第二章 基础知识

修贤超

https://xianchaoxiu.github.io

目录

向量范数的定义

- **定义 2.1** 令记号 $\|\cdot\|: \mathbb{R}^n \to \mathbb{R}^+$ 是一种非负函数, 如果满足
  - **正定性** 对于 $\forall v \in \mathbb{R}^n$, 有 $\|v\| \geqslant 0$, 且 $\|v\| = 0$ 当且仅当 $v = 0_{n\times 1}$
  - **齐次性** 对于 $\forall v \in \mathbb{R}^n$ 和 $\alpha \in \mathbb{R}$, 有 $\|\alpha v\| = |\alpha|\|v\|$
  - **三角不等式** 对于 $\forall v, w \in \mathbb{R}^n$, 均成立 $\|v + w\| \leqslant \|v\| + \|w\|$

  则称 $\|\cdot\|$ 是定义在向量空间 $\mathbb{R}^n$ 上的**向量范数**

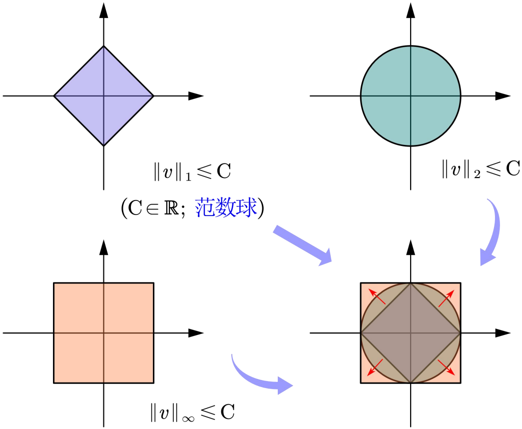
- 最常用的向量范数

$$\|v\|_p = \left(|v_1|^p + |v_2|^p + \cdots + |v_n|^p\right)^{\frac{1}{p}} \quad (p \geq 1)$$

$$\|v\|_\infty = \max_{1\leqslant j\leqslant n} |v_j|$$

向量范数的定义

- **不同范数所度量的距离分别具有怎样的特征呢?**



矩阵范数

- ℓ_1 范数 $\|A\|_1 = \sum_{i,j} |A_{ij}|$
- Frobenius 范数 $\|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2} = \sqrt{\mathrm{Tr}(AA^\top)}$
- 算子范数是一类特殊的矩阵范数，由向量范数诱导得到

$$\|A\|_{(m,n)} = \max_{x\in\mathbb{R}^n, \|x\|_{(n)}=1} \|Ax\|_{(m)}$$

  - $p=1$ 时，$\|A\|_{p=1} = \max_{\|x\|_1=1} \|Ax\|_1 = \max_{1\leqslant j\leqslant n} \sum_{i=1}^m |a_{ij}|$
  - $p=2$ 时，$\|A\|_{p=2} = \max_{\|x\|_2=1} \|Ax\|_2 = \sqrt{\lambda_{\max}(A^\top A)}$，又称为 A 的**谱范数**
  - $p=\infty$ 时，$\|A\|_{p=\infty} = \max_{\|x\|_\infty=1} \|Ax\|_\infty = \max_{1\leqslant i\leqslant m} \sum_{j=1}^n |a_{ij}|$

矩阵范数

- 核范数

$$\|A\|_* = \sum_{i=1}^{r} \sigma_i$$

- 矩阵内积

$$\langle A, B \rangle = \mathrm{Tr}(AB^\top) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}$$

- 命题 2.2 设 $A, B \in \mathbb{R}^{m \times n}$, 则

$$|\langle A, B \rangle| \leqslant \|A\|_F \|B\|_F$$

等号成立当且仅当 A 和 B 线性相关，即柯西不等式

- 性质 同一矩阵空间内，矩阵范数彼此之间是相互等价的

目录

梯度

- 定义 2.2 给定函数 $f : \mathbb{R}^n \to \mathbb{R}$，且 f 在点 x 的一个邻域内有意义，若存在向量 $g \in \mathbb{R}^n$ 满足

$$\lim_{p \to 0} \frac{f(x+p) - f(x) - \langle g, p \rangle}{\|p\|} = 0$$

其中 $\|\cdot\|$ 是任意的向量范数，称 f 在点 x 处可微（或 Fréchet 可微），g 为 f 在点 x 处的梯度，记作

$$\nabla f(x) = [\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots, \frac{\partial f(x)}{\partial x_n}]^\top$$

