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

- 线性规划的下界与对偶

Primal LP:

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

Dual LP:

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

- 再看LP对偶
  - 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

课程回顾

- 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

- 拉格朗日对偶问题

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

$$\max_{u,v} g(u,v)$$
$$s.t. u \geq 0$$

- 软对偶与强对偶

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

哈尔滨工业大学 HARBIN INSTITUTE OF TECHNOLOGY

课程内容

- KKT(Karush-Kuhn-Tucker)条件
- 举例
- 约束形式与拉格朗日形式

KKT条件

➢给定一个一般约束优化问题：

$$\min_{x} \quad f(x)$$

$$s.t.\ \ h_i(x) \le 0, i = 1, \cdots, m$$

$$\ell_j(x) = 0, j = 1, \cdots, r$$

KKT(Karush-Kuhn-Tucker)条件是：

$0 \in \nabla_x\left(f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j \ell_j(x)\right)$ (stationarity)

$u_i \cdot h_i(x) = 0\ for\ all\ i$ (complementary slackness)

$h_i(x) \le 0, \ell_j(x) = 0\ for\ all\ i, j$ (primal feasibility)

$u_i \ge 0\ for\ all\ i$ (dual feasibility)

KKT条件的必要性

- x^*和u^*,v^*分别是0对偶间隙主问题和对偶问题的解（强对偶成立，比如满足Slater’ s condition），则

$$f(x^*) = g(u^*,v^*)$$

$$= \min_x f(x) + \sum_{i=1}^{m} u_i^* h_i(x) + \sum_{j=1}^{r} v_j^* \ell_j(x)$$

$$\le f(x^*) + \sum_{i=1}^{m} u_i^* h_i(x^*) + \sum_{j=1}^{r} v_j^* \ell_j(x^*)$$

$$\le f(x^*)$$

也就是说，所有这些不等式事实上都是等式

必要性

- 从上面的内容总结两点：

1. 点x^*最小化$L(x,u^*,v^*)$。则$L(x,u^*,v^*)$的子微分在$x=x^*$处一定包含0，这就是stationary condition。
2. 我们必须有$\sum_{i=1}^{m} u_i^* h_i(x^*) = 0$, 因为每一项$u_i^* h_i(x^*) \leq 0\ \forall i$, 就意味着$u_i^* h_i(x^*) = 0\ \forall i$. 这就是complementary slackness

必要性：如果x^*和u^*, v^*都分别是原问题和对偶问题的解，并且原最优值和对偶最优值0间隙，那么x^*和u^*, v^*满足KKT条件。

注意：上述陈述并没有假设预先就有的关于问题的凸性，即对于函数f, h_i, ℓ_j要求是凸函数

充分性

- 如果存在满足KKT条件的x^*, u^*, v^*, 那么有

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

第一个等式成立因为stationarity，通过$\nabla_x L = 0$可以解出x^*，然后代入拉格朗日。第二个等式成立因为complementary slackness $u_i^* h_i(x^*) = 0$ 并且 $v_j^* \ell_j(x^*) = 0$

因此，对偶间隙为零(并且x^*和u^*, v^*分别为原问题和对偶问题的可行解) 那么x^*和u^*, v^*分别是原问题和对偶问题的最优解。

充分性：如果x^*和u^*, v^*满足KKT条件，那么x^*和u^*, v^*都分别是原问题和对偶问题的最优解

充分性必要性

- 总结：KKT条件等价于0对偶间隙：
  - 通常是充分的
  - 在强对偶条件下是必要的
- 充分性必要性：

对于一个满足强对偶性的问题(比如，假定满足Slater’s condition：凸问题并且存在一个严格满足非仿射不等式约束的x)，则

x^*和u^*,v^*分别是原问题和对偶问题的解

$\Leftrightarrow$ x^*和u^*,v^*满足KKT条件

KKT 条件最初叫KT(Kuhn-Tucker)条件,在1951年由两位作者首次发表,但是后来人们发现这些条件出现在Karush在1939年未发表的硕士学位论文里

对于一般的优化问题，KKT条件可以通过次梯度研究最优性条件推导出来：

$$0 \in \nabla f(x^*) + \sum_{i=1}^{m} N\{h_i \le 0\}(x^*) + \sum_{i=1}^{r} N\{\ell_j = 0\}(x^*)$$

$N_C(x)$是C在x处的normal cone.

Normal cone: given any set C and point $x \in C$, the normal cone is

$$\mathcal{N}_C(x) = \{g : g^T x \ge g^T y, \text{ for all } y \in C\}$$

This is always a convex cone, regardless of C.

举例：带等式约束的二次规划

➢考虑$Q \succ 0$的二次规划问题：

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

这是一个凸问题，没有不等式约束，根据KKT条件

x是一个解$\Leftrightarrow \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} -c \\ 0 \end{bmatrix}$ for some u.

