# Boyd 的凸优化

刘大爷

May 17, 2025

我去年九月时问起凸集是否定义在向量空间上，未得到有效答复。现看夏道行忆起前事。凸集是泛函分析中常用的概念，在线性空间上半范数的研究需要凸集。对于线性空间中任意半范数 $p(x)$，集 $\{x|p(x) \leq 1\}$ 就是一个凸集，称它为范数 $p$ 所导出的凸集。总之具有许多性质，但我不知道。

## 0.1 凸集

### 0.1.1 仿射集合和凸集

1. 仿射集合。下面的 $x_1, x_2$ 都是任意的（在其域中）

   - 直线与线段
   - 集合是仿射的，如果过集合中任意两个不同的的直线仍在集合中，即：

   $$\theta x_1 + (1-\theta)x_2 \in C \tag{1}$$

   - 仿射组合如果 $\sum_k \theta_k x_k \in C, \sum_k \theta_k = 1$，则称为 $\{x_k\}$ 的仿射组合。
   - 子空间。仿射集合 $C$ 的子空间为：

   $$V = c - x_0 = \{x - x_0 | x \in C\} \tag{2}$$

     易证其对于加法和数乘封闭。

   - 仿射包：

   $$\textbf{aff}C = \{\theta_1 x_1 + ... + \theta_k x_k | x_1, ..., x_k \in C, \theta_1 + ... + \theta_k = 1\} \tag{3}$$

     仿射包是包含 $C$ 的最小仿射集合

   - 仿射维数与相对内部因定义不够清晰，暂时跳过。

2. 凸集

   - 定义

   $$\theta x_1 + (1-\theta)x_2 \in C, \theta \in [0,1] \tag{4}$$

   - 凸组合

     如果 $\sum_k \theta_k x_k \in C, \sum_k \theta_k = 1, \theta_k > 0$，则称为 $\{x_k\}$ 的凸组合。一个集合是凸集等价于集合包含其中所有点的凸组合。

- 凸包：

$$\text{conv}C = \{\theta_1 x_1 + ... + \theta_k x_k | x_1, ..., x_k \in C, \theta_1 + ... + \theta_k = 1, \theta_k \geq 0\} \tag{5}$$

  凸是包含 $C$ 的最小凸集合。

3. 锥

- 凸锥定义

$$\theta_1 x_1 + \theta_2 x_2 \in C, \theta x \in C, \theta \geq 0 \tag{6}$$

  锥不一定是凸的。上述定义是针对凸锥的充要条件。

- 锥组合

  如果 $\sum_k \theta_k x_k \in C, \theta_k \geq 0$，则称为 $\{x_k\}$ 的锥组合。一个集合是凸锥等价于集合包含其中所有点的锥组合。

- 集合 $C$ 的锥包是其中所有元素所有锥组合的集合，是包含 $C$ 的最小凸锥。因此可以说，锥包是凸锥，锥包一定是凸的。

### 0.1.2 重要例子

在此前需要明确一些概念。就一般而言，仿射集合是定义在向量空间上的，但是仿射空间不一定具有加法单位元和逆元，因此更多是一个集合而非向量空间。由仿射集合定义在向量空间上，因此其性质在欧几里得空间也成立；子空间定义在向量空间上，因此也满足仿射集合的性质。

1. 重要例子。直线经过零点，具有加法单位元，才符合子空间的性质。

2. 超平面是集合：

$$\{x | a^T x = b\} \tag{7}$$

   上述定义等价于

$$\{x | a^T (x - x_0) = 0\}$$
$$\{x | a^T (x - x_0) = 0\} = x_0 + a^\perp$$
$$\{v | a^T v = 0\} = a^\perp$$

   半空间为上述取“$\leq$”。

3. 欧几里得球和椭球（感觉太 trivial 了）

$$\epsilon = \{x + Au | \|u\|_2 \leq 1\} \tag{8}$$

显然 $A = r$ 时为球的定义。一般来说，$A$ 对称半正定，但当是奇异时称为退化的。椭球也有如此定义：

$$\{x | (x - x_c)^T P^{-1} (x - x_c) \leq 1\}$$

其中 $P$ 一般是对称正定的

4. 范数锥

$$C = \{(x, t) | \|x\| \leq t\} \tag{9}$$

既然是范数，满足范数的运算定理。范数锥和范数球是凸的。

5. 多面体

$$\mathcal{P} = \{x | a_j^T x \leq b, c_j^T x = d_j\} \tag{10}$$

注意到等式约束 $c_j^T x = d_j$ 可能是冗余的。

6. 单纯形（不考）

$$C = \text{conv}\{v_i\} = \{\sum_{i=0}^{k} \theta_i v_i | \theta_i \geq 0, 1^T \theta = 1\} \tag{11}$$

7. 半正定锥

$$S^n = \{X \in R^{n \times n} | X = X^T\} \tag{12}$$

此为对称矩阵。半正定、正定锥同理。显然半正定矩阵的集合 $S_+^n$ 是凸锥

### 0.1.3 保凸运算

1. 交集。凸集的交集

2. 仿射函数。一个函数是凸的，如果其为一个线性函数和一个常数的和，即为 $f(x) = Ax + b$。其中 $f : \mathrm{R}^n \to \mathrm{R}^m$。因为仿射集定义在向量空间上，因此仿射集的数乘和加法是保凸的。向量空间是仿射的，任何仿射集都是向量空间的子空间，而因此对于 $x \in S_1, y \in S_2, x + y \in V$，仍是仿射的。直积（有序对）和部分和是仿射的，但直和是否是仿射的有待商榷。如果 $S_1, S_2$ 是无关的，则两仿射集的和亦为直和。仿射变换有：伸缩、平移、投影、集合的和、直积、部分和

