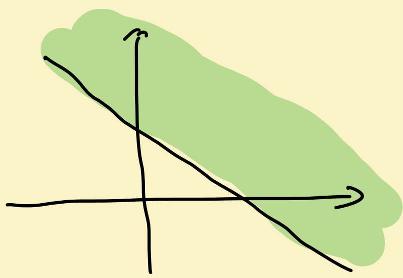
范数　向量 l_p 范数 $\|v\|_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{\frac{1}{p}}$

矩阵 $\|A_1\| = \sum_{ij} |A_{ij}|$　　$\|A\|_F = \sqrt{\sum_{ij} A^2} = \sqrt{tr(AA^T)}$

矩阵诱导 $\|A\|_{(m,n)} = \max\limits_{x\in\mathbb{R}^n, \|x\|_n=1} \|Ax\|_m$

凸集　$\forall\, x,y \in C$　$\forall\, \theta \in [0,1]$　$\theta x + (1-\theta) y \in C$



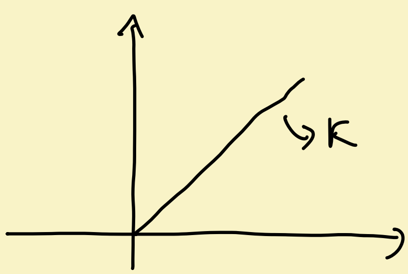
多面体：$\{x \mid Ax \le b,\ Cx = d\}$ 是一个凸集，因为是多个凸集的交

$x_1, x_2 \in \{x \mid a^T x \le b\}$ 半空间是凸集

Proof: $\theta a^T x_1 + (1-\theta) a^T x_2 \le b$

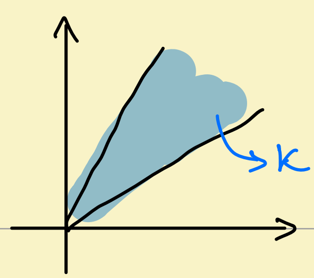
$x_1, x_2 \in \{x \mid c^T x = d\}$ 是一个仿射集，所以是凸集

E.g.:
$$x_1 + 2x_2 \le 5,\quad x_1 \ge 0,\quad x_2 \ge 0 \Rightarrow \underbrace{\begin{bmatrix} 1 & 2 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} \le \underbrace{\begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix}}_{b}$$



锥 cone

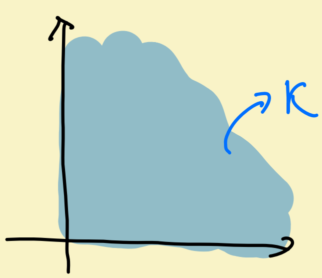
Defn: 一个集合 $K \subseteq \mathbb{R}^n$ 是一个锥，如果 $x \in K \Rightarrow sx \in K, \forall s > 0$



凸锥: convex cone

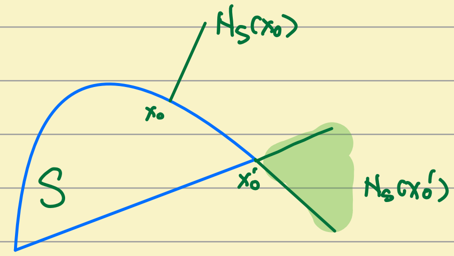
Defn: 一个集合 $K \in \mathbb{R}$ 是一个凸锥，如果 $x, y \in K$

$\Rightarrow \alpha x + \beta y \in K \quad \forall \alpha, \beta > 0$



Thm: 一个锥 K 是一个凸锥，当且仅当 $K + K \subseteq K$.