- 如果对区域 D 上的每一个点 x 都有 $\nabla f(x)$ 存在，则称 f 在 D 上可微

海瑟矩阵

- 定义 2.3 如果函数 $f(x): \mathbb{R}^n \to \mathbb{R}$ 在点 x 处的二阶偏导数 $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ 都存在，则 f 在点 x 处的海瑟矩阵为

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_3} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_3} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \frac{\partial^2 f(x)}{\partial x_n \partial x_3} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

- 当 $\nabla^2 f(x)$ 在区域 D 上的每个点 x 处都存在时，称 f 在 D 上二阶可微
- 若 $\nabla^2 f(x)$ 在 D 上还连续，则称 f 在 D 上二阶连续可微

梯度利普希茨连续

- **定义 2.4** 给定可微函数 f，若存在 $L>0$，对任意的 $x,y\in \mathrm{dom}\ f$ 有

$$\|\nabla f(x)-\nabla f(y)\|\le L\|x-y\|$$

  则称 f 是**梯度利普希茨连续的**，相应利普希茨常数为 L

- **引理 2.1** 设可微函数 $f(x)$ 的定义域为 $\mathbb{R}^n$ 且为梯度 L-利普希茨连续的，则函数 $f(x)$ 有**二次上界**

$$f(y)\le f(x)+\nabla f(x)^\top(y-x)+\frac{L}{2}\|y-x\|^2,\quad \forall\ x,y\in \mathrm{dom}\ f$$

- $f(x)$ 定义域的要求可减弱为凸集

梯度利普希茨连续

- **推论 2.1** 设可微函数 $f(x)$ 的定义域为 $\mathbb{R}^n$ 且存在一个全局极小点 x^*, 若 $f(x)$ 为梯度 L -利普希茨连续的, 则对任意的 x 有

$$\frac{1}{2L}\|\nabla f(x)\|^2 \leq f(x) - f(x^*)$$

证明 由于 x^* 是全局极小点，有

$$f(x^*) \leq f(y) \leq f(x) + \nabla f(x)^\top (y-x) + \frac{L}{2}\|y-x\|^2$$

上式对任意的 y 均成立，因此可对不等号右边取下确界

$$\begin{aligned} f(x^*) &\leq \inf_{y\in\mathbb{R}^n}\{f(x) + \nabla f(x)^\top (y-x) + \frac{L}{2}\|y-x\|^2\} \\ &= f(x) - \frac{1}{2L}\|\nabla f(x)\|^2 \end{aligned}$$

矩阵变量函数的导数

- 对于函数 $f(X)$，若存在矩阵 $G \in \mathbb{R}^{m\times n}$ 满足

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

其中 $\|\cdot\|$ 是任意矩阵范数，称矩阵变量函数 f 在 X 处 Fréchet 可微，G 为 f 在 Fréchet 可微意义下的梯度，记为

$$\nabla f(X) = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2n}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix}$$

矩阵变量函数的导数

- **定义 2.5** 如果对任意方向 $V \in \mathbb{R}^{m\times n}$，存在矩阵 $G \in \mathbb{R}^{m\times n}$ 满足

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

$$\Downarrow$$

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

则称 f 关于 X *Gâteaux* 可微，G 为 f 在 X 处 *Gâteaux* 可微意义下的梯度

- 当 f 是 Fréchet 可微函数时，f 也是 Gâteaux 可微的，且梯度相等

例 2.1

- **线性函数** $f(X)=\mathrm{Tr}(AX^\top B)$

$$\begin{aligned}\lim_{t\to 0}\frac{f(X+tV)-f(X)}{t}&=\lim_{t\to 0}\frac{\mathrm{Tr}(A(X+tV)^\top B)-\mathrm{Tr}(AX^\top B)}{t}\\&=\mathrm{Tr}(AV^\top B)=\langle BA,V\rangle\end{aligned}$$

