凸优化理论与应用

第5章　对偶问题

http://shijuanfeng.blogbus.com/

Outline

- **Lagrange**对偶函数：定义、例子、和共轭函数
- **Lagrange**对偶问题：强对偶，弱对偶

- 最优性条件：互补松弛，**KKT**
- 扰动与灵敏度分析
- 择一定理：弱择一，强择一

- 广义不等式：没在**PPT**上列出

优化问题的拉格朗日函数

- 设优化问题：

$$
\begin{aligned}
\text{minimize} \quad & f_0(x), \quad x \in \mathcal{R}^n \\
\text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \dots, m \\
& h_i(x) = 0, \quad j = 1, \dots, p
\end{aligned}
$$

- 拉格朗日**(Lagrangian)**函数：

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

- 对固定的 x，拉格朗日函数 $L(x, \lambda, \nu)$ 为关于 λ 和 ν 的仿射函数。

拉格朗日对偶函数

- **拉格朗日对偶函数(lagrange dual function)：**

$$g(\lambda,\nu)=\inf_{x\in D} L(x,\lambda,\nu)=\inf_{x\in D}\left(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}\nu_i h_i(x)\right)$$

若拉格朗日函数没有下界，则令

$$g(\lambda,\nu)=-\infty$$

- 拉格朗日对偶函数为凹函数。
- 对$\forall\lambda\succ=0$和 $\forall\nu$，若原最优化问题有最优值p^*，则

$$g(\lambda,\nu)\le p^*$$

Least-squares solution of linear equations

- 原问题：

$$\begin{aligned} &\text{minimize} \quad x^T x, \quad x \in \mathcal{R}^n \\ &\text{subject to} \quad Ax = b \end{aligned}$$

- 拉格朗日函数：

$$L(x,\nu) = x^T x + \nu^T (Ax - b)$$

- 拉格朗日对偶函数：

$$g(\nu) = -\frac{1}{4}\nu^T AA^T \nu - b^T \nu$$

对偶函数与共轭函数

- 共轭函数 $f^*(y) = \sup_{x \in \mathrm{dom} f} (y^T x - f(x))$
- 共轭函数与对偶函数存在密切联系
- 具有线性不等式约束和线性等式约束的优化问题：

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & Ax \preceq b \\ & Cx = d \end{aligned}$$

对偶函数：

$$g(\lambda, \nu) = -b^T \lambda - d^T \nu - f_0^*(-A^T \lambda - C^T \nu)$$

上界

共轭函数

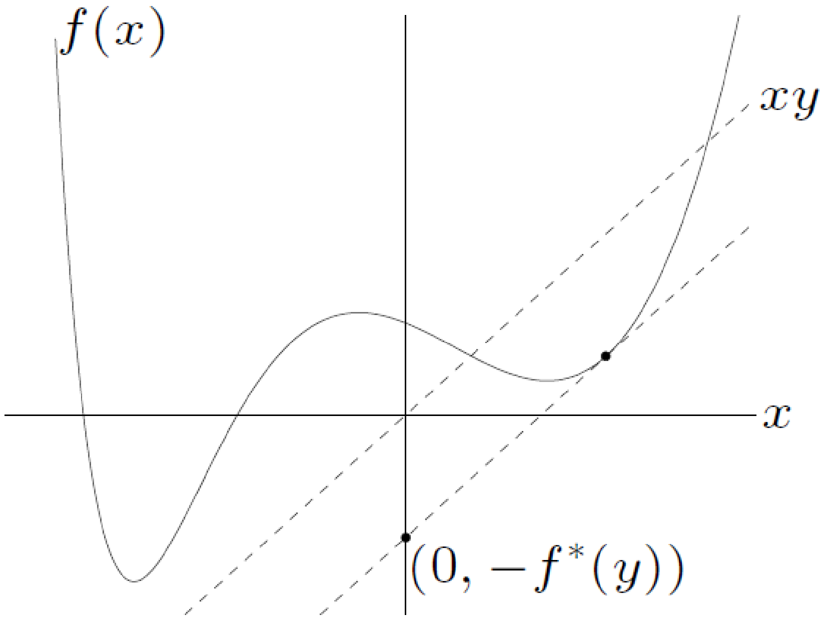
$$f^*(y) = \sup_{x \in \mathrm{dom} f} (y^T x - f(x)).$$

- 凸

将在第五章对偶问题中用到

- a±bi

- ?