3. 线性分式与透视函数。比仿射函数更普遍

- 透视函数
$$P(z,t)=z/t \tag{13}$$
没有对定义域有凸的要求。

- 线性分式函数
$$f(x)=(Ax+b)/(c^Tx+d) \quad \mathbf{dom} f=\{x|c^Tx+d>0\} \tag{14}$$
由透视函数和仿射函数复合而成：$g:\mathrm{R}^n\to\mathrm{R}^{m+1}$，$P:\mathrm{R}^{m+1}\to\mathrm{R}^m$，$f=P\circ g:\mathrm{R}^n\to\mathrm{R}^m$

比如，条件概率。显然 $A$ 与 $c$ 都可为 0。

### 0.1.4 广义不等式

**正常锥**

1. 正常锥的定义
$$N_c(x)=\{v\in\mathrm{R}^n:(v,y-x)\le 0,\forall y\in C\} \tag{15}$$
其中 $C$ 是凸集，$v$ 表示指向凸集外部的向量。$x$ 必须在凸集 $C$ 中。

2. 正常锥的性质：
凸、闭、实（非空内部，注线不是实的）、尖的

3. 正常锥定义偏序关系：
$$x\preceq_K y\longleftrightarrow y-x\in \mathrm{K}$$

**广义不等式**

严格的广义不等式许多性质和一般的广义不等式类似

1. 对于加法保序
$$x+u\preceq_K y+v$$

2. 具有传递性

3. 非负数乘保序

4. 自反的（严格的广义不等式不满足）
$$x\preceq_K x$$

关于正常锥和广义不等式有两个典型例子：非负象限和半正定锥。很 trival 吧。

**最小元与极小元**

1. 最小元 $x$

$$x \preceq_K y \quad y \in S$$

2. 极小元 $x$

$$y \preceq_k x \quad y \in S, y = x$$

### 0.1.5 分离与支撑超平面

1. 超平面 $\{x|a^T x = b\}$ 为集合 $C$ 和 $D$ 的分离超平面，若两者为凸集且交集为空集

2. 如果 $a \neq 0$ 且对任意 $x \in C$ 满足 $a^T x \leq a^T x_0$，则称超平面 $\{x|a^T x = a^T x_0\}$ 为集合 $C$ 在 $x_0$ 处支撑超平面

**对偶锥**

令 $K$ 是一个锥，集合

$$K^* = \{y|x^T y \geq 0, \forall x \in K\}$$

称为 $K$ 的对偶锥，即使 $K$ 不是凸锥

Proof. 若在实向量空间上 $y^T x$ 等价于内积（即满足锥性），因此

$$\begin{aligned}\theta y_1^T x + (1-\theta) y_2^T x &= \theta(y_1, x) + (1-\theta)(y_2, x) \\ &= (\theta y_1 + (1-\theta) y_2, x) \\ &= (y, x)\end{aligned}$$

凸性得证 □

更严谨的说法，对偶锥是一系列过原点的超平面的交

**帕累托最优制造**

1. 生成集合 P 上的极小元对应的制造方法称为帕累托最优的，$P$ 的极小元构成的集合为帕累托最优制造前沿（什么玩意儿）

## 0.2 凸函数

### 0.2.1 基本性质和例子

1. 凸函数
$$f(\theta x+(1-\theta)y)\leq\theta f(x)+(1-\theta)f(y) \tag{16}$$
dom$f$ must be convex set. 如果不等号成立，则为严格的。注意, $\theta\in(0,1), x\neq y$ 时为严格凸的（凸的只需 $\theta\in[0,1]$）

2. 仿射函数既凸又凹

3. 函数是凸的，当且仅当其在与其定义域相交的任何直线上都是凸的。这个命题等价于当且仅当对于任意 $x\in$ dom$f$ 和任意向量 $v$，函数 $g(t)=f(x+tv)$ 是凸的（$\{t|x+tv\in$ dom$f\}$）

4. 扩充实值函数

5. 一阶条件（陈述是充要的），如果 dom$f$ 是凸集且有
$$f(y)\geq f(x)+\nabla f(x)^T(y-x) \tag{17}$$

Proof.
$$\nabla f(x)^T=\frac{f(x+\theta(y-x))-f(x)}{\theta(y-x)}$$
$$f(x+\theta(y-x))=f(x)+\theta\nabla f(x)^T(y-x)$$

根据凸函数的定义可得:
$$\theta f(y)+(1-\theta)f(x)\geq f(\theta y+(1-\theta)x)$$
则有:
$$\begin{aligned}\theta f(y)+(1-\theta)f(x)&\geq f(x)+\theta\nabla f(x)^T(y-x)\\ f(y)&\geq f(x)+\nabla f(x)^T(y-x)\end{aligned}$$
□

注意 $\nabla f(x)$ 是列向量。严格凸的则不等式成立

6. 二阶条件（是充要的），$\forall x \in \mathrm{dom} f$（$\mathrm{dom} f$ 是开集）有

$$\nabla^2 f(x) \succeq 0 \tag{18}$$

其中 $\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$。注意，严格凸函数推出二阶条件是充分不必要的。

7. 凸函数实例（不常见的）

- 负熵.$xlogx$. 二阶导严格凸，二阶充要条件证明其。
- 范数. 都是凸的

  Proof. 由三角不等式

  $$\|x+y\| \leq \|x\| + \|y\|$$

  和齐次性可得

  $$f(\theta x + (1-\theta)y) \leq f(\theta x) + f((1-\theta)y) = \theta f(x) + (1-\theta) f(y)$$

  □

- 仿射函数既凸又凹
- 最大值函数是凸的

  Proof. 由

  $$\theta x + (1-\theta)y \leq \theta x^* + (1-\theta)y^*$$

  （其中 $x^* = \max\{x_n\}, y^* = \max\{y_n^*\}$）可得

  $$\begin{aligned}\theta \max\{x_1, ..., x_n\} + (1-\theta)\max\{y_1, ..., y_n\} &= \max\{\theta x_1, ..., \theta x_n\} + \max\{(1-\theta)y_1, ..., (1-\theta)y_n\} \\ &\geq \max\{\theta x_1 + (1-\theta)y_1, ..., \theta x_n + (1-\theta)y_n\}\end{aligned}$$

