清華大学
Tsinghua University

运筹学

第五讲 凸优化

魏韡

2025年4月16日

主要内容

1. 凸函数
2. 凸优化
3. 拉格朗日对偶
4. KKT最优性条件

1.1 凸函数的定义

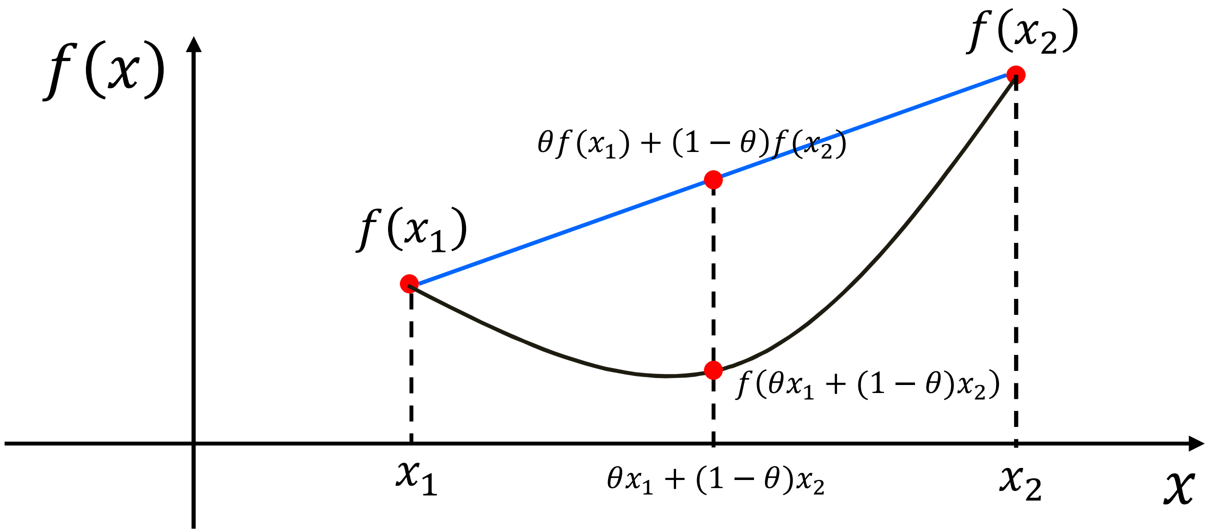
定义1

函数 $f:\mathbb{R}^n \to \mathbb{R}$ 是凸函数，若其定义域 $\mathrm{dom}[f]$ 是凸集，且

$$f(\theta x_1 + (1-\theta)x_2) \le \theta f(x_1) + (1-\theta) f(x_2)$$

$$\forall x_1, x_2 \in \mathrm{dom}[f], \theta \in [0,1]$$

函数 $f:\mathbb{R}^n \to \mathbb{R}$ 是凹函数若 $-f$ 是凸函数



1.1 凸函数的定义

凸性的进一步分类

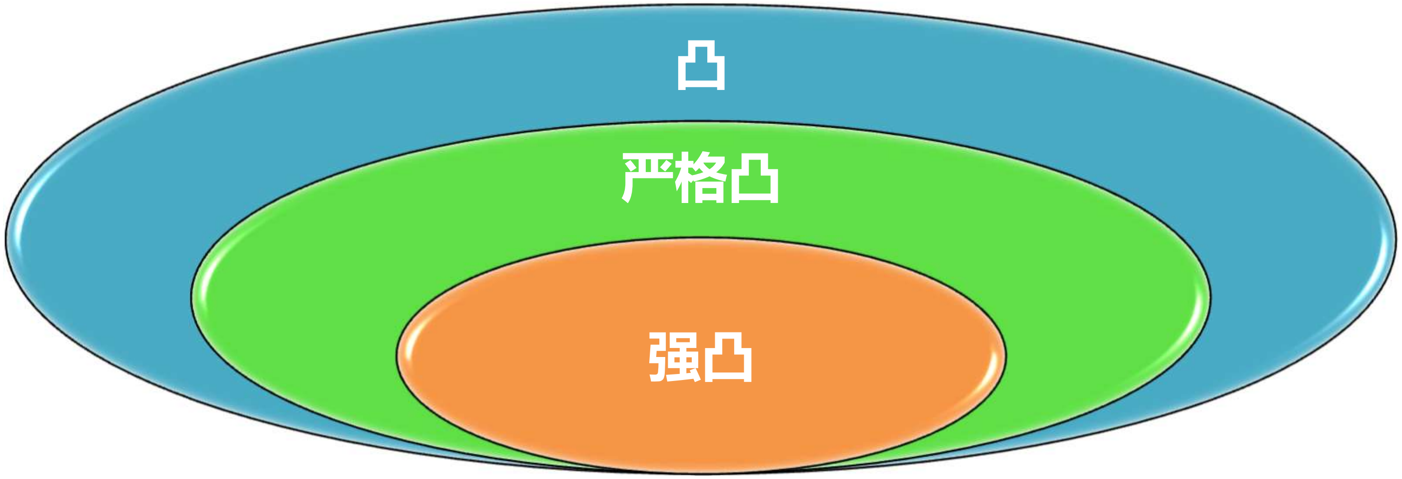
严格凸：函数 $f\colon \mathbb{R}^n \to \mathbb{R}$ 是凸函数，若其定义域 $\mathrm{dom}[f]$ 是凸集，且