E.g.: 法锥 Normal cone : $N_S(x_0) = \{y \in \mathbb{R}^n \mid \langle y, x - x_0 \rangle \le 0, \forall x \in S\}$



适当锥 proper cone

Defn : 一个锥 $K \subseteq \mathbb{R}^n$ 是适当锥，当其满足

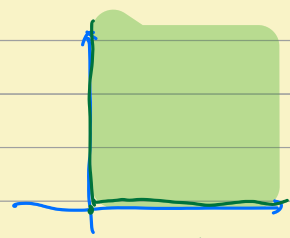
$x, y \in K$

$ax + by \in K \Rightarrow \begin{cases} \lambda x \in K & \lambda > 0 \\ x + y \in K \end{cases}$

① K 是闭集

② K 是实心的，$\mathrm{int}(K) \neq \emptyset$

③ K 是尖的，$x \in K$, $-x \in K$，则 $x = 0$



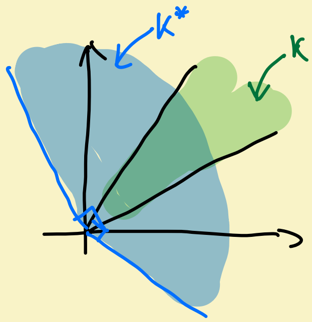
R^n_+

$K = \{X \in \mathbb{R}^{n \times n} \mid X$ 是半正定矩阵$\}$

$\forall y \in \mathbb{R}^n, \quad y^T X y \ge 0$

$y^T X y \ge 0 \Rightarrow \lambda y^T X y \ge 0 \Rightarrow y^T (\lambda X) y \ge 0$

$y^T X y \ge 0, \ y^T Z y \ge 0 \Rightarrow y^T (X + Z) y \ge 0$



对偶锥

Defn: 若 $K \in \mathbb{R}^n$ 是一个锥，则 K 的对偶锥为

$$K^* = \{y \in \mathbb{R}^n \mid \langle x, y\rangle \geq 0,\ \forall x \in K\}$$

E.x. ① $K = \mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x \geq 0\}$
② $K = \{(x,t) \in \mathbb{R}^{n+1} \mid \|x\|_2 \leq t,\ t \geq 0\}$
} 自对偶

① $K^* = \{y \in \mathbb{R}^n \mid \langle x, y\rangle \geq 0,\ \forall x \geq 0\}$

$x^T y = \sum_{i=1}^{n} x_i y_i \geq 0$, and $x \geq 0 \Rightarrow y \geq 0$

$\therefore K^* = \{y \in \mathbb{R}^n \mid y \geq 0\}$

② $K^* = \left\{(y,s) \in \mathbb{R}^{n+1} \mid \left\langle \begin{pmatrix} x \\ t \end{pmatrix}, \begin{pmatrix} y \\ s \end{pmatrix} \right\rangle \geq 0,\ \forall \begin{pmatrix} x \\ t \end{pmatrix} \in K\right\}$

$\left\langle \begin{pmatrix} x \\ t \end{pmatrix}, \begin{pmatrix} y \\ s \end{pmatrix} \right\rangle \geq 0 \Rightarrow x^T y + t \cdot s \geq 0,\ \forall \|x\|_2 \leq t$

Setting $x = -y,\ t = \|y\| \Rightarrow -\|y\|_2^2 + s\|y\|_2 \geq 0$.

$\Rightarrow \|y\|_2 \leq s$

$\therefore K^* = \{(y, s) \in \mathbb{R}^{n+1} \mid \|y\|_2 \leq s\}$

函数的几个概念

$\forall a \in \mathbb{R}\quad -\infty < a < +\infty$

$(+\infty) + (+\infty) = +\infty$

$+\infty + a = +\infty$

① **实值函数**. 映射 $f: \mathbb{R}^n \to \mathbb{R}$ 实值函数.