  □

- 指数和的对数，可视为最大值函数的可微近似
- 最小二乘，线性分式函数都是凸函数（皆用二阶条件）

8. 下水平集。定义：

$$C_\alpha = \{x \in \mathrm{dom} f | f(x) \leq \alpha\} \tag{19}$$

凸函数的下水平集合仍是凸集

Proof. $x, y \in C_\alpha$, 有 $f(x) \le \alpha$,$f(y) \le \alpha$, 则 $f(\theta x + (1-\theta)y) \le \alpha$，即 $\theta x + (1-\theta)y \in C_\alpha$ □

反之不成立，比如 $-e^x$。对上水平集（凹函数）同理。强拟凸函数的任一下水平集都是凸集。

9. 上境图是 $\mathrm{R}^{n+1}$ 的一个子集。定义：

$$\text{epi}f = \{(x,t)|x \in \text{dom}f, f(x) \le t\} \tag{20}$$

几何视角：法向量为 $(\nabla f(x), -1)$ 的超平面在边界的 $(x, f(x))$ 支撑 epi$f$，其中 $(y,t)$ 是半空间中的点。写成超平面形式。

10. Jensen 不等式。凸函数的定义即为基本不等式。若存在 $k$ 个点满足（凸组合,$\sum \theta_i = 1$），则有：

$$f(\theta_1 x_1 + \ldots + \theta_k x_k) \le \theta_1 f(x_1) + \ldots + \theta_k f(x_k) \tag{21}$$

或有：

$$f(\mathbf{E}x) \le \mathbf{E}f(x) \tag{22}$$

利用对数函数可证明算数-几何平均不等式

(a) 算术-几何平均不等式。真的易证

$$\sqrt{ab} \le (a+b)/2$$

(b) Holder 不等式：对于 $\frac{1}{p} + \frac{1}{q} = 1$ 有

$$\langle x, y \rangle \le \|x\|_p \|y\|_q$$

### 0.2.2 保凸运算

验证保凸运算

1. 定义

2. 一二阶条件

3. 上境图

下述保凸运算的前提都是凸函数

1. 非负加权求和。凸函数的集合本身是一个凸锥。

$$f = \theta_1 f_1 + ... + \theta_n f_n$$

2. 复合仿射映射

$$g(x) = f(Ax + b) \tag{23}$$

3. 逐点最大和逐点上确界

$$f(x) = \max\{f_1(x), f_2(x)\} \tag{24}$$

Proof.

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \max\{f_1(\theta x + (1-\theta)y), f_2(\theta x + (1-\theta)y)\} \\ &\leq \max\{\theta f_1(x) + (1-\theta)f_1(y), \theta f_2(x) + (1-\theta)f_2(y)\} \\ &\leq \theta \max\{f_1(x), f_2(x)\} + (1-\theta)\max\{f_1(y), f_2(y)\}(\text{注意，为什么是} \leq) \\ &= \theta f(x) + (1-\theta)f(y) \end{aligned}$$

□

4. 复合。$f = h \circ g$，若 $f$ 是凸的，$h$ 和 $g$ 必须满足的条件。

标量复合

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

下述即使对 $h$ 的扩充实值函数依然成立

- if $h$ is convex and non-decreasing and $g$ is convex, $f$ is convex
- if $h$ is convex and non-increasing, $g$ is concave, $f$ is convex
- if $h$ is concave and non-decreasing, $g$ is concave, $f$ is concave
- if $h$ is concave and non-increasing, $g$ is convex, $f$ is concave

矢量复合.$g(x)$ is vector

$$f(x) = g'(x)^T \nabla^2 h(g(x))g'(x) + \nabla h(g(x))^T g''(x)$$

就是有四种情况。如果 $g_i$ 是凸函数，$h$ 是凸函数且在每维单调非减，则凸; 如果 $g_i$ 是凸函数，$h$ 是凹函数且在每维单调非减，则凹; 如果 $g_i$ 是凹函数，$h$ 是凸函数且在每维单调非增，则凸; 如果 $g_i$ 是凹函数，$h$ 是凸函数且在每维单调非减，则凹

5. 最小化。$f(x,y)$ 是凸函数。

$$g(x) = \inf_{y\in C} f(x,y) \tag{25}$$

Proof.

$$\begin{aligned} g(\theta x_1 + (1-\theta)x_2) &= \inf f(\theta x_1 + (1-\theta)x_2, y) \\ &\le f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \\ &\le \theta f(x_1,y_1) + (1-\theta) f(x_2,y_2) \\ &\le \theta g(x_1) + (1-\theta) g(x_2) + \epsilon \end{aligned}$$

□

6. 透视函数

$$g(x,t) = tf(x/t) \tag{26}$$

负对数的透视函数称为相对熵，是凸的。

### 0.2.3 共轭函数

1. 共轭函数 $f : \mathrm{R}^n \to \mathrm{R}$，无论是否为凸函数，$f^*$ 是凸函数因为其为逐点上确界

$$f^*(y) = \sup_{x\in \mathrm{dom} f} (\langle y,x\rangle - f(x)) \tag{27}$$

Exampl

2. 仿射函数 $f^*(y) = yx - ax - b$ 有界当且仅当 $y = a$, so dom$f = a$

3. 负对数. $f(x) = -\log x$, 仅在 $x = -1/y$ 有上界

有一些细节，主要是在不同区间需要讨论，从而确定 dom$f^*$. 注意，判断方式是对 $x$ 求导而非

基本性质。

- Fenchel 不等式

$$f(x) + f(y)^* \ge x^T y \tag{28}$$

因为共轭函数定义在实数域上，因此 $y^T x = x^T y = \langle x,y\rangle$.