$$f(\theta x_1 + (1-\theta)x_2) < \theta f(x_1) + (1-\theta)f(x_2)$$

$$\forall x_1, x_2 \in \mathrm{dom}[f], x_1 \neq x_2, \theta \in (0,1)$$

强凸：称函数 f 是强凸函数，若$\exists \alpha > 0$使$f(x) - \alpha\|x\|_2^2$仍是凸函数

函数 $f\colon \mathbb{R}^n \to \mathbb{R}$ 严格凹（强凹）若 $-f$ 严格凸（强凸）



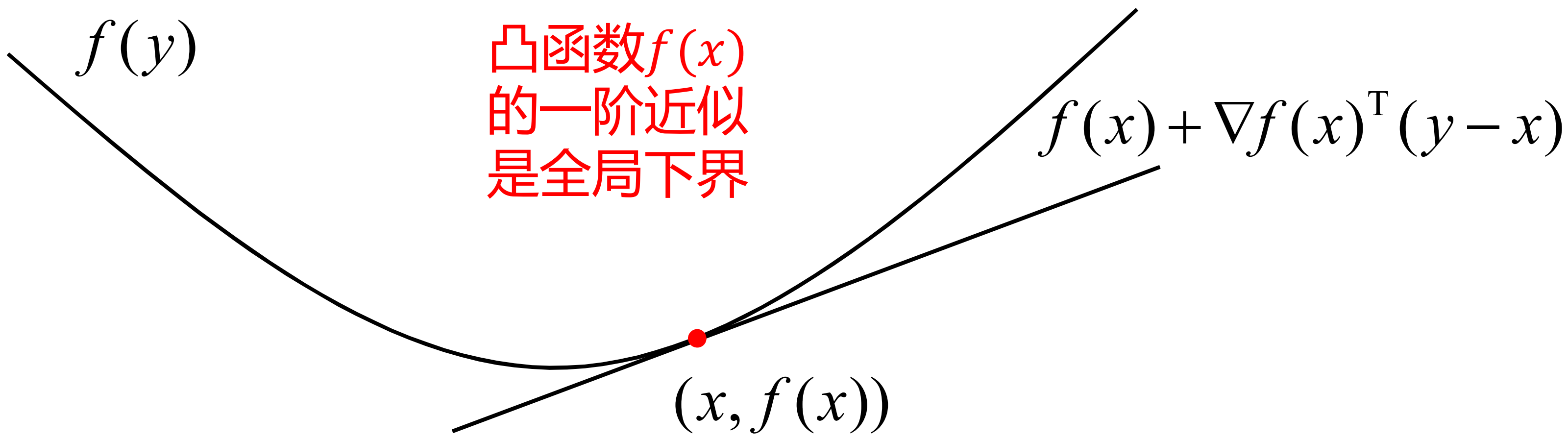
1.2 判断函数的凸性

假设$f(x)$可导，其梯度 $\nabla f(x)$ 在定义域 $\mathrm{dom}[f]$ 上存在.

判断函数凸性的一阶条件

定义域为凸集的可导函数 $f(x)$ 是凸函数当且仅当

$$f(y) \geq f(x) + \nabla f(x)^{\mathrm{T}}(y - x), \forall x, y \in \mathrm{dom}[f]$$



1.2 判断函数的凸性

假设函数 $f(x)$ 二阶可导，其海森矩阵

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

在定义域 $\mathrm{dom}[f]$ 内每个点上都存在

$f(x)=x^4$

凸性与海森矩阵的关系

对于定义域为凸集的二阶可导函数 f：

- f 是凸函数当且仅当$\nabla^2 f(x) \succcurlyeq 0, \forall x \in \mathrm{dom}[f]$；
- 若$\nabla^2 f(x) \succ 0, \forall x \in \mathrm{dom}[f]$，则 f 是严格凸函数；
- f 是强凸函数当且仅当$\exists \varepsilon > 0, \nabla^2 f(x) \succcurlyeq \varepsilon I, \forall x \in \mathrm{dom}[f]$

单变量凸函数举例

- 仿射函数 $f(x) = ax + b, dom[f] = \mathbb{R}$, 对任意 a, b
- 二次函数 $f(x) = ax^2 + bx + c, dom[f] = \mathbb{R}, a > 0$
- 指数函数 $f(x) = e^{ax}$, 对任意 a
- 幂函数 $f(x) = x^a, dom[f] = \mathbb{R}_{++}$, $a \geq 1$ 或 $a \leq 0$
- 负熵函数 $f(x) = x\log(x), dom[f] = \mathbb{R}_{++}$

单变量凹函数举例

- 仿射函数 $f(x) = ax + b, dom[f] = \mathbb{R}$, 对任意 a, b
- 二次函数 $f(x) = ax^2 + bx + c, dom[f] = \mathbb{R}, a < 0$
- 幂函数 $f(x) = x^a, dom[f] = \mathbb{R}_{++}$, $0 < a < 1$
- 对数函数 $f(x) = \log(x), dom[f] = \mathbb{R}_{++}$

1.2 判断函数的凸性

定义在$\mathbb{R}^n, n \geq 2$上的凸函数

- 定义在 $\mathbb{R}^n$ 上的范数是凸函数

$$\begin{aligned} \|tx+(1-t)y\| &\leq \|tx\|+\|(1-t)y\| \quad \text{三角不等式} \\ &= t\|x\|+(1-t)\|y\| \quad \text{齐次性} \end{aligned}$$

- 二次型函数 $f(x)=x^{\mathrm{T}}Qx/2+b^{\mathrm{T}}x+c$

$$\nabla^2 f(x)=Q \qquad Q \succeq 0 \;\Leftrightarrow\; f(x) \text{ is convex}$$

这个二次函数严格凸吗? $f(x) = x_1^2 + x_2^2, x \in \mathbb{R}^3$

- 二次-线性分式函数 $f(x)=x^2/y,\ y>0$

$$\nabla^2 f(x,y)=\frac{2}{y^3}\begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix}=\frac{2}{y^3}\begin{bmatrix} y \\ -x \end{bmatrix}\begin{bmatrix} y \\ -x \end{bmatrix}^{\mathrm{T}} \succcurlyeq 0$$

