对偶理论（Duality）

张伯雷

南京邮电大学 计算机学院、通达学院

https://bolei-zhang.github.io/course/opt.html

回顾

- 凸集
  - 集合中任意两点所组成的**线段**仍然在该集合中
- 凸函数
  - 定义一：如果dom$f$为凸集，且$f(\theta x+(1-\theta)y)\leq\theta f(x)+(1-\theta)f(y)$对所有的$x,y\in\mathrm{dom}f$，$0\leq\theta\leq1$成立
  - 定义二：对于可微函数f，如果dom$f$为凸集，则f为凸函数当且仅当$f(y)\geq f(x)+\nabla f(x)^T(y-x)$，对所有的$x,y\in\mathrm{dom}f$成立
  - 定义三：对于二阶可微函数f，如果dom$f$为凸集，则f为凸函数当且仅当$\nabla^2 f(x)\succcurlyeq 0$，对所有$x\in\mathrm{dom}f$成立
- 凸优化问题
  - 标准形式
  - 局部最优=全局最优
  - 最优条件：$x\in X$最优 $\Leftrightarrow \nabla f_0^T(x)(y-x)\geq 0$ for all $y\in X$

目录

- 拉格朗日与共轭函数
- 对偶间隙
- 最优性条件：Slater条件
- 最优条件的两个解释
  - 几何解释
  - 鞍点解释
- 互补松弛与KKT条件

一般优化问题

- 优化问题

$$\min f_0(x).$$
$$\text{s.t.}\quad f_i(x) \le 0, \qquad i = 1, \dots, m$$
$$h_i(x) = 0, \qquad i = 1, \dots, p$$

  - $x \in R^n$
  - $D = \bigcap_{i=0}^{m} dom f_i \cap \bigcap_{i=1}^{p} dom h_i$
  - $X = \{x \in D$ 且所有约束条件可以满足$\}$
  - p^* 为最优值

不一定是凸优化问题!

拉格朗日函数

- 优化问题

$$
\begin{aligned}
\min \quad & f_0(x). \\
\text{s.t.} \quad & f_i(x) \le 0, \qquad i = 1, \dots, m \\
& h_i(x) = 0, \qquad i = 1, \dots, p
\end{aligned}
$$

- 拉格朗日函数（Lagrangian function）
  - 给该问题中的每一个约束指定一个拉格朗日乘子λ, ν，以乘子为加权系数将约束增加到目标函数中
  - L: $\mathbb{R}^n \times \mathbb{R}^{\mathrm{m}} \times \mathbb{R}^p \rightarrow \mathbb{R}$，其中定义域为$D \times \mathbb{R}^{\mathrm{m}} \times \mathbb{R}^p$，

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

拉格朗日对偶函数

- 拉格朗日对偶函数(Lagrange Function/Dual Function)

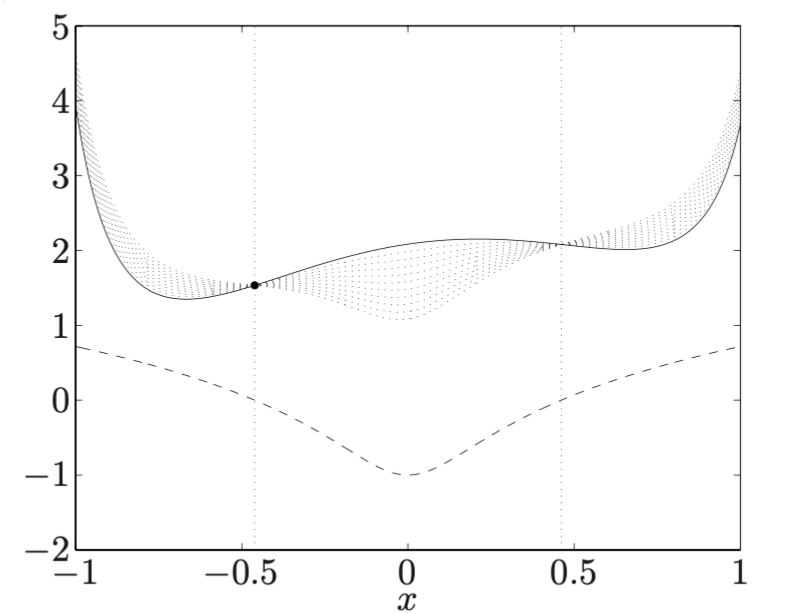
$$g(\lambda, \upsilon) = \inf_{x \in D} L(x, \lambda, \upsilon)$$