- 共轭的共轭
$$f^{**}(x) = \sup_{y \in \mathrm{R}^n} \{y^T x - f^*(y)\} \tag{29}$$
由 Fenchel 不等式可得,
$$f(x) \geq x^T y - f^*(y) = f^{**}(x)$$
若 $f(x)$ 是真的闭凸函数，那么 $f(x) = f^{**}(x)$。证明如下（为自己而已）:

Proof. 取向量 $(x, \gamma) \in \mathrm{epi} f^{**}$，$(y, \omega) \in \mathrm{epi} f^*$ 且 $(x, \gamma) \notin \mathrm{epi} f$，则对于 $\mathrm{epi} f$ 的支持一超平面，其法向量为 $(\alpha, \beta)$（因为对应于上境图，其中 $\alpha, x$ 为 n 维实数空间中的向量）有:
$$(\alpha, \beta)^T (y, \omega) < b < (\alpha, \beta)^T (x, \gamma)$$
即
$$\alpha^T y + \beta\omega < \alpha^T x + \beta\gamma$$
由于 $\omega, \gamma$ 为标量，可以取任意值，而由 $f > f^{**}$ 可得存在 $\omega > \gamma$。因此取 $\beta$ 为负（例如-1）时上式成立:
$$\alpha^T y - \omega < \alpha^T x - \gamma$$
即
$$\alpha^T y - f(y) < \alpha^T x - f^{**}(x)$$
$$f^*(\alpha) < \alpha^T x - f^{**}(x)$$
由于共轭的共轭满足 Fenchel 不等式，即
$$f^*(y) + f^{**}(x) > y^T x$$
因此上式与此矛盾, 不存在这样的向量 $(x, \gamma)$, 也即 $(x, \gamma) \in \mathrm{epi} f$, 则有 $f(x) \leq f^{**}(x)$, 又由 Fenchel 不等式可得，$f(x) = f^{**}(x)$ □

- 伸缩变换和复合仿射变换. $g(x) = af(x) + b$，则有:
$$g^*(y) = af^*(y/a) - b \tag{30}$$

Proof.

$$\begin{aligned}g^*(x) &= \sup\{y^Tx - g(x)\}\\ &= \sup\{y^Tx - af(x) - b\}\\ &= \sup\{y^Tx - af(x)\} - b\\ &= a\sup\{\frac{y^T}{a}x - f(x)\} - b\\ &= af^*(y/a) - b\end{aligned}$$

□

$g(x) = f(Ax+b)$，则有:

$$g^*(y) = f^*(A^{-T}y) - b^TA^{-T}y \tag{31}$$

Proof.

$$\begin{aligned}g^*(x) &= \sup\{y^Tx - g(x)\}\\ &= \sup\{y^Tx - f(Ax+b)\}\end{aligned}$$

令 $Ax+b=u$，则 $x = A^{-1}(u-b)$，有:

$$\begin{aligned}g^*(x) &= \sup\{y^TA^{-1}(u-b) - f(u)\}\\ &= \sup\{y^TA^{-1}u - f(u)\} - y^TA^{-1}b & &= f^*(A^{-T}y) - y^TA^{-1}b\end{aligned}$$

□

- 独立函数的和

$$f^*(w,z) = f_1^*(w) + f_2^*(z) \tag{32}$$

### 0.2.4 拟凸函数

1. 定义与例子. 定义域及其所有下水平集

$$S_\alpha = \{x \in \mathbf{dom} f | f(x) \le \alpha\} \tag{33}$$

都是凸集。如果函数拟凸又拟凹，那么是拟线性函数，即 $\{x|f(x)=\alpha\}$ 和定义域是凸集。

对数函数拟线性，$f(x_1,x_2) = x_1x_2$（定义域正实数）是拟凹函数，线性分式函数拟线性函数，上取整函数是拟凸函数。

2. 基本性质. 充要条件

$$f(\theta x + (1-\theta)y) \leq \max\{f(x), f(y)\} \tag{34}$$

3. 如果函数是拟凸的，等价于满足其中之一 (i) 非减 (ii) 非增 (iii) 全局最小点左右恒非减或非增

### 0.2.5 对数凸与对数凹

1. 定义.$\log f$ 是凸或凹的，则是对数凸或对数凹的。如果对数凹的，有：

$$f(\theta x + (1-\theta)y) \geq f(x)^{\theta} f(y)^{1-\theta} \tag{35}$$

2. 仿射函数是对数凹函数，指数函数既是对数凸函数也是对数凹函数

## 0.3 凸优化问题

### 0.3.1 优化问题

#### 基本术语

方程组为约束，相应的函数为约束函数。定义域为：

$$\mathcal{D} = \bigcap^{m} \mathrm{dom} f_i \cap \bigcap^{p} \mathrm{dom} h_i$$

可行点的集合为可行集。$p^* = \inf\{f_0(x)\}$ 若 $p^* = \infty$ 则不可行，若 $p^* = -\infty$ 则无下界。若 $f_0(x^*) = p^*$ 则 $x^*$ 为最优点，存在则最优值可达，不存在则不可达（常为无下界时）。满足 $f_0(x) \leq p^* + \epsilon$ 的可行解为 $\epsilon$ 次优。局部最优解：

$$f_0(x) = \inf\{f_0(z) | \|z - x\|_2 \leq R\}$$

局部最优是优化问题的解

#### 等价问题

此处有诸多疑难，建议先跳过

1. 变量变换是等价问题。若 $\phi : z \to x$ 是双射，则 $f_0(z)$ 由变量代换得到。即为 $\tilde{f}_i(z) = f_i(\phi(z))$
2. 目标函数和约束函数的变换。此处 $\phi : \mathrm{R} \to \mathrm{R} \tilde{f}_i(x) = \phi_i(f_i(x))$
3. 松弛变量。
4. 消除等式约束。所谓消除等式约束，就是定义这样一个映射 $\phi : z \to x$ 使得 $\{0\} \subset \mathrm{null}\phi$。以线性等式约束 $Ax = b$ 就是将通解（如果存在）代入优化问题：$x = Fz + x_0$
5. 引入等式约束。引入 $y = Ax + b$
6. 优化部分变量。由 $\inf_{x,y} f(x, y) = \inf_x \tilde{f}(x)$
7. 上境图

$$\begin{aligned} \min \quad & t \\ \text{s.t} \quad & f_0(x) \leq t \end{aligned}$$