1.2 判断函数的凸性

保凸运算

- 凸函数 $f_1, \cdots, f_m$ 的非负加权和
$$f = \lambda_1 f_1 + \cdots + \lambda_m f_m \quad \lambda_1, \cdots, \lambda_m \geq 0$$
- 凸函数和仿射函数的复合： $g(x) = f(Ax + b)$
- 凸函数 $f_1, \cdots, f_m$ 逐点求极大：
$$f(x) = \max\{f_1(x), \cdots, f_m(x)\}$$

举例:

➢ 线性不等式的对数障碍函数

$$f(x) = -\sum_i \log(b_i - a_i^{\mathrm{T}} x) \quad \mathrm{dom}\, f = \{x \mid Ax < b\}$$

➢ 2范数与仿射函数的复合

$$f(x) = \|Ax + b\| \qquad \nabla^2 f(x) = 2A^{\mathrm{T}}A \succcurlyeq 0$$

1.2 判断函数的凸性

证明：若 $f_1,\cdots,f_m$ 是凸函数，则$\max\{f_1,\cdots,f_m\}$是凸函数

$$f(tx+(1-t)y)=\max_i\{f_i(tx+(1-t)y)\}\ \le \max_i\{tf_i(x)+(1-t)f_i(y)\}$$

$$\le t\max_i\{f_i(x)\}+(1-t)\max_i\{f_i(y)\}=tf(x)+(1-t)f(y)$$

实例

- 如下形式的分段线性函数

$$f(x)=\max\{a_1^{\mathrm{T}}x+b_1,\cdots,a_m^{\mathrm{T}}x+b_m\}$$

- 对阵矩阵的最大特征值

$$f(X)=\lambda_{\max}(X)=\sup\left\{y^{\mathrm{T}}Xy\ |\ \|y\|_2=1\right\}$$

将$y^{\mathrm{T}}Xy=\mathrm{Tr}\left(yy^{\mathrm{T}}\cdot X\right)$视为以给定的 y 为系数的关于X的线性函数，$f(X)$是系数满足$\|y\|_2=1$的所有线性函数$y^{\mathrm{T}}Xy$逐点求极大的结果

主要内容

1. 凸函数
2. 凸优化
3. 拉格朗日对偶
4. KKT最优性条件

2.1 凸优化问题的一般形式

$$
\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & g_i(x) \le 0, i = 1, \cdots, m \\
& Ax = b
\end{aligned}
$$

- $x \in \mathbb{R}^n$ 是决策变量
- 目标函数 $f: \mathbb{R}^n \to \mathbb{R}$ 是凸函数(或max凹函数)
- 约束函数 $g_i: \mathbb{R}^n \to \mathbb{R}$ 是凸函数
- 等式约束是线性方程 (为什么只能是线性的？)
- 最优值: $v^* \le f(x), \forall x \in X$
- 若极小化问题不可行，通常记 $v^* = \infty$
- 若极小化问题无界，通常记 $v^* = -\infty$
- 局部极小解 x^* 与局部极小值 v^*:

$$v^* \le f(x^*),\ \forall x \in X \cap B(x^*, \varepsilon)$$

2.1 凸优化问题的一般形式

等价转换

1. 目标函数的单调变换不影响最优解($\Gamma: \mathbb{R} \to \mathbb{R}$ 严格单增)

$$\begin{aligned} \min\ & f(x) \\ \text{s.t.}\ & g(x) \le 0 \\ & Ax = b \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} \min\ & \Gamma(f(x)) \\ \text{s.t.}\ & g(x) \le 0 \\ & Ax = b \end{aligned}$$

2. 上境图转换

$$\begin{aligned} \min\ & f(x) \\ \text{s.t.}\ & g(x) \le 0 \\ & Ax = b \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} \min\ & t \\ \text{s.t.}\ & g(x) \le 0,\ f(x) \le t \\ & Ax = b \end{aligned}$$

- 任何优化问题都可以转化为优化一个线性目标函数
- 任何非凸优化问题都可以转化成一个“凸优化”问题

$$\min_{x,t} \left\{ t \,\middle|\, \operatorname{conv}\left(\left\{ (x,t) \,\middle|\, x \in X, t \ge f(x) \right\} \right) \right\}$$

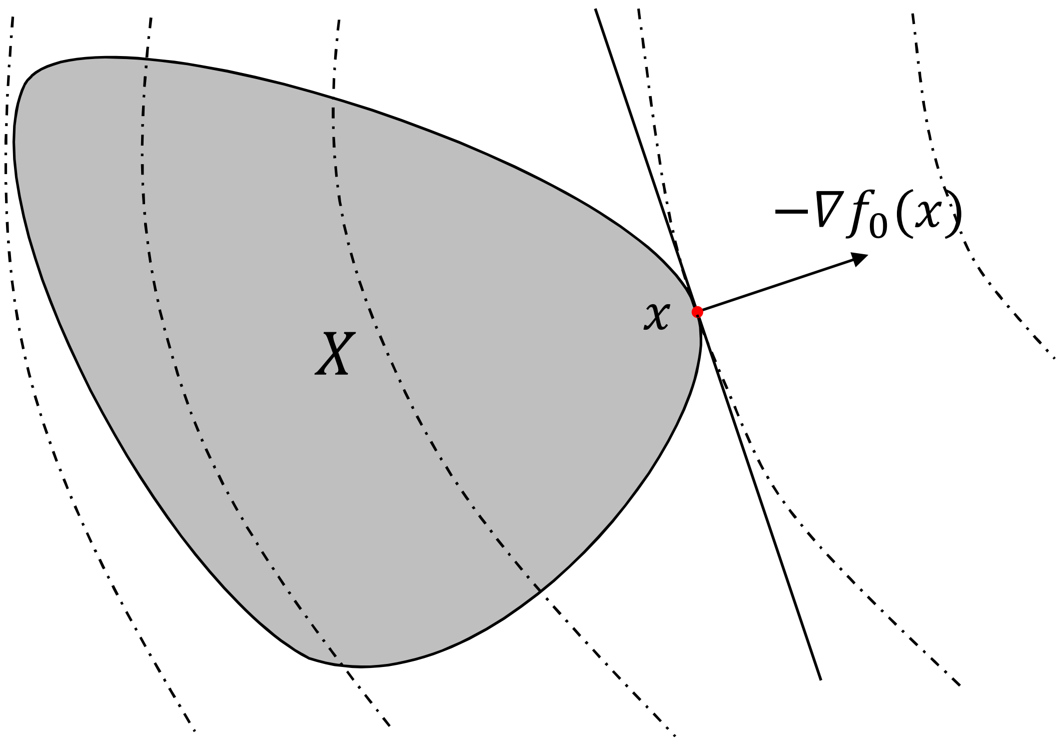
但凸包表达式未知!