$$\Rightarrow\quad \nabla f(X)=BA$$

- **二次函数** $f(X,Y)=\|XY-A\|_F^2$

$$\begin{aligned}f(X,Y+tV)-f(X,Y)&=\|X(Y+tV)-A\|_F^2-\|XY-A\|_F^2\\&=2\langle tXV,XY-A\rangle+t^2\|XV\|_F^2\\&=2t\langle V,X^\top(XY-A)\rangle+\mathcal{O}(t^2)\end{aligned}$$

$$\Rightarrow\quad \frac{\partial f}{\partial Y}=2X^\top(XY-A)$$

The Matrix Cookbook

[http://matrixcookbook.com]

Kaare Brandt Petersen
Michael Syskind Pedersen

Version: November 15, 2012

2.5.2 Second Order

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}^2) = 2\mathbf{X}^T \tag{106}$$

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}^2\mathbf{B}) = (\mathbf{XB} + \mathbf{BX})^T \tag{107}$$

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}^T\mathbf{B}\mathbf{X}) = \mathbf{BX} + \mathbf{B}^T\mathbf{X} \tag{108}$$

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{B}\mathbf{X}\mathbf{X}^T) = \mathbf{BX} + \mathbf{B}^T\mathbf{X} \tag{109}$$

目录

广义实值函数

- 在最优化领域，经常涉及量取 inf（sup） 操作，可能为无穷
- 定义 2.6 令 $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ 为广义实数空间，则映射

$$f:\ \mathbb{R}^n \to \overline{\mathbb{R}}$$

  称为广义实值函数
- 规定
  - $-\infty < \alpha < \infty, \forall \alpha \in \mathbb{R}$
  - $(+\infty) + (+\infty) = +\infty, \quad (+\infty) + \alpha = +\infty, \forall \alpha \in \mathbb{R}$

适当函数

- **定义 2.7** 给定广义实值函数 f 和非空集合 $\mathcal{X}$，如果存在 $x \in \mathcal{X}$ 使得 $f(x) < +\infty$，并且对任意的 $x \in \mathcal{X}$ 都有 $f(x) > -\infty$，则称函数 f 是关于集合 $\mathcal{X}$ 的**适当函数**
- 具体含义
  - 至少有一处取值不为正无穷
  - 处处取值不为负无穷
- 对于适当函数 f，规定其定义域

$$\operatorname{dom} f = \{x \mid f(x) < +\infty\}$$

- **若无特殊说明，定理中所讨论的函数均为适当函数**

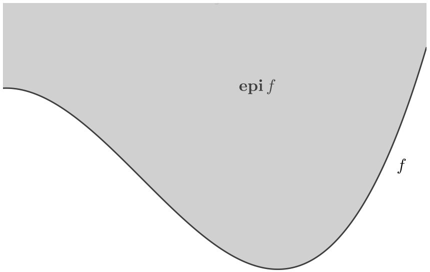
闭函数

- **定义 2.8** 设 f 为广义实值函数，α-**下水平集**定义为

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

- **定义 2.9** 设 f 为广义实值函数，**上方图**定义为

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

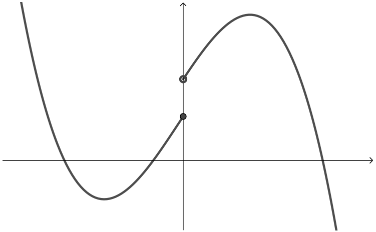


下半连续函数

- **定义 2.10** 设 f 为广义实值函数，若 $\operatorname{epi} f$ 为闭集，则称 f 为闭函数
- **定义 2.11** 设 f 为广义实值函数，若对任意的 $x \in \mathbb{R}^n$，有

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

则 $f(x)$ 为下半连续函数



闭函数与下半连续函数

- **定理 2.2** 设广义实值函数 $f:\ \mathbb{R}^n \to \overline{\mathbb{R}}$，则以下命题等价
  - $f(x)$ 的任意 α-下水平集都是闭集
  - $f(x)$ 是下半连续的
  - $f(x)$ 是闭函数
- **性质**
  - 若 f 与 g 均为适当的闭（下半连续）函数，并且 $\mathrm{dom}\ f \cap \mathrm{dom}\ g \neq \emptyset$，则 $f+g$ 也是闭（下半连续）函数
  - 若 f 为闭（下半连续）函数, 则 $f(Ax+b)$ 也为闭（下半连续）函数