上面的等式系统整合了stationary和primal feasibility条件

（complementary slackness 和 dual feasibility在该问题中都是没有意义的）

举例：Water-filling (B & V page 245)

➢考虑如下问题：

$$\min_{x} -\sum_{i=1}^{n} \log(\alpha_i + x_i)$$
$$s.t.\ \ x \geq 0, 1^T x = 1$$

信息理论：把$\log(\alpha_i + x_i)$考虑为第i个信道的通信率。

上述问题的KKT条件是：

$$-\frac{1}{\alpha_i + x_i} - u_i + v = 0, i = 1, \cdots, n$$
$$u_i \cdot x_i = 0, i = 1, \cdots, n,$$
$$x \geq 0, 1^T x = 1$$
$$u \geq 0$$

消去u，得到

$$\frac{1}{\alpha_i + x_i} \leq v, i = 1, \cdots, n \qquad (*)$$
$$x_i \left(v - \frac{1}{\alpha_i + x_i} \right) = 0, i = 1, \cdots, n \qquad (**)$$
$$x \geq 0, 1^T x = 1$$

举例：Water-filling (

$$
\begin{aligned}
&-\frac{1}{\alpha_i + x_i} - u_i + v = 0, i = 1, \cdots, n \\
&u_i \cdot x_i = 0, i = 1, \cdots, n, \\
&x \geq 0, 1^T x = 1 \\
&u \geq 0
\end{aligned}
$$

)

1. 如果$v^* < \frac{1}{\alpha_i}$, 条件(*)只有在$x_i^* > 0$时成立，再由条件(**)得到$v^* = \frac{1}{\alpha_i + x_i^*}$. 求解$x_i^*$，就得到$x_i^* = \frac{1}{v^*} - \alpha_i$ 条件是$v^* < \frac{1}{\alpha_i}$
2. 如果$v^* \geq \frac{1}{\alpha_i}$，那么$x_i^* > 0$是不可能的，因为这意味着$v^* \geq \frac{1}{\alpha_i} > \frac{1}{\alpha_i + x_i^*}$，这和complementarity slackness 条件相悖。所以，当$v^* \geq \frac{1}{\alpha_i}$时，有$x_i^* = 0$

综上得到

$$
x_i^* = \begin{cases} \frac{1}{v^*} - \alpha_i, & v^* < \frac{1}{\alpha_i} \\ 0, & v^* \geq \frac{1}{\alpha_i} \end{cases} = \max\left\{0, \frac{1}{v} - \alpha_i\right\}, i = 1, \cdots, n
$$

将上式代入$1^T x = 1$, 就得到

$$
\sum_{i=1}^{n} \max\left\{0, \frac{1}{v} - \alpha_i\right\} = 1
$$

左边是$\frac{1}{v}$的分段线性函数，断点在α_i, 所以该等式有一容易确定的唯一解

举例：Water-filling (B & V page 245)

- 跟water-filling的关系

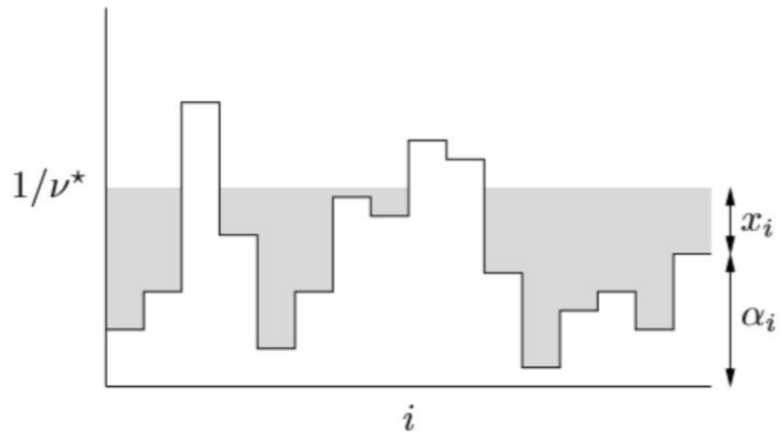


Figure 5.7 Illustration of water-filling algorithm. The height of each patch is given by α_i. The region is flooded to a level $1/\nu^\star$ which uses a total quantity of water equal to one. The height of the water (shown shaded) above each patch is the optimal value of $x_i^\star$.

将α_i视为patch i上方的高度，然后用水填充到$\frac{1}{v}$高度，如上图示，由于总用水量为$\sum_{i=1}^n \max\{0，\frac{1}{v^*} - \alpha_i\}$，通过增加水位，直到总用水量为1，那么patch i以上的水深就是最优值，即x_i^*

举例：支持向量机SVM

➤给定$y \in \{-1,1\}^n$ 和$X \in R^{n\times p}$, 支持向量机问题可表示为：

$$\min_{\beta,\beta_0,\xi} \quad \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{n}\xi_i$$

$$s.t. \quad \xi_i \geq 0, i = 1,\cdots,n$$

$$y_i\left(x_i^T\beta + \beta_0\right) \geq 1 - \xi_i, i = 1,\cdots,n$$

引入对偶变量$v, w \geq 0$. KKT stationary condition:

$$0 = \sum_{i=1}^{n} w_i y_i, \qquad \beta = \sum_{i=1}^{n} w_i y_i x_i, \qquad w = C1 - v$$