8. 隐式与显式约束。拿衣舞。

### 0.3.2 凸优化

**标准形式**

$$\begin{aligned} &\min f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0 \\ & a_i^T x = b_i \end{aligned}$$

$f_i$ 必须为凸函数，即满足：1）目标函数是凸的；2）不等式约束函数是凸的；3）等式约束函数是仿射的。因此可行集是凸的，由 m 个下水平集和 p 个超平面的交集组成。

上述为凸优化问题的标准形式，实际情况下，许多不是标准形式。

### 0.3.3 局部最优与全局最优

凸优化问题的局部最优同时也是全局最优。这个理论乍一看很显然，但实际上也很显然，用分析的语言来描述。

$$f_o(x^*) = \inf\{f_0(x) || x - x^*| < \epsilon\}$$

如果 $x^*$ 仅为局部最优而非全局最优的，那么存在一个点 $y$ 为最优的。由凸函数的性质可得，

$$x = (1-\theta)x^* + \theta y$$

$$f(x) \le (1-\theta) f(x^*) + \theta f(y)$$

取 $\theta = \frac{\epsilon}{2|y-x^*|}$，将其代入可得：

$$\begin{aligned} x =& x^* + \theta(y - x^*) \\ =& x^* + \frac{\epsilon}{2} \end{aligned}$$

故 $x$ 又在局部中，和凸性矛盾。证毕。显然这个在度量空间也成立，除非凸性在度量空间不成立。

**可微函数的最优性准则**

如果 $x$ 是最优解，当且仅当：

$$\nabla f_0(x)^T (y - x) \ge 0 \quad \forall x, y \in X \tag{36}$$

Think of 凌青。注意证明同样采用了上面的反正-构造方法。

1. 无约束问题的最优解充要条件为：
$$\nabla f_0(x) = 0$$
思考一下无约束二次规划解的三种情况（其实就是可逆性）

2. 只含等式约束的问题。这个比较复杂，当时提及了实向量空间上的伴随算子的矩阵等价于算子的转置，因此这个证明是容易的。不过还要再看，其他部分的论述很不清晰。建议学习凌青的证明。$f_0(x)$ 的梯度方向与 $Ax+b$ 垂直，也即 $Ax+b$ 与 $f_0(x)$ 相切
$$\nabla f_0(x)^T \perp Av$$
或
$$\nabla f_0(x) + A^T v = 0$$

3. 非负象限中的极小化。由 $x \succeq 0, \forall y \in X$ 与最优性准则可得 $-\nabla f_0(x)^T x \geq 0$。结论就是
$$\sum \nabla f_0(x)_i x_i = 0$$
实际上是互补性。

**等价的凸问题**

1. 消除等式约束。这个要看一下板书
2. 引入等式约束。比如原 $x = Ay + b$ 辄加入此约束
3. 松弛变量。$f_i$ 必须是仿射的。
4. 上境图形式。新的目标函数 $t$ 和约束函数 $f_0(x) - t$
5. 极小化部分变量

### 0.3.4 线性规划问题

目标函数和约束函数都必须是仿射的。具有如下形式：
$$\begin{aligned} \min \quad & c^T x + d \\ s.t \quad & Gx \preceq h \\ & Ax = b \end{aligned}$$

可行集是多面体。

1. 标准形式

$$\begin{aligned} \min \quad & c^T x \\ s.t \quad & Ax = b \\ & x \succeq 0 \end{aligned}$$

2. 没有等式约束，辄为不等式形式线性规划

$$\begin{aligned} \min \quad & c^T x \\ s.t \quad & Ax \preceq b \end{aligned}$$

3. 转化为标准形，即加入松弛变量，都是等式约束。后面的例子太多，等十六周和做题遇到才看。

4. 清楚分片线性极小化采用上境图转化。分片线性极小化：

$$f(x) = \max_{i=1\ldots m} (a_i^T xx + b_i)$$

构造为上境图为：

$$\begin{aligned} \min \quad & t \\ \text{s.t} \quad & a_i^T x + b_i \leq t \end{aligned}$$

### 0.3.5 二次优化问题

**二次规划**

$$\begin{aligned} \min \quad & (1/2)x^T P x + q^T x + r \\ s.t \quad & Gx \preceq h \\ & Ax = b \end{aligned}$$

注意到 $P$ 是对称半正定的矩阵。上述为 QP。如果不等式约束也是凸二次型，辄称为 QCQP。

**典型例子**

1. 最小二乘及回归。约束为 $l_i \leq x_i \leq u_i$

2. 多面体距离。$\|x_1 - x_2\|_2^2$ 若最优解为 0 则相交，无可性解多面体为空

### 二阶锥规划

此类问题为 SOCP

$$
\begin{aligned}
\min \quad & f^T x \\
s.t \quad & \|A_i x + b_i\|_2 \preceq c_i^T x + d_i \\
& Fx = g
\end{aligned}
$$

显然，当 $c_i = 0$ 是等同于 QCQP。

1. 鲁棒线性规划。约束条件

$$\sup\{a_i^T x | a_i \in \epsilon_i\} = \bar{a}_i^T x + \|P_i^T x\|_2 \leq b_i$$

# 0.4 对偶

## 0.4.1 Lagrange 函数的定义

优化问题是标准形式的，但不要求是凸的。Lagrange 函数用于将原始优化问题的目标函数与约束结合，形式如下：

$$L(x,\lambda,\nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \nu_j h_j(x),$$

其中：

- $f(x)$ 是目标函数；
- $g_i(x) \le 0$ 是不等式约束；
- $h_j(x) = 0$ 是等式约束；
- $\lambda_i \ge 0$ 是与不等式约束相关的拉格朗日乘子；
- $\nu_j \in \mathbb{R}$ 是与等式约束相关的拉格朗日乘子。