- f*(y)是线性函数yx和f(x)之间最大的差值

Equality constrained norm minimization

- 问题描述：

$$\begin{array}{ll} \text{minimize} & \|x\| \\ \text{subject to} & Ax = b \end{array}$$

- 共轭函数：

$$f_0^*(y) = \begin{cases} 0 & \|y\|_* \le 1 \\ \infty & otherwise \end{cases}$$

- 对偶函数：

$$g(\nu) = -b^T\nu - f_0^*(-A^T\nu) = \begin{cases} -b^T\nu & \|-A^T\nu\|_* \le 1 \\ -\infty & otherwise \end{cases}$$

拉格朗日对偶问题

- **拉格朗日对偶问题的描述：**

$$\begin{aligned} &\text{maximize} \quad g(\lambda, \nu) \\ &\text{subject to} \quad \lambda \succ= 0 \end{aligned}$$

- **对偶可行域** $\lambda \succ= 0$

$$g(\lambda, \nu) > -\infty$$

- **最优值** d^*

- **最优解** (λ^*, ν^*)

LP问题的对偶问题

- **标准LP问题**

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \succ= 0 \end{aligned}$$

- **对偶函数**

$$g(\lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & otherwise \end{cases}$$

- **对偶问题**

$$\begin{aligned} \text{minimize} \quad & g(\lambda, \nu) \\ \text{subject to} \quad & \lambda \succ= 0 \end{aligned}$$

- **等价描述**

$$\begin{aligned} \text{minimize} \quad & g(\lambda, \nu) \\ \text{subject to} \quad & A^T \nu - \lambda + c = 0 \\ & \lambda \succ= 0 \end{aligned}$$

弱对偶性

- 定理（弱对偶性）：设原始问题的最优值为 p^*，对偶问题的最优值为d^*，则 $d^* \le p^*$ 成立。
- **optimal duality gap**

$$p^* - d^*$$

- 可以利用对偶问题找到原始问题最优解的下界。

强对偶性

- 定义（强对偶性）：设原始问题的最优值为 p^*，对偶问题的最优值为 d^*。若 $d^* = p^*$ 成立，则称原始问题和对偶问题之间具有**强对偶性**。
- 强对偶性并不是总是成立的。
- 凸优化问题**通常（但并不总是）**具有强对偶性。
- **Slater**定理：若凸优化问题存在严格可行解，即存在 $x \in \operatorname{relint} D$，满足

$$f_i(x) < 0, i = 1, \ldots, m,$$

$$Ax = b$$

则优化问题具有强对偶性。该条件称为**Slater条件**。

强对偶性

- 弱化的**Slater**条件：若不等式约束条件的前 k 个为线性不等式约束条件，则**Slater**条件可以弱化为：

  存在 $x \in \operatorname{rel\,int} D$，满足

$$
\begin{aligned}
& f_i(x) \le 0, i = 1, \dots, k, \\
& f_i(x) < 0, i = k+1, \dots, m, \\
& Ax = b
\end{aligned}
$$

Least-squares solution of linear equations

- 原问题：

$$\begin{array}{ll} \text{minimize} & x^T x, \quad x \in \mathcal{R}^n \\ \text{subject to} & Ax = b \end{array}$$

- 对偶问题：

$$\text{maximize} \quad g(v) = -\frac{1}{4} v^T A A^T v - b^T v$$

- 具有强对偶性

对偶可行解的不等式

- 对于一优化问题，若 x 为一可行解，(λ,ν) 为对偶问题可行解，则有如下不等式：

$$f_0(x)-p^* \le f_0(x)-g(\lambda,\nu)$$

  x 为$\varepsilon-$次优解，其中

$$\varepsilon = f_0(x)-g(\lambda,\nu)$$

- 不等式可以用于对次优解的精度估计

互补松弛条件

- 设 x^* 为原始优化问题的最优解, (λ^*, ν^*) 为对偶问题的最优解, 若两者具有强对偶性，则

$$
\begin{aligned}
f_0(x^*) = g(\lambda^*, \nu^*) \\
&= \inf_x \left(f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_i \nu_i^* h_i(x) \right) \\
&\leq f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_i \nu_i^* h_i(x^*) \\
&\leq f_0(x^*)
\end{aligned}
$$

