0$ 并且 $\ell_1(x) = 0, \cdots, \ell_r(x) = 0$

则强对偶成立

实际上，对于非仿射的h_i只需要严格不等式成立即可

对偶间隙(Dual Gap)

- 给定主问题可行点x和对偶问题可行点u, v，项$f(x) - g(u, v)$称为x和u, v之间的对偶间隙。

- 已知$f(x) - f^* \leq f(x) - g(u, v)$, 所以当对偶间隙为0时，$x$就是主最优的，相应的， u, v是对偶最优的

- 从算法角度来看，这也提供了一个终止判别标准：如果$f(x) - g(u, v) \leq \epsilon$, 那么我们可以保证$f(x) - f^* \leq \epsilon$

参考文献

- 运筹学与最优化方法，第6章，Page 142-145
- 最优化方法，第1章
- Convex Optimization, CMU, theory II Duality and optimility