2.2 最优解的性质

定理1 （一阶最优性条件）

假设凸优化目标函数可微，则$x^* \in X$ 是最优解当且仅当

$$\nabla f(x^*)^{\mathrm{T}}(y - x^*) \geq 0, \forall y \in X$$



2.2 最优解的性质

例1: 论证 $x^* = (1,0.5,-1)$ 是如下优化问题的最优解

$$\begin{aligned} \min \quad & 0.5x^{\mathrm{T}}Px + q^{\mathrm{T}}x \\ \text{s.t.} \quad & -1 \le x_1, x_2, x_3 \le 1 \end{aligned}$$

其中

$$P = \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix} \qquad q = \begin{bmatrix} -22.0 \\ -14.5 \\ 13.0 \end{bmatrix}$$

计算目标函数的梯度

$$\nabla f(x^*) = Px + q = (-1, 0, 2)^\top$$

最优性条件

$$\nabla f(x^*)^\top (y - x) = -1\left(y_1 - 1\right) + 2\left(y_3 + 1\right) \ge 0$$

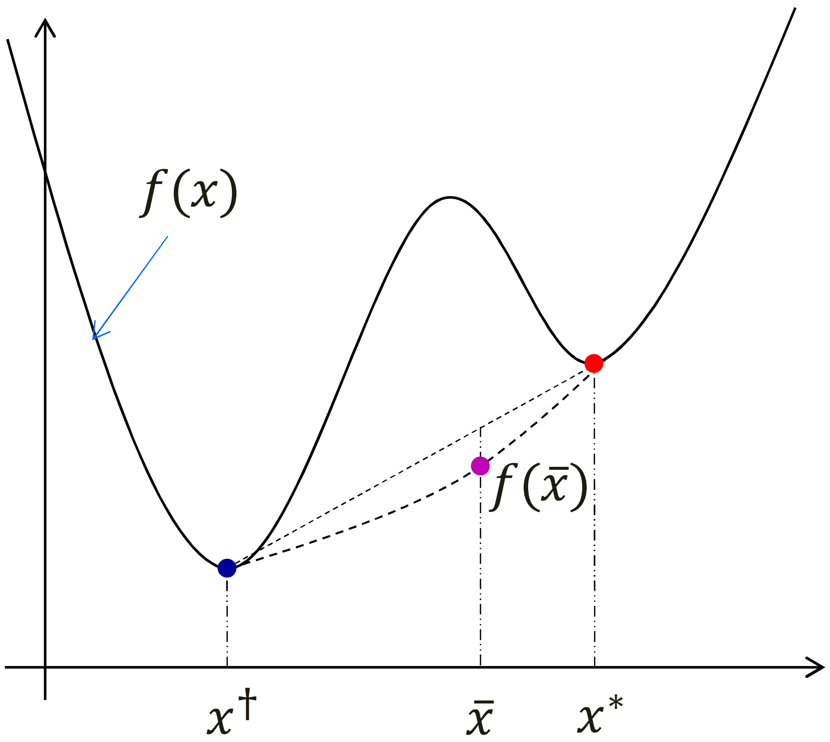
对任意 $y, -1 \le y_1, y_2, y_3 \le 1$ 成立。故 x^* 是最优解。

2.2 最优解的性质

定理2

凸优化问题的局部最优解也是全局最优解

证明 反证法。假设全局最优解为 $x^\dagger$，而 $x^* \neq x^\dagger$ 是一个局部最优解，这意味着 $f(x^\dagger) < f(x^*)$。当 $\alpha \in (0,1)$ 时，点 $\bar{x} = \alpha x^* + (1-\alpha)x^\dagger$在连接 $x^\dagger$ 和 x^* 的线段上移动。由于可行域为凸集，线段上的任何一点都是可行解。根据凸函数的性质



$$
\begin{aligned}
f(\bar{x}) &\leq \alpha f(x^*) + (1-\alpha) f(x^\dagger) \\
&\leq \alpha f(x^*) + (1-\alpha) f(x^*) = f(x^*)
\end{aligned}
$$

令 $\alpha \to 1, \bar{x} \to x^*$, 但 $f(\bar{x}) \leq f(x^*)$。这与 x^* 是局部最优解矛盾。

3.1 拉格朗日函数

$$
\begin{aligned}
\min\ & f(x) \\
\text{s.t. } & g_i(x) \le 0, i = 1, \cdots, m \\
& h_i(x) = 0, i = 1, \cdots, p
\end{aligned}
$$

