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

- 只有等式约束的KT (Kuhn-Tucker)条件
- 只有不等式约束的KT条件
- 举例
- 一般约束的KT条件
- 凸规划的一阶充分条件

课程内容

- 对偶范数
- 对偶函数
- 对偶锥
- 对偶技巧

对偶范数

- $\|x\|$是一个范数，例如

ℓ_p范数：$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$ for $p \geq 1$

迹范数(Trace norm)：$\|X\|_{tr} = \sum_{i=1}^{r} \sigma_i(X)$，$\sigma(X)$表示矩阵$X$的奇异值

对偶范数$\|x\|_*$定义为

$$\|x\|_* = \max_{\|z\| \leq 1} z^T x$$

含义：对于一个范数小于等于1的向量z, z与x的内积最大值就是x的对偶范数

得到不定式$|z^T x| \leq \|z\| \|x\|_*$（泛化的Holder不等式），对于上面的两个例子，有

ℓ_p对偶范数：$\left(\|x\|_p\right)_* = \|x\|_q$，$\frac{1}{p} + \frac{1}{q} = 1$

迹范数对偶范数：$(\|X\|_{tr})_* = \|X\|_{op} = \sigma_1(X)$

对偶范数

- 对偶范数的对偶是其本身：$\|x\|_{**} = \|x\|$

证明：问题 $\min\limits_{y}\|y\| \quad s.t.\ y = x$的最优值是$\|x\|$

拉格朗日$L(y,u) = \|y\| + u^T(x-y) = \|y\| - y^T u + x^T u$

利用对偶范数的定义，

$$\min_{y}\{\|y\| - y^T u\} = \begin{cases} 0, & \|u\|_* \le 1 \\ -\infty, & \|u\|_* > 1 \end{cases}$$

因此，拉格朗日对偶问题变为：

$$\max_{\mathrm{u}} u^T x \quad s.t. \|u\|_* \le 1$$

这定义了对偶范数的范数，利用强对偶$f^* = g^*$, 即$\|x\|_{**} = \|x\|$

如果$\|u\|_* > 1$，那么存在$\|y\| \le 1$使得$y^T u = \|u\|_* > 1$, 从而$\|y\| - \mathrm{y}^T u \to -\infty$, 当$y^T u \to \infty$

如果$\|u\|_* \le 1$，那么存在$\|y\| \le 1$使得$y^T u = \|u\|_* \le 1$，从而$\|y\| - \mathrm{y}^T u \ge \|y\| - \|y\|\|u\|_* \ge 0$,从而当$y = 0$时，$\|y\| - \mathrm{y}^T u = 0$

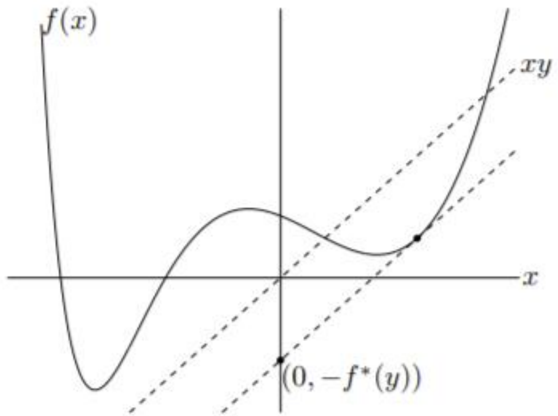
共轭函数

- 给定一个函数$f: R^n \to R$, 定义它的共轭函数$f^*: R^n \to R$,

$$f^*(y) = \max_x y^T x - f(x)$$

注意f^*通常是凸函数，为什么？

因为它是对于y的凸函数的逐点最大化



$f^*(y)$: 线性函数$y^T x$和f(x)之间的最大间隙

对于可微的f，共轭称作勒让德变换(Legendre transform)

共轭函数

- 性质：
  - 芬切尔不等式(Fenchel’s inequality)：对于任意的x,y,
    $f(x)+f^*(y)\geq x^Ty$
  - 共轭的共轭f^{**}满足$f^{**}\leq f$
  - 如果f是封闭（所有下水平集都是封闭）且凸的，那么，$f^{**}=f$
  - 如果f是封闭且凸的，那么对于任意的x,y,有

$$x\in\partial f^*(y)\Leftrightarrow y\in\partial f(x)$$
$$\Leftrightarrow f(x)+f^*(y)=x^Ty$$