目录

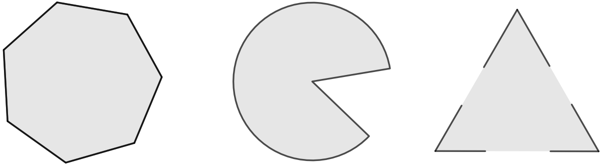
凸集的几何定义

- **定义 2.12** 若过集合 $\mathcal{C}$ 中的任意两点的直线都在 $\mathcal{C}$ 内, 则称 $\mathcal{C}$ 为仿射集, 即

$$x_1, x_2 \in \mathcal{C} \quad \Rightarrow \quad \theta x_1 + (1-\theta)x_2 \in \mathcal{C}, \forall \theta \in \mathbb{R}$$

- **定义 2.13** 若连接集合 $\mathcal{C}$ 中的任意两点的线段都在 $\mathcal{C}$ 内, 则称 $\mathcal{C}$ 为凸集, 即

$$x_1, x_2 \in \mathcal{C} \quad \Rightarrow \quad \theta x_1 + (1-\theta)x_2 \in \mathcal{C}, \forall 0 \leqslant \theta \leqslant 1$$



(a) (b) (c)

凸集的性质

- 若 $\mathcal{S}$ 是凸集, 则 $k\mathcal{S} = \{ks \mid k \in \mathbb{R}, s \in \mathcal{S}\}$ 是凸集
- 若 $\mathcal{S}$ 和 $\mathcal{T}$ 均是凸集, 则 $\mathcal{S} + \mathcal{T} = \{s + t \mid s \in \mathcal{S}, t \in \mathcal{T}\}$ 是凸集
- 若 $\mathcal{S}$ 和 $\mathcal{T}$ 均是凸集, 则 $\mathcal{S} \cap \mathcal{T}$ 是凸集

  证明 设 $x, y \in \mathcal{S} \cap \mathcal{T}$ 且 $\theta \in [0, 1]$. 由于 $\mathcal{S}$ 和 $\mathcal{T}$ 均为凸集, 则

  $$\theta x + (1 - \theta) y \in \mathcal{S} \cap \mathcal{T}$$

- 凸集的内部和闭包都是凸集
- 任意多凸集的交都是凸集

凸组合和凸包

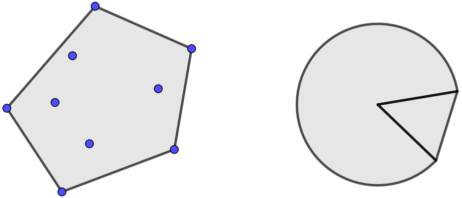
- 形如

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$

$$\theta_1 + \cdots + \theta_k = 1, \theta_i \geqslant 0, i = 1, \cdots, k$$

的点称为 $x_1, \cdots, x_k$ 的凸组合

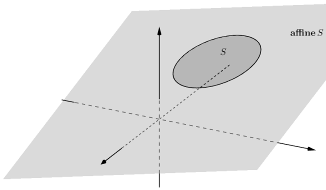
- 集合 $\mathcal{S}$ 的所有点的凸组合构成的点集为 $\mathcal{S}$ 的凸包, 记为 $\operatorname{conv} \mathcal{S}$



- $\operatorname{conv} \mathcal{S}$ 是包含 $\mathcal{S}$ 的最小凸集

仿射组合和仿射包

- 定义 2.14 形如
$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$
$$\theta_1 + \cdots + \theta_k = 1, \theta_i \in \mathbb{R}, i = 1, \cdots, k$$
的点称为 $x_1, \cdots, x_k$ 的仿射组合
- 集合 $\mathcal{S}$ 的所有点的仿射组合构成的点集为 $\mathcal{S}$ 的仿射包, 记为 affine $\mathcal{S}$