## 0.4.2 Lagrange 对偶函数的定义

对偶函数是通过最小化 Lagrange 函数 $L(x,\lambda,\nu)$（对 $x$ 变量）而得：

$$g(\lambda,\nu) = \inf_x L(x,\lambda,\nu).$$

其中：

- $g(\lambda,\nu)$ 是拉格朗日对偶函数；
- $\inf_x$ 表示在所有 $x$ 上取下确界。
- 称 $\lambda \succeq 0, (\lambda, v) \in \text{dom} g$ 的 $(\lambda, v)$ 是对偶可行的

## 0.4.3 性质

### (1) 对偶性 (Weak Duality)

对于任何 $\lambda \ge 0$ 和任意 $\nu$，都有：

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

其中 $f^*$ 是原始优化问题的最优值。即有：

$$\begin{aligned} g(\lambda, v) &= \inf(f_0(x) + \lambda^T(Ax - b) + v^T(Cx - d)) \\ &= -b^T\lambda - d^T v - \inf(f_0(x) + (A^T\lambda + C^T v)^T x) \\ &= -b^T\lambda - d^T v - f_0^*(-(A^T\lambda + C^T v)^T x) \end{aligned}$$

注意为负是因为虽然 $\langle y, x\rangle$ 是正的，但是 $f_0(x)$ 前是负的。试求熵的最大化问题

**强对偶性** (Strong Duality)

若原始优化问题满足一定条件（如凸优化问题满足 Slater 条件），则有：

$$\sup_{\lambda \geq 0, \nu} g(\lambda, \nu) = f^*.$$

**凸性**

Lagrange 对偶函数 $g(\lambda, \nu)$ 是关于 $\lambda$ 和 $\nu$ 的凹函数，即使原始问题不是凸优化问题。

### 0.4.4 Lagrange 对偶问题

**定义：** 对偶问题是指通过最大化对偶函数 $g(\lambda, \nu)$ 得到：

$$\sup_{\lambda \geq 0, \nu} g(\lambda, \nu).$$

**性质：**

- 对偶问题是一个凸优化问题，即使原始问题不是凸的。
- 解对偶问题时，需满足 KKT 条件以保证解的可行性和最优性。
- 不等式和等式约束的线性规划的对偶函数是等价的

**弱对偶性：**

- $d^* \leq p^*$ 恒成立，无论原问题是否为凸问题，也无论 $d^*, p^*$ 是否无限
- 原问题无下界，对偶问题不可行；对偶问题无上界，原问题不可行
- 对偶间隙:$p^* - d^*$

### 0.4.5 Slater 条件与强对偶性

#### 强对偶性

对偶间隙为零则成立。若原问题是凸问题，那么强对偶性通常成立

#### Slater 条件

Slater 条件是凸优化中特殊的一种可行性条件，描述优化问题的强对偶性是否成立。对于问题：

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} \ f(x) \\ &\text{s.t. } g_i(x) \le 0,\ i = 1, 2, \ldots, m, \\ &\quad\ \ h_j(x) = 0,\ j = 1, 2, \ldots, p, \end{aligned}$$

当目标函数 $f(x)$ 和所有约束 $g_i(x)$ 是凸函数，等式约束 $h_j(x)$ 是仿射函数时，Slater 条件要求存在一个严格可行解 $x$，使得：

$$g_i(x) < 0,\ i = 1, 2, \ldots, m, \quad h_j(x) = 0,\ j = 1, 2, \ldots, p.$$

#### 1. 定义

互补松弛性是 KKT 条件中的一个重要部分，描述了不等式约束的松弛与其拉格朗日乘子的关系。具体地：

$$\lambda_i g_i(x) = 0, \quad \forall i = 1, \ldots, m,$$

其中：

- $\lambda_i \ge 0$ 是不等式约束 $g_i(x) \le 0$ 的拉格朗日乘子；
- $g_i(x) \le 0$ 是约束条件。

### 0.4.6 互补松弛性

- **必要性：** 若 $x^*$ 是原问题的最优解，且强对偶性成立，则互补松弛性是必要的。
- **直观意义：** 互补松弛性表示，当一个不等式约束是活跃的（即 $g_i(x^*) = 0$）时，其对应的拉格朗日乘子 $\lambda_i$ 可以是正值；当约束不活跃时（即 $g_i(x^*) < 0$），则 $\lambda_i = 0$。
- **在凸优化中的充分性：** 对于凸优化问题，互补松弛性与其他 KKT 条件一起是充分条件。

### 0.4.7 KKT 条件的定义

对于一个带约束的优化问题：

$$\begin{aligned}
&\min_{x\in\mathbb{R}^n}\ f(x)\\
&\text{s.t. } g_i(x)\le 0,\ i=1,2,\ldots,m,\\
&\quad\ \ h_j(x)=0,\ j=1,2,\ldots,p,
\end{aligned}$$

如果 $f(x)$ 和所有 $g_i(x)$ 是可微函数，KKT 条件包括以下几个部分：

1. **可行性条件：**

$$g_i(x^*)\le 0,\ i=1,2,\ldots,m,\quad h_j(x^*)=0,\ j=1,2,\ldots,p.$$

2. **拉格朗日条件：** 存在拉格朗日乘子 $\lambda_i\ge 0$ 和 $\nu_j$，使得：

$$\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i\nabla g_i(x^*)+\sum_{j=1}^{p}\nu_j\nabla h_j(x^*)=0.$$

3. **互补松弛条件：**

$$\lambda_i g_i(x^*)=0,\ i=1,2,\ldots,m.$$

4. **非负性条件：**

$$\lambda_i\ge 0,\ i=1,2,\ldots,m.$$

### 0.4.8 KKT 条件与 Slater 条件的区别

1. **适用范围：**
   - KKT 条件适用于一般优化问题，包括非凸优化问题。
   - Slater 条件主要用于凸优化问题。
2. **目标函数要求：**
   - KKT 条件不要求目标函数必须是凸函数。
   - Slater 条件要求目标函数和不等式约束函数是凸函数。
3. **对偶性关系：**
   - KKT 条件是最优性条件，与目标函数的凸性无关。

