最优化理论基础

Hailiang ZHAO @ ZJU-CS
http://hliangzhao.me

2021 年 7 月 14 日

本 *slide* 所依赖的基本数学知识请参见http://hliangzhao.me/math/math.pdf；《最优化简介》请参见http://hliangzhao.me/math/opt1.pdf。

向量范数

范数是高维空间中**计量事物自身长度**的一种方式。

定义

向量范数 是一个从向量空间 $\mathbb{R}^n$ 到实数域 $\mathbb{R}$ 的非负函数 $\|\cdot\|$，且满足

- 正定性：$\forall v \in \mathbb{R}^n$，$\|v\| \geq 0$ 且 $\|v\| = 0 \Longleftrightarrow v = 0$
- 齐次性：$\forall v \in \mathbb{R}^n, \alpha \in \mathbb{R}$，$\|\alpha v\| = |\alpha| \|v\|$
- 三角不等式：$\forall v, w \in \mathbb{R}^n$，$\|v + w\| \leq \|v\| + \|w\|$

l_p 范数（$p \geq 1$）：$\|v\|_p \triangleq \left(|v_1|^p + ... + |v_n|^p\right)^{\frac{1}{p}}$，默认 $p = 2$

由正定阵 A 诱导的范数：$\|v\|_A \triangleq \sqrt{v^{\mathrm{T}} A v}$

定理

柯西不等式 $\forall a, b \in \mathbb{R}^n$，$|a^{\mathrm{T}} b| \leq \|a\| \|b\|$。

矩阵范数

矩阵的 l_p 范数（$p \geq 1$）：

$$\|A\|_p \triangleq \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^p\right)^{\frac{1}{p}}$$

$p=2$ 时，称之为矩阵的 Frobenius 范数，记为 $\|\cdot\|_F$：

$$\|A\|_F \triangleq \sqrt{\sum_{i,j} a_{ij}^2} = \sqrt{\mathrm{Tr}\left(A^\mathrm{T}A\right)}$$

定理

正交不变性 对于任意的正交矩阵 $U \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^{n\times n}$，

$$\|UAV\|_F^2 = \mathrm{Tr}\left(UAVV^\mathrm{T}A^\mathrm{T}U^\mathrm{T}\right) = \|A\|_F^2$$

这里运用了性质 $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$。

矩阵范数

算子范数（由向量范数诱导的矩阵范数）:
给定 m 维和 n 维空间的向量范数 $\|\cdot\|_{(m)}$ 和 $\|\cdot\|_{(n)}$,

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

显然有结论:

$$\|Ax\|_{(m)} \le \|A\|_{(m,n)}\|x\|_{(n)}$$

（矩阵范数和给定的向量范数**相容**）

若 $\|\cdot\|_{(m)}$ 和 $\|\cdot\|_{(n)}$ 均取相应空间的 l_p 范数，如 $p=2$，则可得到**矩阵的 2 范数**:

$$\|A\|_2 \triangleq \max_{\forall x\in\mathbb{R}^n, \|x\|_2=1} \|Ax\|_2 = \max_{r=1,\ldots,\mathrm{rank}(A)} \sigma_r (\text{最大奇异值})$$

严格区分矩阵的 2 范数和矩阵的 F 范数（l_2 范数）!

矩阵范数

核范数 (非零奇异值之和):

$$\|A\|_* \triangleq \sum_{r=1}^{\operatorname{rank}(A)} \sigma_r$$

内积用来表征两个矩阵（或其张成的空间）之间的夹角。常用的内积是 Frobenius 内积:

$$\langle A, B\rangle \triangleq \operatorname{Tr}(AB^{\mathrm{T}}) = \sum_{i,j} a_{ij} b_{ij}$$

显然，$\langle A, A\rangle = \|A\|_F^2$。

定理

矩阵范数的柯西不等式 $\forall A, B \in \mathbb{R}^{m\times n}$，$|\langle A, B\rangle| \le \|A\|_F \|B\|_F$。在 A 和 B 线性相关的时候取等。

导数

定义

梯度 若函数 $f:\mathbb{R}^n\to\mathbb{R}$ 在点 x 的一个邻域内有意义且存在 $g\in\mathbb{R}^n$ 使得

$$\lim_{p\to 0}\frac{f(x+p)-f(x)-g^{\mathrm{T}}p}{\|p\|}=0,$$

则称 f 在点 x 处**可微**。我们称 g 为 f 在点 x 处的梯度，记为 $\nabla f(x)$。若区域 D 上每一个点 x 都有 $\nabla f(x)$ 存在，则称 f 在 D 上可微。

基于该定义，若依次对每一个维度 i 取 $p=\varepsilon e_i$（e_i 为第 i 维的标准基），则可得到

$$\nabla f(x)=\left[\frac{\partial f(x)}{\partial x_1},\dots,\frac{\partial f(x)}{\partial x_n}\right]^{\mathrm{T}}$$

二阶导数

定义

海瑟矩阵 若函数 $f:\mathbb{R}^n\to\mathbb{R}$ 在点 x 的二阶偏导数 $\left\{\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right\}_{i,j=1,\ldots,n}$ 都存在，则

$$\nabla^2 f(x) \triangleq \left[\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right]_{i,j=1,\ldots,n} \in \mathbb{R}^{n\times n}$$