- affine $\mathcal{S}$ 是包含 $\mathcal{S}$ 的最小仿射集

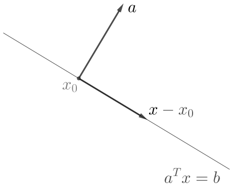
锥组合和凸锥

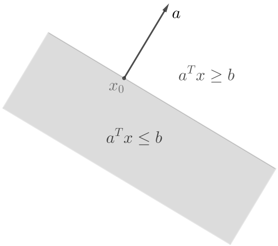
- 形如
$$x = \theta_1 x_1 + \cdots + \theta_k x_k, \theta_i > 0, i = 1, \cdots, k$$
的点称为 $x_1, \cdots, x_k$ 的**锥组合**
- 若集合 $\mathcal{S}$ 中任意点的锥组合都在 $\mathcal{S}$ 中, 则称 $\mathcal{S}$ 为**凸锥**

- 锥组合不要求系数的和为 1

超平面和半空间

- 任取非零向量 $a \in \mathbb{R}^n$, 称 $\{x \mid a^\top x = b\}$ 为**超平面**, $\{x \mid a^\top x \leqslant b\}$ 为**半空间**
- 满足线性等式和不等式组点的集合 $\{x \mid Ax \leqslant b, Cx = d\}$ 称为**多面体**





- 超平面是仿射集和凸集, 半空间是凸集但不是仿射集
- 多面体是有限个半空间和超平面的交

分离超平面定理

- 定理 2.3 如果 $\mathcal{C}$ 和 $\mathcal{D}$ 是不相交的凸集, 则存在非零向量 a 和常数 b, 使得

$$a^\top x \leqslant b, \forall x \in \mathcal{C} \quad \text{且} \quad a^\top x \geqslant b, \forall x \in \mathcal{D}$$

  即超平面 $\{x \mid a^\top x = b\}$ 分离了 $\mathcal{C}$ 和 $\mathcal{D}$

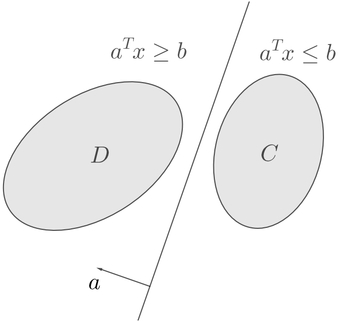
- 定理 2.4 如果存在非零向量 a 和常数 b, 使得

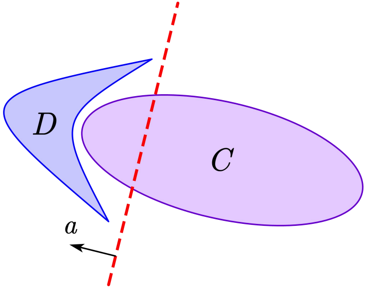
$$a^\top x < b, \forall x \in \mathcal{C} \quad \text{且} \quad a^\top x > b, \forall x \in \mathcal{D}$$

  即超平面 $\{x \mid a^\top x = b\}$严格分离了 $\mathcal{C}$ 和 $\mathcal{D}$

分离超平面的示意

- 在 $\mathbb{R}^2$ 中的 2 个凸集使用超平面即可轻松划分，但遇到非凸集合就必须使用更加复杂的平面





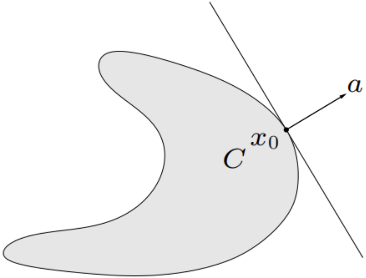
支撑超平面

- **定义 2.5** 给定集合 $\mathcal{C}$ 及其边界点 x_0, 如果 $a \neq 0$ 满足 $a^\top x \leqslant a^\top x_0, \forall x \in \mathcal{C}$, 则称集合

$$\{x \mid a^\top x = a^\top x_0\}$$

为 $\mathcal{C}$ 在边界点 x_0 处的**支撑超平面**

- **定理 2.5** 若 $\mathcal{C}$ 是凸集, 则 $\mathcal{C}$ 的任意边界点处都存在支撑超平面



球和椭球

- 称空间中到点 x_c 的距离小于等于定值 r 的集合为(欧几里得) 球, 即

$$\mathrm{B}(x_c, r) = \{x \mid \|x - x_c\|_2 \leqslant r\} = \{x_c + ru \mid \|u\|_2 \leqslant 1\}$$