- 决策变量 $x \in \mathbb{R}^n$, 最优值为 p^*

拉格朗日函数: $L : \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}^p \to \mathbb{R}$

$$L(x, \lambda, v) = f(x) + \lambda^{\mathrm{T}} g(x) + v^{\mathrm{T}} h(x)$$

- 是目标函数和约束函数的加权组合
- $\lambda \geq 0$ 是对应于≤型不等式约束的乘子向量
- v无符号限制，是对应于等式约束的乘子向量

3.2 对偶函数与弱对偶性

拉格朗日对偶函数 $F: \mathbb{R}_+^m \times \mathbb{R}^p \to \mathbb{R}$

$$F(\lambda,v)=\inf_{x\in\mathrm{R}^n} L(x,\lambda,v)$$

min v.s. inf

$$=\inf_{x\in\mathrm{R}^n}\left\{f(x)+\lambda^{\mathrm{T}}g(x)+v^{\mathrm{T}}h(x)\right\}$$

- F 是 (λ, v) 的凹函数，对某些 λ, v 可能无界

定理 3

对偶函数值与原问题目标函数值存在如下关系：

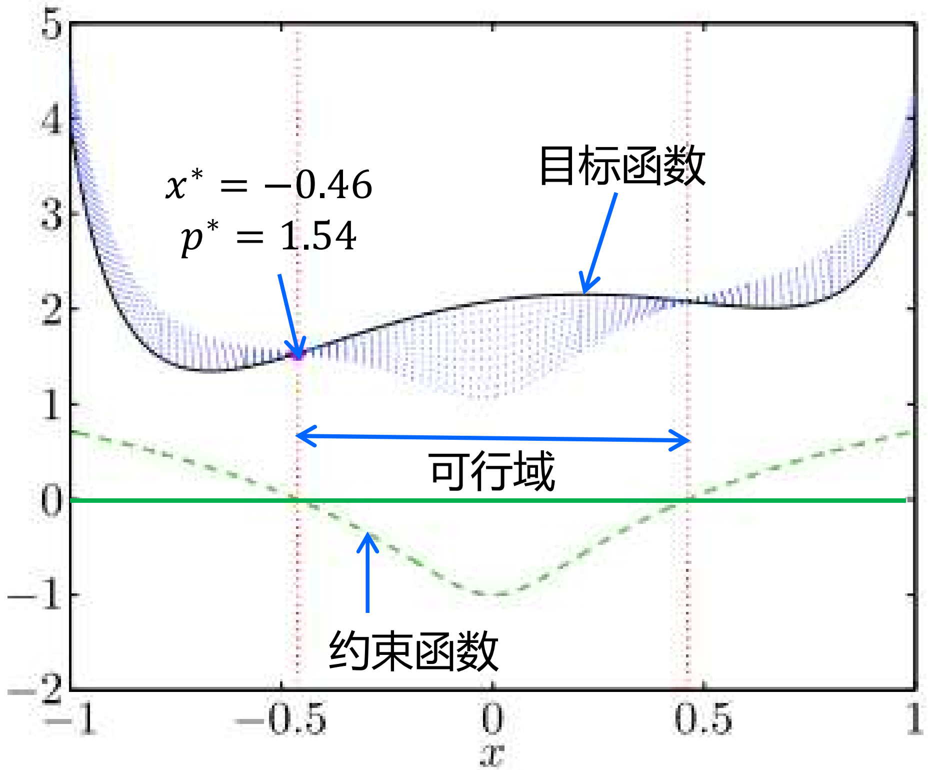
$$F(\lambda,v)\le p^*\le f(x)$$

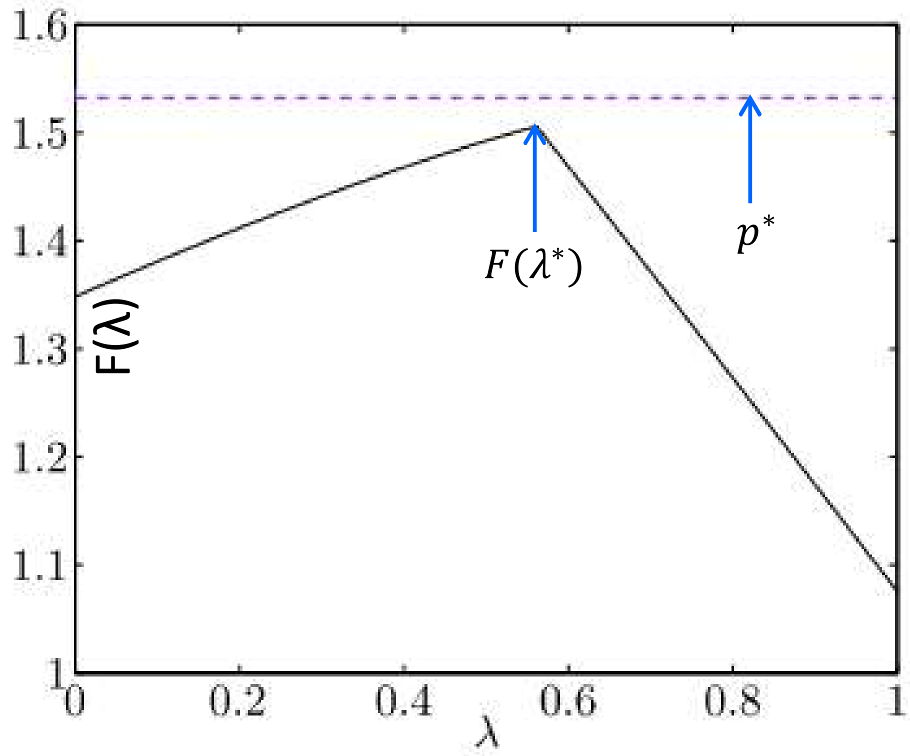
证明: 假设 $F(\lambda,v)$ 有界，原问题的最优解是 x^*，满足约束条件 $g(x^*)\le 0, h(x^*)=0$ 。对任意 $\lambda\ge 0$

$$p^*=f(x^*)\ge f(x^*)+\underbrace{\lambda^{\mathrm{T}}g(x^*)}_{\le 0}+\underbrace{v^{\mathrm{T}}h(x^*)}_{=0}=L(x^*,\lambda,v)\ge F(\lambda,v)$$

$F(\lambda,v)$的定义

3.2 对偶函数与弱对偶性





当 λ 分别取 $0.1, 0.2, \cdots, 1.0$ 等10个值时的拉格朗日函数 $L(x,\lambda)$ 每种情况下 $L(x,\lambda)$ 关于x 的极小值都小于原问题的最优值 p^*.