  - 如果$f(u,v)=f_1(u)+f_2(v)$, 那么$f^*(w,z)=f_1^*(w)+f_2^*(z)$

独立函数的和的共轭等于独立函数的共轭的和

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

共轭函数

- 举例：
- 简单的二次规划：$f(x)=\frac{1}{2}x^TQx$，$Q\succ 0$，

$y^Tx-\frac{1}{2}x^TQx$对于y是严格凹的，并且在$x^*=Q^{-1}y$处取得最大值，从而

$$f^*(y)=\frac{1}{2}y^TQ^{-1}y$$

- 指示函数：如果$f(x)=I_C(x)$，那么它的共轭是

$f^*(y)=\max y^Tx-I_C(x)=\max_{x\in C} y^Tx= I_C^*(y)$

被称作C的支撑函数

- 范数：如果$f(x)=\|x\|$，那么它的共轭是

$$f^*(y)=I_{\{z:\|z\|_*\le 1\}}(y)$$

这里$\|\ \|_*$是$\|\ \|$的对偶范数

HARBIN INSTITUTE OF TECHNOLOGY

举例：Lasso dual

- 给定$y \in \mathbb{R}^n, X \in \mathbb{R}^{n\times p}$, Lasso(Least absolute shrinkage and selection operator)问题即：

$$\min_{\beta} \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

它的对偶函数是一个常数f^*。我们把原问题进行变换，得

$$\min_{\beta,z} \quad \frac{1}{2}\|y - z\|_2^2 + \lambda\|\beta\|_1$$
$$s.t. \quad z = X\beta$$

哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

举例：Lasso dual

对偶函数是

$$
\begin{aligned}
g(u) &= \min_{\beta,z} \frac{1}{2}\|y - z\|_2^2 + \lambda\|\beta\|_1 + u^T(z - X\beta) \\
&= \min_{z}\left(\frac{1}{2}\|y - z\|_2^2 + u^T z\right) \\
&\quad + \min_{\beta}(\lambda\|\beta\|_1 - (X^T u)^T\beta) \\
&= \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|y - u\|_2^2 \\
&\quad + \min_{\beta} \lambda\left(\|\beta\|_1 - \frac{(X^T u)^T}{\lambda}\beta\right) \\
&= \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|y - u\|_2^2 - \lambda I_{\{v:\|v\|_\infty \le 1\}}\left(\frac{X^T u}{\lambda}\right)
\end{aligned}
$$

举例：Lasso dual

因此, Lasso对偶问题是$\max\limits_{u} \frac{1}{2}(\|y\|_2^2 - \|y - u\|_2^2) \; s.t. \|X^T u\|_\infty \le \lambda \Leftrightarrow \min\limits_{u} \|y - u\|_2^2 \quad s.t. \|X^T u\|_\infty \le \lambda$

对偶问题是通过作变换$z = X\beta$推导出来的，故两者等价。
Slater 条件成立，则强对偶也成立。

注意：最优一个问题的最优值不一定是原问题的最优值

对于z的KKT稳定性条件给出了对偶解，任何对偶解都满足
$X\beta = y - u$

举例：Lasso dual

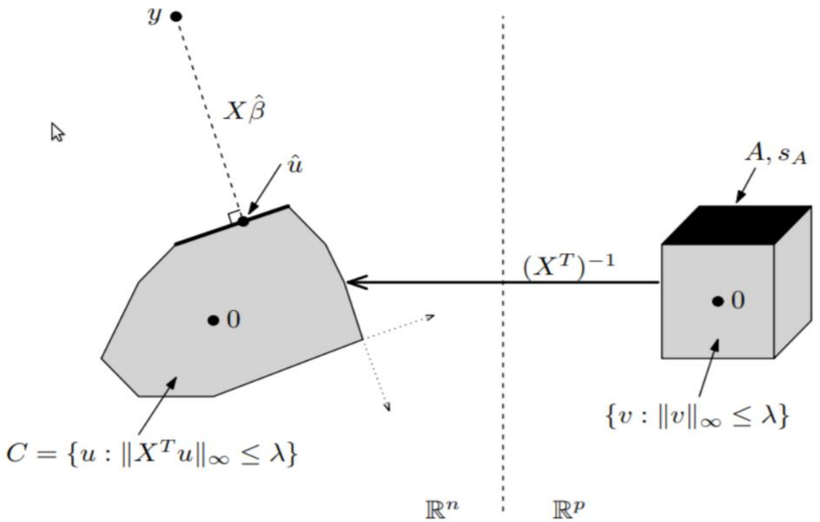


Figure 13.2: Visualization of Lasso Dual Problem, where its solution looks like projection of y onto set C, a polyhedron mapped from a hypercube.