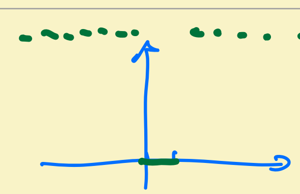
② **广义实值函数**, 映射 $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} := \overline{\mathbb{R}}$

③ **适当函数**

Defn: 给定广义实值函数 f 和非空集合 $\mathcal{X}$，如果 $\exists x \in \mathcal{X}$ 使得 $f(x) < +\infty$ 且 $\forall x \in \mathcal{X}$，都有 $f(x) > -\infty$，那么称 **f 是 $\mathcal{X}$ 上的适当函数**

E.g. $f(x) = \begin{cases} \ln(x) & x > 0 \\ -\infty & x = 0 \end{cases}$ $\qquad I_\Omega(x) = \begin{cases} 0 & x \in \Omega \\ +\infty & \text{otherwise} \end{cases}$

$f(x) = \begin{cases} -\ln(x) & x > 0 \\ +\infty & x = 0 \end{cases}$ $\qquad \Omega = [0,1]$

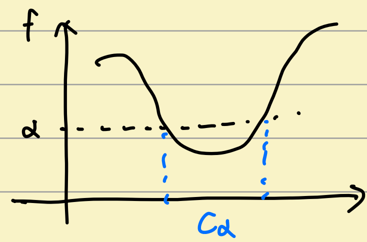


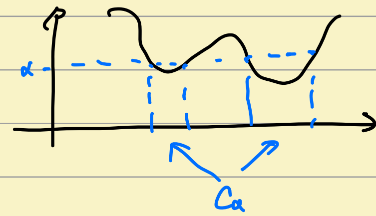
④ 闭函数

Defn: 设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ 为广义实值函数，若 epi f 为闭集，则称f为闭函数.

Defn (下水平集) $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ $\alpha\in\mathbb{R}$，f的α-下水平集为

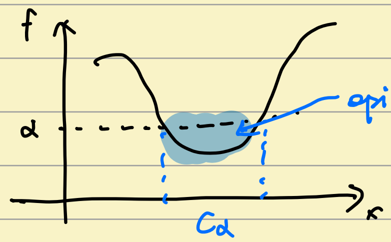
$$C_\alpha=\{x\mid f(x)\le\alpha\}$$





Defn (上图) $f:\mathbb{R}^n\to\overline{\mathbb{R}}$，f的上图为　　epigraph 上图

$$\text{epi}\, f=\{(x,\alpha)\in\mathbb{R}^{n+1}\mid f(x)\le\alpha\}.$$

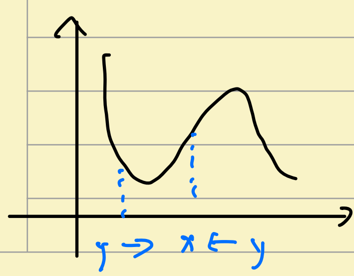


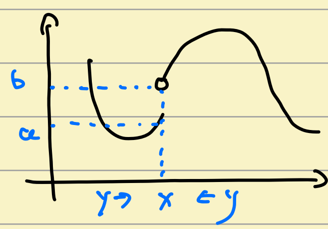
⑤ 下半连续函数

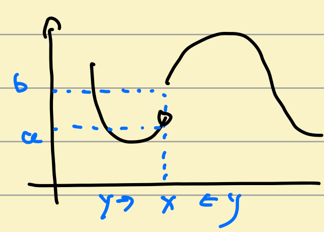
Defn: 设 $f:\mathbb{R}^n\to\overline{\mathbb{R}}$，若对于任意的 $x\in\mathbb{R}^n$，有

$$\liminf_{y\to x} f(y)\ge f(x)$$

则称 f(x) 为下半连续函数.







$\liminf_{y\to x} f(y)=a$

$f(x)=a$

$\liminf_{y\to x} f(y)=a$

$f(x)=b$

Thm (下半连续函数与闭函数) 设 $f: \mathbb{R}^n \to \overline{\mathbb{R}}$，以下的命题等价.

① $f(x)$ 的任意 α-下水平集是闭集

② $f(x)$ 是下半连续的

③ $f(x)$ 是闭函数.

导数.