- 目标函数$f(x)$和约束函数$g(x)$都不是凸函数，但对偶函数 $F(\lambda)$ 是关于 λ 的凹函数
- 本例中对偶间隙大于0

3.2 对偶函数与弱对偶性

通过对罚函数近似理解弱对偶性

$$\min f_0(x) + \sum_{i=1}^{m} I^{-}(g_i(x)) + \sum_{i=1}^{m} I^{0}(h_i(x))$$

罚函数 $I^{-}(u) = \begin{cases} 0 & u \le 0 \\ \infty & u > 0 \end{cases}$ $\quad I^{0}(u) = \begin{cases} 0 & u = 0 \\ \infty & u \neq 0 \end{cases}$

线性函数是罚函数的全局下界

$$I^{-}(u) = \lambda u \qquad I^{0}(u) = \mu u$$

$$g(\lambda, \mu) = \min_x L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{m} \mu_i h_i(x)$$

3.3 对偶问题

拉格朗日对偶问题

$$\max_{\lambda,v} \ F(\lambda,v)$$
$$\text{s.t.} \ \ \lambda \geq 0$$

- 对偶问题是凸优化，不论原问题是否为凸优化
- 对偶最优解: λ^*, v^*
- 对偶最优值: F^*

原始-对偶最优值的关系

弱对偶性: $F^* \leq p^*$ (对凸优化和非凸优化都成立)

强对偶性: $F^* = p^*$ (在一定条件下对凸优化成立)

3.4 强对偶性

约束规范: 这里指强对偶在凸优化问题中成立需要的条件

Slater条件: 强对偶对以下凸优化问题成立

$$
\begin{aligned}
\min \quad & f(x) \\
s.t. \quad & g_i(x) \le 0, i = 1, \cdots, m \\
& Ax = b
\end{aligned}
$$

若其严格可行，即 $\exists x:\ g(x) < 0, i = 1, \cdots, m,\ Ax = b$

- 对非线性不等式约束存在严格可行解
- 线性不等式约束不需要存在严格可行解
- Slater条件是一个常见的，并非唯一的约束规范

主要内容

1. 凸函数
2. 凸优化
3. 拉格朗日对偶
4. KKT最优性条件

4.1 互补松弛

假设强对偶成立, x^* 是原始最优解, (λ^*, ν^*) 是对偶最优解

原始最优值		对偶最优值	
$\cancel{f(x^*)}$	$=$	$\cancel{f(x^*)} + \lambda^{*\mathrm{T}} g(x^*) + \nu^{*\mathrm{T}} h(x^*)$	=0

$$\boxed{\begin{aligned} &\lambda^{*\mathrm{T}} g(x^*) = 0 \\ &\lambda^* \ge 0,\ g(x^*) \le 0 \end{aligned}}$$

$$\lambda_i^* g_i(x^*) = 0,\ i = 1, \cdots, m$$

$$\lambda_i^* > 0 \Rightarrow g_i(x^*) = 0, \quad g_i(x^*) < 0 \Rightarrow \lambda_i^* = 0$$

4.2 Karush-Kuhn-Tucker (KKT)条件

KKT条件

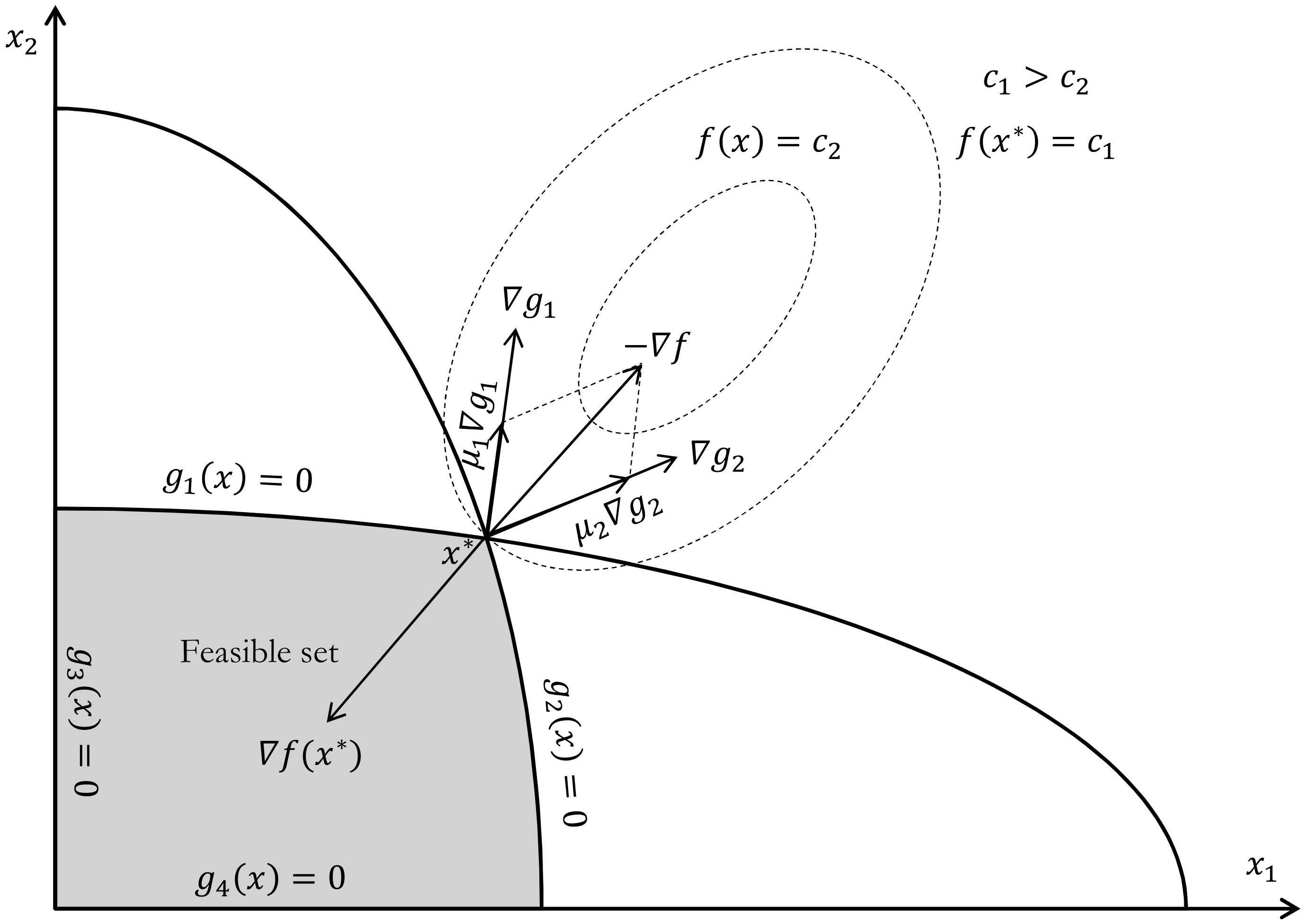
- 原始可行 $g(x) \leq 0,\ h(x) = 0$
- 对偶可行 $\lambda \geq 0$
- 互补松弛 $\lambda_i g_i(x) = 0, \forall i$ (或 $0 \leq \lambda \perp g(x) \leq 0$)
- 驻点条件 $\nabla f(x) + \lambda^{\mathrm{T}} \nabla g(x) + \nu^{\mathrm{T}} \nabla h(x) = 0$