- 设形如

$$\{x \mid (x - x_c)^\top P^{-1}(x - x_c) \leqslant 1\} = \{x_c + Au \mid \|u\|_2 \leqslant 1\}$$

的集合为椭球, 其中 x_c 为椭球中心, P 为对称正定, 且 A 非奇异

- 令 $\|\cdot\|$ 是任意一个范数, 称

$$\{x \mid \|x - x_c\| \leqslant r\}$$

为中心为 x_c 半径为 r 的范数球

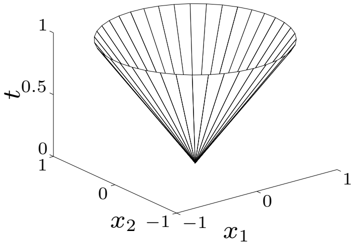
范数锥

- 形如

$$\{(x,t) \mid \|x\| \leqslant t\}$$

的集合为范数锥

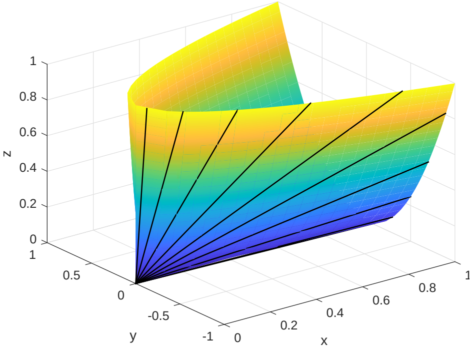
- 使用 $\|\cdot\|_2$ 度量距离的锥为二次锥，也称冰淇淋锥



- 范数球和范数锥都是凸集

(半) 正定锥

- 记 $\mathcal{S}^n$ 为**对称矩阵**的集合，即 $\mathcal{S}^n = \{X \in \mathbb{R}^{n\times n} \mid X^\top = X\}$
- 记 $\mathcal{S}^n_+$ 为**半正定矩阵**的集合，即 $\mathcal{S}^n_+ = \{X \in \mathcal{S}^n \mid X \succeq 0\}$
- 记 $\mathcal{S}^n_{++}$ 为**正定矩阵**的集合，即 $\mathcal{S}^n_{++} = \{X \in \mathcal{S}^n \mid X \succ 0\}$

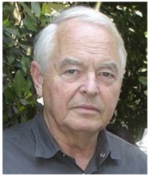


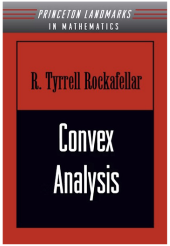
对于矩阵 $\begin{pmatrix} x & y \\ y & z \end{pmatrix}$，其特征值应全部大于等于 0

$$\Downarrow$$

$$\{(x,y,z) \mid x \geqslant 0, z \geqslant 0, xz \geqslant y^2\}$$

凸分析





R T Rockafellar

FOLLOW

Professor of Mathematics, University of Washington
Verified email at uw.edu - Homepage

optimization　convex analysis　variational analysis　economic equilibrium　risk and reliability

TITLE	CITED BY	YEAR
Solving problems in convex optimal control by progressive decoupling in the dynamics RT Rockafellar Mathematical Control and Related Fields, 0-0		2024
Primal–Dual Stability in Local Optimality M Benko, RT Rockafellar Journal of Optimization Theory and Applications, 1-30	2	2024
Generalized Nash Equilibrium from a Robustness Perspective in Variational Analysis RT Rockafellar Set-Valued and Variational Analysis 32 (2), 19	1	2024
Distributional robustness, stochastic divergences, and the quadrangle of risk RT Rockafellar Computational Management Science 21 (1), 34	2	2024
Data poisoning attacks on traffic state estimation and prediction F Wang, X Wang, Y Hong, RT Rockafellar, XJ Ban Transportation Research Part C: Emerging Technologies, 104577	1	2024
Generalizations of the proximal method of multipliers in convex optimization RT Rockafellar Computational Optimization and Applications 87 (1), 219-247	6	2024

Q&A

Thank you!

感谢您的聆听和反馈