向量函数梯度

Defn (梯度) 给定 $f: \mathbb{R}^n \to \mathbb{R}$，且 f 在 x 的邻域内有定义. 若存在向量 $g \in \mathbb{R}^n$ 满足

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - g^T h}{\|h\|_2} = 0$$

nabla

则称 f 在 x 处可微. 称 g 是 f 在 x 处的梯度. 记 $\nabla f(x)$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} \qquad \nabla^2 f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1}\frac{\partial}{\partial x_1} & \frac{\partial f(x)}{\partial x_1}\frac{\partial}{\partial x_2} & \cdots & \frac{\partial f(x)}{\partial x_1}\frac{\partial}{\partial x_n} \\ & \vdots & & \\ \frac{\partial f(x)}{\partial x_n}\frac{\partial}{\partial x_1} & \frac{\partial f(x)}{\partial x_n}\frac{\partial}{\partial x_2} & \cdots & \frac{\partial f(x)}{\partial x_n}\frac{\partial}{\partial x_n} \end{bmatrix}$$

$x \in \mathbb{R}^n$ $\qquad \nabla f(x) \in \mathbb{R}^n$ $\qquad \nabla^2 f(x) \in \mathbb{R}^{n \times n}$

矩阵函数梯度

Fréchet 可微:

Defn:
$$\lim_{V \to 0} \frac{f(X+V) - f(X) - \langle V, G \rangle}{\|V\|} = 0$$

Gâteaux 可微:

Defn:
$$\lim_{t \to 0} \frac{f(X+tV) - f(X) - t\langle V, G \rangle}{t} = 0$$

$$\lim_{t \to 0} \frac{f(X+tV) - f(X)}{t} = \langle V, G \rangle$$

方向导数:

Defn:
$$\frac{df}{dX}[V] = \lim_{t \to 0} \frac{f(X+tV) - f(X)}{t}$$

$tr(AB) = tr(BA)$

$tr(A^T B) = \langle A, B \rangle$

E.x: $f(X) = a^T X b$ $\quad \dfrac{f(X+tV) - f(X)}{t} = \dfrac{a^T(X+tV)b - a^T X b}{t}$

$= \dfrac{\cancel{a^T X b} + \cancel{t} a^T V b - \cancel{a^T X b}}{\cancel{t}} = a^T V b = \langle V, G \rangle$

$a^T V b = tr(a^T V b) = tr(b a^T V) = \langle a^T b, V \rangle$ $\qquad \nabla f(x) = a^T b$

$tr(A \pm B) = tr(A) \pm tr(B)$

Ex: $f(x) = tr(X^T A X)$

$\dfrac{f(X+tV) - f(X)}{t} = \dfrac{tr((X+tV)^T A (X+tV)) - tr(X^T A X)}{t}$

$(X+tV)^T A (X+tV) = X^T A X + t^2 V^T A V + t(V^T A X + X^T A V)$

$\lim_{t \to 0} \dfrac{\cancel{tr(X^T A X)} + t^2 tr(V^T A V) + t \cdot tr(V^T A X + X^T A V) - \cancel{tr(X^T A X)}}{t}$

$= \lim_{t \to 0} \underbrace{\cancel{t} \cdot tr(V^T A V)}_{0} + \underbrace{tr(V^T A X + X^T A V)}$ $\qquad tr(A) = tr(A^T)$

$= \langle V, AX \rangle + \langle V, A^T X \rangle = \langle V, (A + A^T) X \rangle$

$\nabla f(x) = (A + A^T) X$

$\det(X_1 X_2) = \det(X_1)\cdot\det(X_2)$

X eigenvalue : $\lambda_1 \cdots \lambda_n$

$\det X = \prod_{i=1}^{n} \lambda_i$

Ex: $f(X) = \ln(\det X)$

$\det(X+tV) = \det\left(X^{\frac{1}{2}} I X^{\frac{1}{2}} + t X^{\frac{1}{2}} X^{-\frac{1}{2}} V X^{-\frac{1}{2}} X^{\frac{1}{2}}\right)$