所以

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

即

$$\lambda_i^* f_i(x^*) = 0, i = 1, \ldots, m$$

KKT优化条件

- 设优化问题中，函数 $f_0(x),\dots,f_m(x),h_0(x),\dots,h_p(x)$ 可微。设x^*为原始优化问题的最优解，(λ^*,ν^*) 为对偶问题的最优解，且两者具有强对偶性，则(x^*,λ^*,ν^*) 满足如下条件：

1. $f_i(x^*)\le 0, i=1,\dots,m$

2. $h_i(x^*)=0, i=1,\dots,p$

3. $\lambda_i^*\ge 0, i=1,\dots,m$

4. $\lambda_i^* f_i(x^*)=0, i=1,\dots,m$

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

KKT条件为必要条件！

凸优化问题的KKT条件

- 设原始问题为凸优化问题中，函数

$$f_0(x),\dots,f_m(x),h_0(x),\dots,h_p(x)$$

可微。设$(\tilde{x},\tilde{\lambda},\tilde{v})$ 满足**KKT**条件，则 $\tilde{x}$ 为原始问题的最优解，而 $(\tilde{\lambda},\tilde{v})$ 为对偶问题的最优解，且两者具有强对偶性。

例

- 原始凸优化问题 minimize $-\sum_i \log(\alpha_i + x_i)$

  subject to $x \succ= 0$

  $1^T x = 1$

  KKT条件 $x^* \succ= 0, 1^T x^* = 1, \lambda^* \succ= 0,$

  $\lambda_i^* x_i^* = 0, i = 1, \ldots, n,$

  $-1/(\alpha_i + x_i^*) - \lambda_i^* + \nu^* = 0, i = 1, \ldots, n$

凸优化问题的对偶求解

- 设原始优化问题与对偶问题具有强对偶性，且(λ^*, ν^*) 为对偶问题的最优解。$L(x, \lambda^*, \nu^*)$ 存在唯一的最小解，即

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x)$$

存在唯一解 x^* 。若 x^* 为原始问题的可行解，则x^* 即为原始问题的最优解；若 x^* 不是原始问题的可行解，则原始问题不存在最优解。

扰动问题

- 扰动问题：

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le u_i, \quad i=1,\dots,m \\ & h_j(x) = v_j, \quad j=1,\dots,p \end{aligned}$$

- 当 $u=0, v=0$ 时即为原始问题。
- 若 u_i 为正，则第 i 个不等式约束被放宽；若 u_i 为负，则第 i 个不等式约束被收紧。
- 记 $p^*(u,v)$ 为扰动问题的最优解。若扰动问题无最优解，则记

$$p^*(u,v)=\infty$$

灵敏度分析

- 设对偶问题存在最优解，且与原始问题具有强对偶性，若非干扰问题的最优对偶解为(λ^*, ν^*)，则有

$$p^*(u,v) \geq p^*(0,0) - \lambda^{*T} u - \nu^{*T} v$$

- 若 $p^*(u,v)$ 在$u=0, v=0$ 处可微，则

$$\lambda_i^* = -\frac{\partial p^*(0,0)}{\partial u_i} \qquad \nu_i^* = -\frac{\partial p^*(0,0)}{\partial v_i}$$

弱择一定理

- 定义：若两个不等式（等式）系统，至多有一个可解，则称这两个系统具有弱选择性。

- 设原始问题的约束条件：

$$f_i(x) \le 0, i = 1,\dots,m, h_j(x) = 0, j = 1,\dots,p$$

- 对偶问题

$$g(\lambda,\nu) = \inf(\sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x))$$

- 对偶不等式组

$$\lambda \succ= 0, g(\lambda,\nu) > 0$$

- 原始问题的约束条件与对偶不等式组具有弱选择性。

强择一定理

- 定义：若两个不等式（等式）系统，恰有一个可解，则称这两个系统具有强选择性。
- 设原始问题为凸优化问题，其严格不等式约束条件为：

$$f_i(x) < 0, i = 1, \ldots, m, Ax = b$$

- 对偶不等式组 $\lambda \succ= 0, \lambda \neq 0, g(\lambda, v) > 0$
- 若存在 $x \in \mathrm{renlint} D$ ，满足 $Ax = b$，则上述两不等式约束系统具有强选择性。

Summary

- **Lagrange**对偶函数：定义、例子、和共轭函数
- **Lagrange**对偶问题：强对偶，弱对偶

- 最优性条件：互补松弛，**KKT**
- 扰动与灵敏度分析
- 择一定理：弱择一，强择一

- 广义不等式：没在**PPT**上列出