对偶问题的解看起来像把y投影到集合$C = \{u: \|X^T u\|_\infty \le \lambda\} = (X^T)^{-1}\{v: \|v\|_\infty \le \lambda\}$. 这个概念呈现如上图。集合$C$是一个从立方体映射过来的多面体。这是线性映射下的逆像(inverse image)。 多面体的每一面对应Lasso选择的active set。 当把y投影到一个面的时候，所有的y_i共享Lasso解的符号的active set。 这些符号的active sets 是局部常数，我们可把lasso看做像变量选择器一样稳定。

共轭和对偶问题的关系

➢共轭常常出现在对偶问题的推导中，如通过

$$-f^*(u) = \min_x f(x) - u^T x$$

最小化拉格朗日

例如

$$\min_x f(x) + g(x)$$

等价于

$$\begin{aligned} \min_{x,z} \quad & f(x) + g(z) \\ s.t. \quad & x = z \end{aligned}$$

对偶函数是：

$$\begin{aligned} g(u) &= \min_x f(x) + g(z) + u^T(z - x) \\ &= -f^*(u) - g^*(-u) \end{aligned}$$

相应的，对偶问题：

$$\max_u -f^*(u) - g^*(-u)$$

共轭和对偶问题的关系

➢举例：

➢指示函数(Indicator function)：

Primal: $\min_x f(x) + I_C(x)$

Dual: $\max_u -f^*(u) - I_C^*(-u)$

这里的I_C^*是C的支撑函数

➢范数(Norms)：

Primal: $\min_x \quad f(x) + \|x\|$

Dual: $\max_u \quad -f^*(u)$

$s.t. \quad \|-u\|_* \le 1$

线性变化的切换(Shifting linear transformations)

➢对偶公式可以帮助我们把一个目标函数的一部分变为对偶函数的一部分。问题

$$\min_{\mathrm{x}} f(x)+g(Ax)$$

等价于

$$\begin{aligned}\min_{x,z}\quad & f(x)+g(z)\\ s.t.\quad & Ax=z\end{aligned}$$

与之前一样，其对偶为($L=f(x)+g(z)+u^T(z-Ax)$)

$$\max_{\mathrm{u}} -f^*(A^Tu)-g^*(-u)$$

举例：对于一个范数和它的对偶范数, $\|\cdot\|$与$\|\cdot\|_*$：

Primal: $\min_{x} \quad f(x)+\|Ax\|$

Dual: $\max_{u} \quad -f^*(A^Tu)$

$s.t. \quad \|-u\|_* \leq 1$

$$\min_{x,z} f(x)+\|z\|+u^T(z-Ax)$$

$$=\min_{x} f(x)-(A^Tu)^Tx+\min_{z}\|z\|+u^Tz$$

$$=-f^*(A^Tu)-I_{\{v:\|v\|_*\leq 1\}}(-u)$$

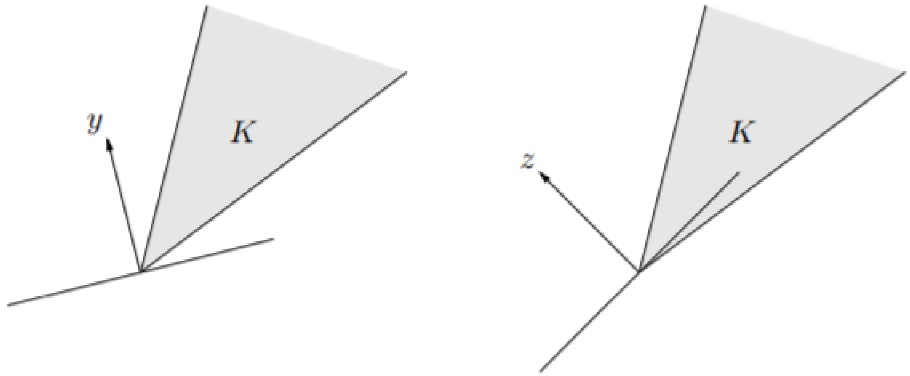
哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