- Slater 条件保证强对偶性，即原问题的最优值等于对偶问题的最优值。

4. **可行性要求：**

    - KKT 条件要求只需满足一般可行性条件。任意满足 KKT 条件的点分别是原问题、对偶问题的最优解
    - Slater 条件要求存在一个严格可行解。

# 0.5 优化方法

## 0.5.1 无约束优化问题

### 求解方法

1. 首先要是二次可微凸函数，因此最优点满足：

$$\nabla f(x^*) = 0$$

2. 极小化点列

$$f(x^{(k)}) \to P^*$$

当 $f(x^{(k)} - p^* \leq \epsilon)$ 时停止

### 例子

1. 二次凸优化

$$x^* = -P^{-1}q$$

如果无解则优化问题无下界。当是最小二乘时，有 $A^TAx^* = A^Tb$

2. 无约束几何规划形如 $f(x) = \log(\sum \exp(a^Tx_i + b_i))$ 没有解析解，只能采用迭代算法。

## 0.5.2 下降方法

优化点列：

$$x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x^{(k}$$

$\Delta x$ 为搜索方向。$f(x^{(k)})$ 给出下水平集 $S$，因此屡次迭代在缩小这个集合，或皆处于 $S$ 中。搜索方向必须满足：

$$\langle \nabla f(x^{(k)})^T, \Delta x^{(k)} \rangle < 0$$

和梯度方向的夹角必须是钝角，此类方向为下降方向。

### 通用下降方法

通用下降方法。$t$ 决定从直线 $x + t\Delta x$ 上某点开始下一步迭代，停止时 $\|\nabla f(x)\|_2 \leq \epsilon$

### 0.5.3 精确直线搜索 (Exact Line Search)

**定义：** 在优化中，精确直线搜索是指沿给定方向 $d_k$ 寻找一个步长 $\alpha_k$，使得：

$$\alpha_k = \arg\min_{\alpha>0} f(x_k + \alpha d_k).$$

**流程：**

1. 确定当前点 $x_k$ 和搜索方向 $d_k$。

2. 求解 $\alpha_k$ 满足上述优化问题。

3. 更新 $x_{k+1} = x_k + \alpha_k d_k$。

### 0.5.4 回溯直线搜索 (Backtracking Line Search)

**定义：** 回溯直线搜索是一种启发式方法，通过逐步减少步长 $\alpha$，找到一个满足 Wolfe 条件的步长：

$$f(x_k + \alpha d_k) \leq f(x_k) + c_1 \alpha \nabla f(x_k)^\top d_k,$$

其中 $c_1 \in (0, 1)$ 是常数。**流程：**

1. 初始化 $\alpha = \alpha_0$ 和缩放因子 $\beta \in (0, 1)$。

2. 逐步减小 $\alpha$: $\alpha \leftarrow \beta\alpha$，直到满足 Wolfe 条件。

3. 更新 $x_{k+1} = x_k + \alpha d_k$。

### 0.5.5 梯度下降法 (Gradient Descent)

**定义：** 梯度下降法使用负梯度方向作为搜索方向 $d_k = -\nabla f(x_k)$，通过调整步长 $\alpha_k$ 进行迭代。**流程：**

1. 确定当前点 $x_k$。

2. 计算搜索方向 $d_k = -\nabla f(x_k)$。

3. 使用直线搜索方法确定步长 $\alpha_k$。

4. 更新 $x_{k+1} = x_k + \alpha_k d_k$。

### 0.5.6 最速下降法 (Steepest Descent)

**定义：** 最速下降法是梯度下降法的特例，步长 $\alpha_k$ 是通过精确直线搜索确定的。

$$\Delta x_{nsd} = \text{argmin}\{\nabla f(x)^T v | \|v\| \leq 1\}$$

**流程：** 与梯度下降法类似，只是步长 $\alpha_k$ 采用精确直线搜索。

### 0.5.7 Newton 步径 (Newton Step)

**定义：** Newton 步径是基于二阶泰勒展开，利用 Hessian 矩阵 $\nabla^2 f(x_k)$ 和梯度 $\nabla f(x_k)$ 定义的搜索方向：

$$d_k = -\left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k).$$

如果函数 $f$ 是二次的。那么 $x + \Delta x_{nt}$ 是精确最优解。**流程：**

1. 确定当前点 $x_k$。
2. 计算 Hessian 矩阵 $\nabla^2 f(x_k)$ 和梯度 $\nabla f(x_k)$。
3. 求解方向 $d_k = -\left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k)$。
4. 使用直线搜索方法确定步长 $\alpha_k$。
5. 更新 $x_{k+1} = x_k + \alpha_k d_k$。
6. 计算 Newton 减量
$$\lambda(x) = (\Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt})^{1/2}$$

### 0.5.8 Newton 方法 (Newton's Method)

**定义：** Newton 方法是一种二阶优化方法，采用 Newton 步径作为更新方向，并通常使用固定步长 $\alpha_k = 1$。又称为阻尼 Newton 方法或者谨慎 Newton 方法 **流程：**

1. 确定当前点 $x_k$。
2. 计算 Hessian 矩阵 $\nabla^2 f(x_k)$ 和梯度 $\nabla f(x_k)$。
3. 求解更新公式：
$$x_{k+1} = x_k - \left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k).$$