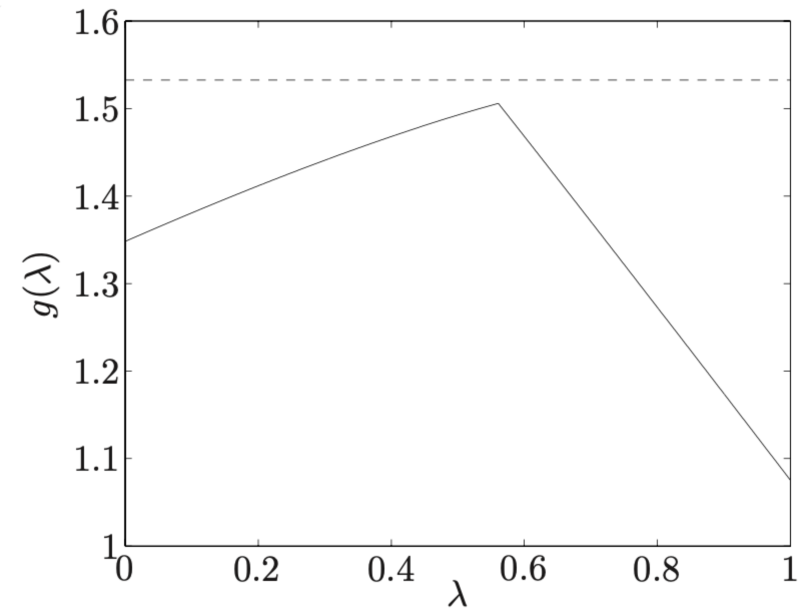
$$= \inf_{x \in D} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \upsilon_i h_i(x)\right)$$

- $\lambda \in \mathbb{R}^m, \upsilon \in \mathbb{R}^p$

对偶函数性质一

- 弱对偶理论：如果$\lambda \geq 0$，则$g(\lambda, \upsilon) \leq p^*$





实线：目标函数f_0

虚线：$f_1 \leq 0$

点虚线：拉格朗日$f_0 + \lambda f_1$:$\lambda = 0.1, 0.2, \ldots, 1.0$

可行解集：[-0.46, 0.46]

对偶函数性质二

- 性质：对偶函数为凹函数

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

- 命题：若$f_1,\ldots,f_m$为凸函数，则$f(x)=\max\{f_1(x),\ldots,f_m(x)\}$为凸函数
- 命题：若$f(x,y)$对所有$y\in A$为关于x的凸函数，则$f(x)=\sup_{y\in A} f(x,y)$ 为凸函数

例1

$$\min \quad x^T x$$

$$\text{s.t.} \quad \mathrm{Ax} = b$$

$$x \in \mathbb{R}^n, b \in \mathbb{R}^p, A \in \mathbb{R}^{p \times n}$$

例2

$$\min \quad C^T x$$

$$\text{s.t.} \quad Ax = b\,, x \geq 0$$

例3

$$
\begin{aligned}
\min \quad & x^T W x \\
\text{s.t.} \quad & x_i = \pm 1, i = 1, \ldots, m
\end{aligned}
$$