若Slater条件满足，x 是最优解当且仅当存在 λ, ν 满足KKT条件

- 将无约束优化的最优性条件 $\nabla f(x) = 0$ 推广到约束优化
- 将优化问题转化为方程问题（计算上变简单了吗？）
- 验证简单求解难！
- 对非凸优化问题，KKT条件是必要条件

4.2 KKT条件

KKT条件的几何解释



4.3 线性规划对偶

$$\begin{aligned}\min\quad & c^{\mathrm{T}}x \\ \text{s.t.}\quad & Ax \le b : \lambda\end{aligned} \quad \overset{\text{对偶}}{\Longleftrightarrow} \quad \begin{aligned}\max\quad & b^{\mathrm{T}}\lambda \\ \text{s.t.}\quad & A^{\mathrm{T}}\lambda = c : x \\ & \lambda \le 0\end{aligned}$$

拉格朗日函数: $L(x,\lambda) = c^{\mathrm{T}}x + \lambda^{\mathrm{T}}\left(Ax - b\right), \lambda \ge 0$

对偶函数: $g(\lambda) = \inf_x L(x,\lambda) = -b^{\mathrm{T}}\lambda + \inf_x (A^{\mathrm{T}}\lambda + c)^{\mathrm{T}}x$

$$= \begin{cases} -b^{\mathrm{T}}\lambda & A^{\mathrm{T}}\lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

对偶问题为

$$\begin{aligned}\max\quad & -b^{\mathrm{T}}\lambda \\ \text{s.t.}\quad & A^{\mathrm{T}}\lambda + c = 0, \lambda \ge 0\end{aligned} \quad \text{或} \quad \begin{aligned}\max\quad & b^{\mathrm{T}}\lambda \\ \text{s.t.}\quad & A^{\mathrm{T}}\lambda = c, \lambda \le 0\end{aligned}$$

4.3 线性规划对偶

线性规划的最优性条件

$$\begin{aligned} \min \quad & c^{\mathrm{T}} x \\ s.t. \quad & Cx \le d : \lambda \\ & Ax = b : v \end{aligned}$$

KKT 条件

$L(x,\lambda,v) = c^{\mathrm{T}}x + \lambda^{\mathrm{T}}(Cx-d) + v^{\mathrm{T}}(Ax-b)$

$c + C^{\mathrm{T}}\lambda + A^{\mathrm{T}}v = 0$

$Cx \le d,\ Ax = b$

$\lambda \ge 0$

$\lambda^{\mathrm{T}}(Cx-d) = 0$

原-对偶条件

$$\begin{aligned} \max \quad & d^{\mathrm{T}}\lambda + b^{T}v \\ \text{s.t.} \quad & C^{\mathrm{T}}\lambda + A^{\mathrm{T}}v = c, \lambda \le 0 \end{aligned}$$

$c - C^{\mathrm{T}}\lambda - A^{\mathrm{T}}v = 0$

$Cx \le d,\ Ax = b$

$\lambda \le 0$

$d^{\mathrm{T}}\lambda + b^{\mathrm{T}}v = c^{\mathrm{T}}x$

例2 考虑如下优化问题

$$\min\left\{x_1^2 + x_2^2 : x_1 + x_2 \geq 4\right\}$$

（1）找出所有满足KKT条件的点

（2）写出并求解对偶问题

（1）写出KKT条件

$$\boxed{\begin{aligned} 2x_1 - \lambda &= 0 \\ 2x_2 - \lambda &= 0 \end{aligned}}$$

$$x_1 + x_2 \geq 4,\ \ \lambda \geq 0$$

$$\boxed{\lambda(4 - x_1 - x_2) = 0}$$

若 $x_1 + x_2 > 4$, 则 $\lambda = 0$

$$\Rightarrow \begin{aligned} x_1 &= 0 \\ x_2 &= 0 \end{aligned} \quad \Rightarrow \quad x_1 + x_2 \ngeq 4$$