### 0.5.9 方法之间的区别

1. **方向计算：**
   - 梯度下降法和最速下降法使用负梯度方向。
   - Newton 方法使用二阶信息构造的方向。
2. **步长选择：**
   - 精确直线搜索求解最优步长，回溯直线搜索通过试探找到满足条件的步长。
   - 最速下降法采用精确直线搜索，而梯度下降法可以结合回溯直线搜索。
   - Newton 方法通常使用固定步长 $\alpha_k = 1$。
3. **收敛速度：**
   - 梯度下降法和最速下降法为一阶方法，收敛较慢。
   - Newton 方法为二阶方法，收敛速度快（在问题条件适当时）。
4. **计算复杂度：**
   - Newton 方法需要计算 Hessian 矩阵，计算复杂度较高。
   - 梯度下降法和最速下降法只需计算梯度，计算复杂度较低。

## 0.6 注意事项

### 0.6.1 纰漏

1. 一个既凸又凹的函数在进行非负加权和时可以作为既凸又凹的；若凸函数为负后是凹的，那么视之为凹
2. 判断矩阵正定/负定往往有两种简单的方法：主子式与特征值
3. $t^\alpha$ 往往是凹函数若有 $\alpha < 1$
4. 欧几里得球与椭球中 $P^{1/2} = A$
5. 常见的对数凹-凸函数：指数函数（对数凸与凹），仿射函数（对数凹函数）
6. 注意上境图与下水平集为 $f(x) \leq t$ 与 $f(x) \leq \alpha$。函数为凸当且仅当上境图是凸集。函数为拟凸若下水平集是凸集，但不一定凸。

7. 如果 $f_i(x, y)$ 关于 $x$ 是凸函数，那么其 max,sup,inf（$y \in A$）也是凸函数，但没有对 min 声明，故对分片线性函数 $a_i^T x$ 的最小不一定保凸

8. 互补松弛条件仅在强对偶性成立时成立，同理 KKT 条件。这两者是必须成立。而 slater 条件是用于确保强对偶性成立的（前提是原问题是凸问题）。如果原问题是凸问题，那么满足 KKT 的点同时也是原问题的最优解

9. 关于凸函数的解答题，首要说明凸函数的定义域是凸的

10. 关于凸优化中，要声明下标从几到几

11. 对偶函数是凹函数，对偶问题是凸优化问题

12. $\Delta x$ 不是某个具体的 $\|x^{k+1} - x^k\|$，而代表一个搜索方向，比如可能为 $-\nabla f(x)$

### 0.6.2 例子概述

1. 射线 $\{x + \theta v | \theta \geq 0\}$ 不是仿射的，除非经过原点

2. 线性矩阵不等式和有序线性不等式（$a^T x = x_1 a_1 + x_2 a_2 + ... + x_n a_n$）

3. 线性分式函数由透视函数和仿射函数复合而来，其中 $g(x) = [A, c^T]^T x + [b, d]^T$ 是仿射函数，自然 $P$ 是透视函数

4. 正常锥（广义不等式）的简易例子：非负象限、半正定锥

5. 对偶锥的例子：对偶锥是其正交补；非负象限；半正定锥；范数的对偶

6. 逐点最大的例子：分片线性 $\max_i(a_i^T x + b_i)$

7. 共轭函数（我的亲娘），搞定了

8. 拟凸函数例子:$\log x$，上取整 $\inf\{z | z \geq x\}$ 是拟线性，$x_1 x_2$ 是拟凹函数

9. 最优性条件例子：无约束 $\nabla f_0(x)^T$, 只含等式约束 $\nabla f_0(x) + A^T v = 0$, 非负象限中的极小化 $x_i(\nabla f_0(x))_i = 0$

| **条件** | **凸问题要求** | **保证的性质** | **说明** |
|---|---|---|---|
| KKT 条件 | 否 | 局部最优解的必要条件（一般问题） | 原问题为凸时，KKT 是全局最优的充要条件。 |
| Slater 条件 | 是 | 强对偶性，KKT 充分性 | 是凸问题强对偶性的充分条件。 |
| 互补松弛条件 | 否 | 局部最优解的必要条件（一般问题） | 原问题为凸时，是全局最优的充要条件之一。 |
| 强对偶性 | 是 | $p^* = d^*$ | Slater 条件保证凸问题中强对偶性成立。 |

Table 1: KKT、Slater、互补松弛、强对偶性之间的关系

# 常见矩阵求导规则

## 一、x 为向量的常见规则

1. **标量对向量求导：** 若 $f = \mathrm{x}^T\mathrm{A}\mathrm{x}$，其中 A 是对称矩阵，则
$$\frac{\partial f}{\partial \mathrm{x}} = 2\mathrm{A}\mathrm{x}.$$
2. **标量对向量求导（线性形式）：** 若 $f = \mathrm{b}^T\mathrm{x}$，其中 b 是常向量，则
$$\frac{\partial f}{\partial \mathrm{x}} = \mathrm{b}.$$
3. **标量对向量求导（内积形式）：** 若 $f = \|\mathrm{x}\|_2^2 = \mathrm{x}^T\mathrm{x}$，则
$$\frac{\partial f}{\partial \mathrm{x}} = 2\mathrm{x}.$$

## 二、X 为矩阵的常见规则

1. **矩阵对矩阵求导：** 若 $f = \mathrm{tr}(\mathrm{X}^T\mathrm{A})$，其中 A 是常矩阵，则
$$\frac{\partial f}{\partial \mathrm{X}} = \mathrm{A}.$$
2. **矩阵的 Frobenius 范数：** 若 $f = \|\mathrm{X}\|_F^2 = \mathrm{tr}(\mathrm{X}^T\mathrm{X})$，则
$$\frac{\partial f}{\partial \mathrm{X}} = 2\mathrm{X}.$$
3. **矩阵对矩阵的乘积求导：** 若 $f = \mathrm{XY}$，其中 X 和 Y 是矩阵，则
$$\frac{\partial (\mathrm{XY})}{\partial \mathrm{X}} = \mathrm{Y}^T.$$
4. **矩阵对行列式求导：** 若 $f = \det(\mathrm{X})$，其中 X 是方阵，则
$$\frac{\partial \det(\mathrm{X})}{\partial \mathrm{X}} = \mathrm{X}^{-1}\mathrm{X}^T.$$