函数的共轭

- 函数共轭:

$$f^*\text{是}f\text{的共轭，若}f^*(y)=\sup_{x\in dom\, f}(y^Tx-f(x))$$

- 拉格朗日对偶函数(Lagrange Function/Dual Function)

$$g(\lambda,v)=\inf_{x\in D}L(x,\lambda,v)$$

$$=\inf_{x\in D}(f_0(x)+\sum_{i=1}^{m}\lambda_if_i(x)+\sum_{i=1}^{p}v_ih_i(x))$$

例

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x = 0 \end{aligned}$$

- 解

例

$$\min \quad f_0(x)$$
$$s.t. \quad Ax \le b, Cx = d$$

- 解

对偶理论（Duality）

张伯雷

南京邮电大学 计算机学院、通达学院

https://bolei-zhang.github.io/course/opt.html

目 录

- 拉格朗日函数

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

- 对偶函数

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

  - 弱对偶理论：如果$\lambda\geq 0$，则$g(\lambda,v)\leq p^*$
  - 性质：对偶函数为凹函数

对偶问题（dual problem）

- 原问题（Primal problem）

$$\begin{aligned} \min \quad & f_0(x). \\ \text{s.t.} \quad & f_i(x) \le 0, \qquad i = 1, \dots, m \\ & h_i(x) = 0, \qquad i = 1, \dots, p \end{aligned}$$

- 对偶问题（Dual problem）

$$\begin{aligned} \max \quad & g(\lambda, \upsilon) \\ \text{s.t.} \quad & \lambda \ge 0 \end{aligned}$$

  - 通过最大化$g(\lambda, \upsilon)$，可以不断逼近于p^*的下界

$$\max_{\lambda \ge 0, \upsilon} g(\lambda, \upsilon) = \max_{\lambda \ge 0, \upsilon} \inf_{x \in D} L(x, \lambda, \upsilon) \le p^*$$

  - 该问题为一个凸优化问题，最优值为d^*，最优解为(λ^*, υ^*)

对偶间隙（duality gap）

- 对偶问题为凸优化问题
- $p^* - d^*$：对偶间隙（duality gap）
- 弱对偶（weak duality）： $d^* \leq p^*$
  - 对凸优化问题与非凸优化问题都成立
  - 用于找到一个原问题的下界
- 强对偶（strong duality）： $d^* = p^*$

相对内点集（relative interior）

定义 5.13 (相对内点集) 给定集合 $\mathcal{D}$，记其仿射包为 $\textbf{affine}\,\mathcal{D}$（见定义2.14）. 集合 $\mathcal{D}$ 的**相对内点集**定义为

$$\textbf{relint}\,\mathcal{D} = \{x \in \mathcal{D} \mid \exists\, r > 0,\ 使得\ B(x,r) \cap \textbf{affine}\,\mathcal{D} \subseteq \mathcal{D}\}.$$

- $B(x,r)$：以x为中心的球

Slater条件

- 强对偶成立的充分条件
  - 对于一个凸优化问题

$$\min f_0(x).$$
$$\text{s.t.}\quad f_i(x) \le 0, \qquad i = 1, \dots, m$$
$$\mathrm{Ax} = \mathrm{b}$$

  - Slater条件：对以上的凸优化问题，存在$x \in \operatorname{relint} D$，满足

$$f_i(x) < 0,\ i = 1, \dots, m, \quad \mathrm{Ax} = \mathrm{b}$$

    - 则称为Slater条件，此时$d^* = p^*$
    - 对大多数凸优化问题，slater条件都成立；对非凸问题，强对偶也可能存在
  - 弱Slater条件：若不等式约束为仿射时，只要可行域非空，必有$d^* = p^*$

例1

- 线性约束的最小二乘问题

$$\begin{aligned} \min \quad & x^T x \\ \text{s.t.} \quad & \mathrm{A}\mathrm{x} = b \end{aligned}$$