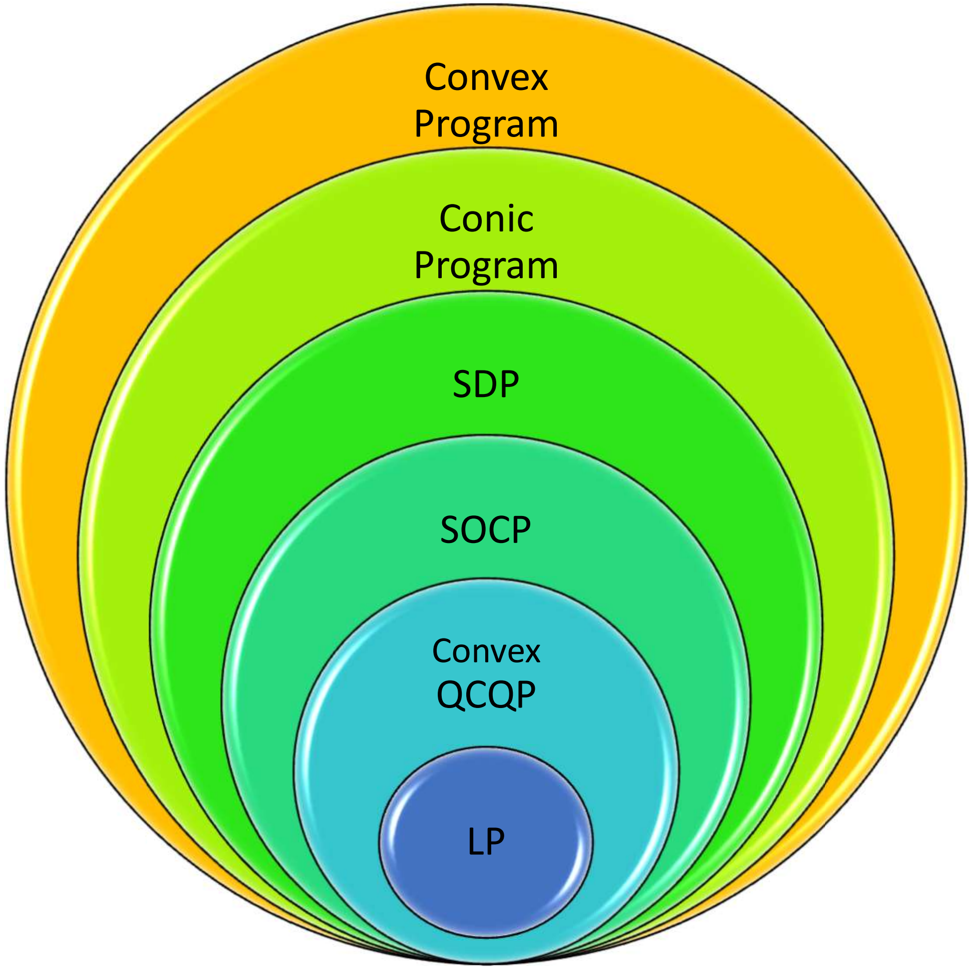
若 $\lambda > 0$, 则 $x_1 + x_2 = 4$

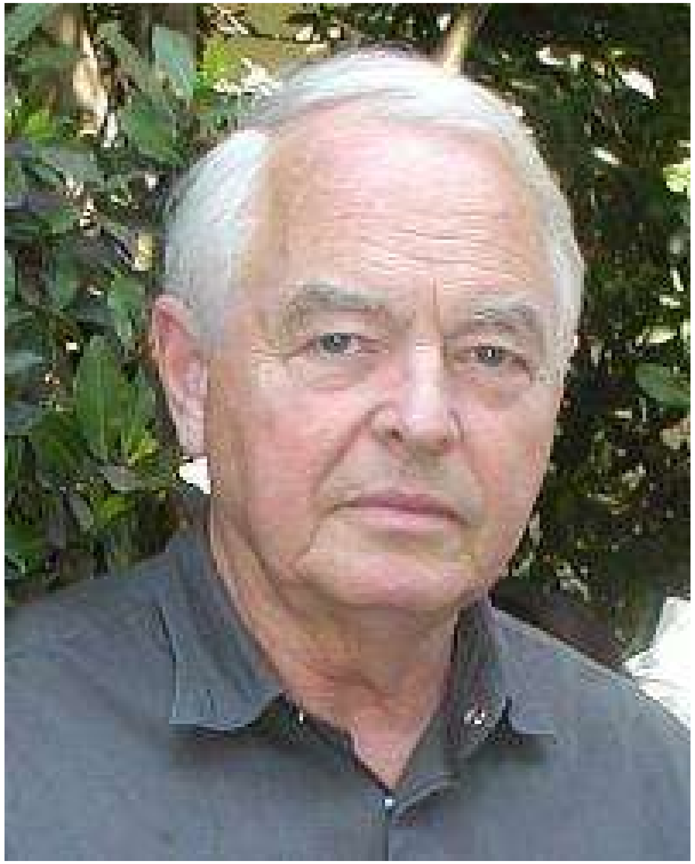
$$\Rightarrow x_1 = x_2 = 2, \quad \lambda = 4$$

（2）拉格朗日函数为 $L(x, \lambda) = x_1^2 + x_2^2 - \lambda(x_1 + x_2 - 4)$

对偶函数为 $F(\lambda) = \min\limits_{x} L(x, \lambda) = -\lambda^2/2 + 4\lambda$

对偶最优解为 $\lambda = 4$ ， 最优值为8，对偶间隙为0





The great watershed in optimization isn't between linearity and nonlinearity, but convexity and non-convexity.

-- R. Tyrrell Rockafellar