对偶锥

- 对于一个锥$K \subseteq \mathbb{R}^n$ $(x \in K, t \geq 0 \Rightarrow tx \in K)$, 其对偶锥定义为：

$$K^* = \{y: y^T x \geq 0 \ \ for\ all\ x \in K\}$$

这通常是一个凸锥(即便K不是凸的)



注意：$y \in K^* \Leftrightarrow$ 半空间$\{x: y^T x \geq 0\}$包含K

重要性质：如果K是一个封闭的凸锥，那么$K^{**} = K$

对偶锥

- 举例:
- 线性空间(Linear subspace): 一个线性空间V的对偶锥是$V^{\perp}$，它的正交补，如$\left(row(A)^* = null(A)\right)$
- 范数锥(Norm Cone): 范数锥

$$K = \{(x,t) \in \mathbb{R}^{n+1}: \|x\| \le t\}$$

的对偶锥是它的对偶范数的范数锥,

$$K^* = \{(y,s) \in \mathbb{R}^{n+1}: \|y\|_* \le s\}$$

正半定锥(Positive Semidefinite cone): 凸锥$\mathbb{S}_+^n$是自对偶的(self-dual),意味着$(\mathbb{S}_+^n)^* = \mathbb{S}_+^n$. 为什么?

$$Y \succcurlyeq 0 \Leftrightarrow tr(YX) \ge 0 \; for \; all \; X \succcurlyeq 0$$

通过检查X的特征值分解

$$row(A) = the\ set\ of\ A^T x;\ \ null(A) = the\ set\ of\ y\ such\ that\ Ay = 0;$$
$$< y, A^T x > = y^T(A^T x) = x^T(Ay) = x^T 0 = 0$$

对偶锥和对偶问题

➢考虑锥约束问题：

$$\min_{x} f(x)$$
$$s.t. Ax \in K$$

它的对偶问题是：

$$\max_{u} -f^*(A^T u) - I_K^*(-u)$$

这里$I_K^*(y) = \max_{z \in K} z^T y$, K的支撑函数。如果K是一个锥，那么这就是

$$\max_{u} -f^*(A^T u)$$
$$s.t. u \in K^*$$

这里K^*是K的对偶锥，因为$I_K^*(-u) = I_{K^*}(u)$

这是一个有用的观察，因为很多不同形式的约束可以视作锥约束

对偶技巧

- 通常，我们会把对偶转化为等价的问题并称之为对偶问题。在强对偶条件下，我们可以用对偶问题的解计算原问题的解。

警告：转换的对偶问题的最优值未必是原问题的最优值

- 推导无约束问题对偶的一个常见技巧是增加一个哑变量(dummy variable)和等式约束来转换原问题

不同的选择会产生不同的对偶问题

双对偶(Double dual)

➢考虑带线性约束的一般最小化问题：

$$\begin{aligned} \min_{x} \quad & f(x) \\ s.t. \quad & Ax \le b \\ & Cx = d \end{aligned}$$

其拉格朗日是

$$L(x,u,v) = f(x) + (A^T u + C^T v)^T x - b^T u - d^T v$$

因此，其对偶问题为：

$$\begin{aligned} \max_{u,v} \quad & -f^*(-A^T u - C^T v) - b^T u - d^T v \\ s.t. \quad & u \ge 0 \end{aligned}$$

回顾共轭的性质：当f是封闭且凸时，$f^{**} = f$.
在这里，对偶的对偶就是原问题

双对偶(Double dual)

- 事实上，对偶的对偶与共轭之间的关联不局限于线性约束。

考虑如下问题：

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

如果f和$h_1,\cdots,h_m$都是封闭和凸的，$\ell_1,\cdots,\ell_r$是仿射的，那么对偶的对偶就是原问题

这可以通过双函数(bifunction)的最小化问题进行证明。在此框架下，对偶函数对应于双函数的共轭(更多内容，可以参考Rockafellar 书的29和30章)。

参考文献

- Convex Optimization, CMU, theory II Duality and optimality
- S. Boyd and L. Vandenberghe (2004), “Convex optimization”, Chapter 2,3, 5
- R. T. Rockafellar (1970), “Convex analysis”, Chapters14,16, 28-30