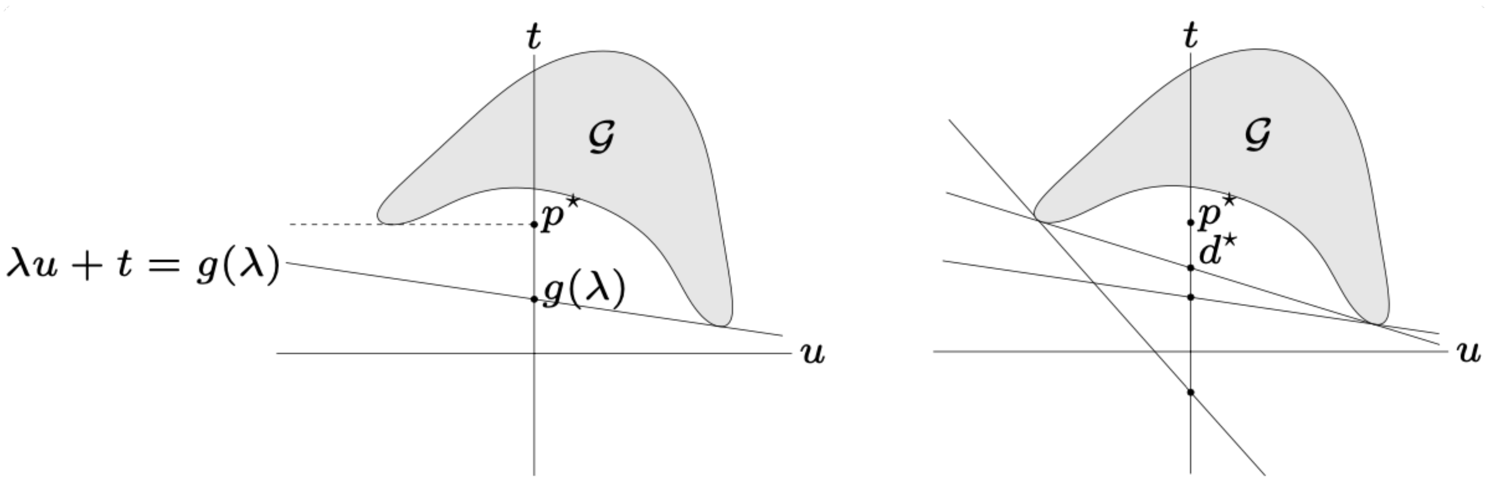
$d^* \leq p^*$

- 几何解释

$$\min f_0(x)\,.$$

$$\text{s.t.}\quad f_1(x) \leq 0, \qquad i = 1, \dots, m$$

  - $G = \{(f_1(x), f_0(x)) | x \in D\}$
  - $p^* = \inf\{t | (u, t) \in G, u \leq 0\}$
  - $g(\lambda) = \inf\{\lambda u + t | (u, t) \in G\}$



$d^* = p^*$

- 几何解释

$$\min f_0(x).$$
$$\text{s.t.}\quad f_1(x) \le 0, \qquad i = 1, \dots, m$$

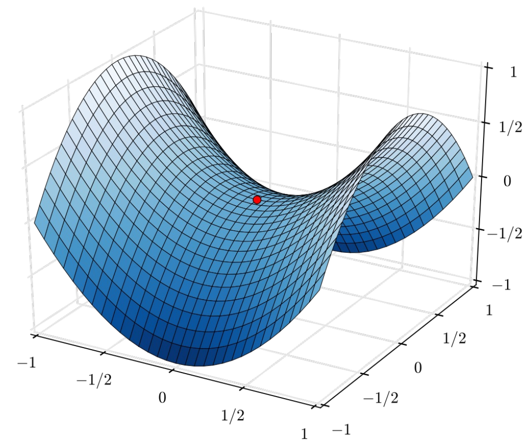
- $G = \{(f_1(x), f_0(x)) | x \in D\}$
- $p^* = \inf\{t | (u, t) \in G, u \le 0\}$
- $g(\lambda) = \inf\{\lambda u + t | (u, t) \in G\}$