Complementary slackness:

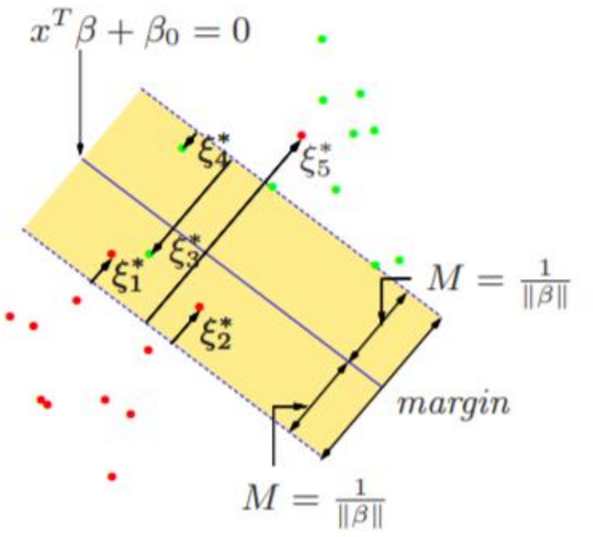
$$v_i\xi_i = 0, \qquad w_i\left(1 - \xi_i - y_i(x_i^T\beta + \beta_0)\right) = 0, i = 1,\cdots,n$$

因此，在最优解处有$\beta = \sum_{i=1}^{n} w_i y_i x_i$, 并且$w_i$是非零的仅当$y_i\left(x_i^T\beta + \beta_0\right) = 1 - \xi_i$. 满足这些条件的点称为支撑点（support points）

哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

举例：支持向量机SVM

- 对于支撑点i, 如果$\xi_i = 0$, 那么x_i就落在间隔的边上，并且$w_i \in (0, C]$;
- 对于支撑点i, 如果$\xi_i \neq 0$, 那么x_i就落在间隔的错的一面，并且w_i =C



KKT条件事实上并没有给寻找最优解的途径，但是给了一个比较好的理解

事实上，我们在执行最优化之前，可以用KKT条件去掉非支撑点

约束形式和拉格朗日形式

- 通常，在统计和机器学习领域，我们会在约束形式和拉格朗日形式之间进行切换。
- 约束形式：

$$\min_{x} \ f(x) \quad s.t. \quad h(x) \leq t \qquad \text{(C)}$$

这里$t \in R$就是一个要调的参数

- 拉格朗日形式：

$$\min_{x} f(x) + \lambda \cdot h(x) \qquad \text{(L)}$$

这里$\lambda \geq 0$是一个可调的参数

两种形式(C)和(L)是等价的，要求$f(x)$和$h(x)$是函数。

(C)到(L)：如果C严格可行，那么强对偶成立，如果再存在一个$\lambda \geq 0$，那么(C)中的任意解可最小化

$$f(x^*) + \lambda \cdot (h(x^*) - t)$$

因此，x^*也是形式(L)的一个解

约束形式和拉格朗日形式

- (L) 到 (C)：如果x^*是形式 (L) 中的一个解，令$t = h(x^*)$，形式 (C) 的KKT条件就得到满足，从而x^*就是形式 (C) 的一个解。
- 结论：

$$\bigcup_{\lambda \geq 0} \{(L)\text{的解}\} \subseteq \bigcup_{t \geq 0} \{(C)\text{的解}\}$$

$$\bigcup_{\lambda \geq 0} \{(L)\text{的解}\} \supseteq \bigcup_{\text{令}(C)\text{严格可行的}t} \{(C)\text{的解}\}$$

注意：可以产生一个可行但不严格可行约束集t的唯一值是$t = 0$, 我们也可以得到完美的等价

比如：如果$h \geq 0$并且问题 (C) 与 (L) 对于$t \geq 0, \lambda \geq 0$分别可行，那么我们就可以得到完美的等价

哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

关于对偶

- 对偶的重要应用：在强对偶条件下，可以通过求解对偶问题解达到求解原问题解的目的

在强对偶条件下，KKT条件对于最优性是必要的。给定对偶解u^*, v^*，任何原问题解x^*满足stationarity condition

$$0 \in \nabla f(x^*) + \sum_{i=1}^{m} u_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} v_i^* \nabla \ell_j(x^*)$$

即：x^*是$\min_x L(x, u^*, v^*)$

特别是，如果该解x^*唯一满足即问题$\min_x L(x, u^*, v^*)$有唯一最小解，那么相应的解就是原问题最优解。这在对偶问题比原问题容易求解的时候非常有用

参考文献

- 运筹学与最优性条件，第6章，Page133-144
- Convex Optimization, CMU, theory II Duality and optimality
- S. Boyd and L. Vandenberghe (2004), “Convex optimization”, Chapter 5
- R. T. Rockafellar (1970), “Convex analysis”, Chapters 28 - 30