$= \det\left(X^{\frac{1}{2}}(I + t X^{-\frac{1}{2}} V X^{-\frac{1}{2}}) X^{\frac{1}{2}}\right)$

$f(X+tV) - f(X)$

$= \ln(\det(X+tV)) - \ln(\det(X))$

$= \ln\left(\det\left(X^{\frac{1}{2}}(I + t X^{-\frac{1}{2}} V X^{-\frac{1}{2}}) X^{\frac{1}{2}}\right)\right) - \ln(\det(X))$

$= \ln\left(\det(X^{\frac{1}{2}}) \det(I + t X^{-\frac{1}{2}} V X^{-\frac{1}{2}}) \det(X^{\frac{1}{2}})\right) - \ln(\det(X))$

$= \ln\left(\det(X)\det(I + t X^{-\frac{1}{2}} V X^{-\frac{1}{2}})\right) - \ln(\det(X))$

$= \ln\left(\det(I + t X^{-\frac{1}{2}} V X^{-\frac{1}{2}})\right) + \cancel{\ln(\det(X))} - \cancel{\ln(\det(X))}$

$\lim_{t\to 0} \ln\left(\det(I + t X^{-\frac{1}{2}} V X^{-\frac{1}{2}})\right)$

$X^{-\frac{1}{2}} V X^{-\frac{1}{2}}$: eigenvalue $\lambda_1 \cdots \lambda_n$

$= \lim_{t\to 0} \ln\left(\prod_{i=1}^{n}(1 + t\lambda_i)\right)$

$\ln(1+x) \sim x + O(t^2)$

$= \lim_{t\to 0} \sum_{i=1}^{n} \ln(1 + t\lambda_i)$

$\operatorname{tr}(A) = \sum_{i=1}^{n} \lambda_i$

$= \lim_{t\to 0} t \sum_{i=1}^{n} \lambda_i + O(t^2)$

$= \lim_{t\to 0} t\cdot \operatorname{tr}(X^{-\frac{1}{2}} V X^{-\frac{1}{2}}) + O(t^2)$

$\operatorname{tr}(X^{-\frac{1}{2}} V X^{-\frac{1}{2}}) = \operatorname{tr}(X^{-\frac{1}{2}} X^{-\frac{1}{2}} V)$

$= \dfrac{\lim_{t\to 0} t\langle (X^{-1})^T, V\rangle + O(t^2)}{t} = \dfrac{f(X+tV) - f(X)}{t}$

$\nabla f(X) = (X^{-1})^T$

矩阵微分：$df = \langle \nabla f(X), dX \rangle$

① $d(X \pm Y) = dX \pm dY$

⊙：逐元素相乘

② $d(XY) = dXY + XdY$

③ $dX^T = (dX)^T$

④ $d\,tr(X) = tr(dX)$

⑤ $d\sigma(X) = \sigma'(X) \odot dX$

Ex: $f(x) = a^T X b$

$df = a^T dX b = \langle ab^T, dX \rangle$

$\Rightarrow \nabla f(x) = ab^T$

Ex: $f(x) = tr(X^T A X)$

$df = d\,tr(X^T A X)$

$= tr((dX)^T A X) + tr(X^T A dX)$

$= \langle AX, dX \rangle + \langle A^T X, dX \rangle$

$= \langle (A + A^T) X, dX \rangle$

$\Rightarrow \nabla f(x) = (A + A^T) X$

$d\sigma(X) = \sigma'(X) \odot dX$

Ex: $f(x) = a^T \exp(Xb)$

$df = a^T d(\exp(Xb))$

$= a^T (\exp(Xb) \odot d(Xb))$

$= tr(a^T (\exp(Xb) \odot d(Xb)))$

$= tr(b (a \odot \exp(Xb))^T dX)$

$= \langle a \odot \exp(Xb) b^T, dX \rangle$

$\Rightarrow \nabla f(x) = a \odot \exp(Xb) b^T$

$tr(A^T (B \odot C))$

$= tr((A \odot B)^T C)$