$d^* = p^*$

- 鞍点解释
  - 极小极大不等式

$$\sup_{\lambda \geq 0} \inf_{x \in D} L(x, \lambda) \leq \inf_{x \in D} \sup_{\lambda \geq 0} L(x, \lambda)$$

$$d^* \leq p^*$$



  - 鞍点定理
    - 对一般的优化问题
    - 拉格朗日函数有鞍点⇔该点为原始/对偶问题的最优解，对偶间隙为0

根据对偶间隙求ϵ次优解

- 如果x是原始问题的可行解，λ,ν是对偶问题的可行解，则存在

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

- ε-次优解集（ε – suboptimal set）

$$X_\varepsilon = \{x | x \in \mathrm{X}_f, f_0(x) \le p^* + \varepsilon\}$$

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

- 原始问题可行解x与对偶问题可行解λ,υ对应的对偶间隙

$$p^* \in [g(\lambda,\nu), f_0(x)]$$

$$d^* \in [g(\lambda,\nu), f_0(x)]$$

算法

- 假设一个算法可以产生序列的可行解:
  - $x^{(k)}, (\lambda^{(k)}, \nu^{(k)})$, for $k = 1,2, \dots$
- 预先设定一个期望的间隙值ε_{abs}
- 不断生成可行解，直至

$$f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)}) \le \epsilon_{\rm abs}$$

- 则$x^{(k)}$是ε_{abs}-次优解

互补松弛（complementary slackness）

- 假设强对偶成立，x^* 为原问题的最优解，λ^*, ν^* 为对偶问题的最优解

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

- 结论1： x^* 最小化 $L(x, \lambda^*, \nu^*)$.
- 结论2： $\lambda_i^* f_i(x^*) = 0, i = 1, \ldots, m.$ （互补松弛 complementary slackness）

互补松弛（complementary slackness）

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

- 如果$\lambda_i^* > 0$，则$f_i(x^*) = 0$
- 如果$f_i(x^*) < 0$，则$\lambda_i^* = 0$

KKT条件（一般可微优化问题）

- 假设$f_0, f_1, \dots, f_m, h_1, \dots, h_p$都是可微的，但不一定为凸函数
- 假设强对偶成立，x^*为原问题的最优解，λ^*, ν^*为对偶问题的最优解
- 由于x^*最小化$L(x, \lambda^*, \nu^*)$，因此该点的梯度为0

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

- KKT条件：

$$\begin{array}{rcll} f_i(x^\star) & \leq & 0, & i = 1, \dots, m \\ h_i(x^\star) & = & 0, & i = 1, \dots, p \\ \lambda_i^\star & \geq & 0, & i = 1, \dots, m \\ \lambda_i^\star f_i(x^\star) & = & 0, & i = 1, \dots, m \\ \nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(x^\star) & = & 0, & \end{array}$$

KKT条件（可微凸优化问题）

- 假设$f_0, f_1, \dots, f_m, h_1, \dots, h_p$都是可微的，且该问题为**凸优化问题**
- KKT条件是强对偶的充分必要条件

- 证明：假设 $(\tilde{x}, \tilde{\lambda}, \tilde{v})$ 为满足KKT条件一个可行解

$$
\begin{aligned}
g(\tilde{\lambda}, \tilde{\nu}) &= L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \\
&= f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{x}) \\
&= f_0(\tilde{x}),
\end{aligned}
$$

- 可以得到：$\tilde{x}, (\tilde{\lambda}, \tilde{v})$ 满足强对偶，**同时为原始与对偶问题的最优解**

KKT条件的意义

- 对一般可微的优化问题（不一定为凸）
  - 对偶间隙为0⇒KKT条件成立
- **对可微的凸优化问题**
  - 对偶间隙为0⟺ KKT条件成立
- 意义:
  - 如果一个凸优化问题，存在很多约束条件，则往往难以求解
  - 很多凸优化问题可以通过求解KKT条件来解决