称为 f 在点 x 处海瑟矩阵。若区域 D 上每一个点 x 都有 $\nabla^2 f(x)$ 存在，则称 f 在 D 上可微。若 $\nabla^2 f(x)$ 在 D 上还连续，则称 f 在 D 上二阶连续可微，此时 $\nabla^2 f(x)$ 是一个对称阵。

定义

雅可比矩阵 对于向量值函数 $f:\mathbb{R}^n\to\mathbb{R}^m$，可类比**梯度**定义它的雅可比矩阵：

$$J(x) = \left[\frac{\partial f_i(x)}{\partial x_j}\right]_{ij} \in \mathbb{R}^{n\times m}$$

泰勒展开

若 $f:\mathbb{R}^n \to \mathbb{R}$ 是连续可微的，则

$$f(x+p) = f(x) + \nabla f(x+tp)^{\mathrm{T}} p,$$

其中 $0 < t < 1$，$p \in \mathbb{R}^n$。

若 $f:\mathbb{R}^n \to \mathbb{R}$ 是二阶连续可微的，则

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp) p dt \qquad (\text{对上式求导})$$

$$f(x+p) = f(x) + \nabla f(x)^{\mathrm{T}} p + \frac{1}{2} p^{\mathrm{T}} \nabla^2 f(x+tp) p,$$

其中 $0 < t < 1$，$p \in \mathbb{R}^n$。

梯度 Lipschitz 连续

定义

梯度 Lipschitz 连续 对于可微函数 f，若存在 $L > 0$ 对于任意 $x, y \in \mathbf{dom} f$ 有

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|,$$

则称 f 是 L-Lipschitz 连续的。

L-Lipschitz 连续表明 $\nabla f(x)$ 的变化可以被自变量 x 的变化所控制，这在算法的收敛性证明中将发挥重要的作用。

接下来两页的内容常常出现在许多算法的收敛性分析中。

梯度 Lipschitz 连续

定理

二次上界 *L-Lipschitz* 连续的函数 $f:\mathbb{R}^n\to\mathbb{R}$ 有二次上界：

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

证明.

通过构造辅助函数 $g(t)=f\big(x+t(y-x)\big),t\in[0,1]$ 来推导

$$f(y)-f(x)-\nabla f(x)^{\mathrm{T}}(y-x)$$

的上界。 □

这说明 L-Lipschitz 连续函数被一个二次函数的上界所控制，即 $f(x)$ 的增长速度不超过二次。（$\mathbf{dom}f$ 仅需是凸集即可，不需要是 $\mathbb{R}^n$。）

梯度 Lipschitz 连续

定理

差距的下界 若 L-*Lipschitz* 连续的函数 $f:\mathbb{R}^n\to\mathbb{R}$ 的定义域为 $\mathbb{R}^n$ 且存在一个全局极小点 x^*，则

$$\frac{1}{2L}\|\nabla f(x)\|^2 \le f(x)-f(x^*), \forall x\in\mathbb{R}^n.$$

证明.

根据上一定理的结论可得

$$\begin{aligned} f(x^*) &\le \inf_{y\in\mathbb{R}^n}\left\{f(x)+\nabla f(x)^{\mathrm{T}}(y-x)+\frac{L}{2}\|y-x\|^2\right\} \\ &= f(x)-\frac{1}{2L}\|\nabla f(x)\|^2. \qquad \triangleright -\frac{b}{2a} \end{aligned}$$

$\square$

矩阵变量函数的导数

定义

Frechet **可微** 对于函数 $f:\mathbb{R}^{m\times n}\to\mathbb{R}$，若存在矩阵 $G\in\mathbb{R}^{m\times n}$ 满足

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

则称 f 在 X 处 Frechet 可微，G 是相应意义下的梯度，记为 $\nabla f(X)$。同样有：

$$\nabla f(X)=\left[\frac{\partial f}{\partial x_{ij}}\right]_{ij}$$

矩阵变量函数的导数

定义

Gateaux **可微** 对于函数 $f:\mathbb{R}^{m\times n}\to\mathbb{R}$，若存在矩阵 $G\in\mathbb{R}^{m\times n}$ 对任意方向 $V\in\mathbb{R}^{m\times n}$ 和任意 $t\in\mathbb{R}$ 满足

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

则称 f 在 X 处 Gateaux 可微，G 是相应意义下的梯度，仍然记为 $\nabla f(X)$。

Gateaux 可微其实是方向导数的某种推广。此外，若 f 是 *Frechet* 可微的，则 f 是 *Gateaux* 可微的，且二者意义下的梯度相等。

我们主要关注 Gateaux 可微。

利用 Gateaux 可微的定义计算梯度

以下内容利用 $\mathrm{Tr}(A+B)=\mathrm{Tr}(A)+\mathrm{Tr}(B)$、$\mathrm{Tr}(AB)=\mathrm{Tr}(BA)$ 以及 $\|\cdot\|_F$ 的定义来计算。

示例 1：考虑线性函数 $f(X)=\mathrm{Tr}(AX^{\mathrm{T}}B)$，其中 $A\in\mathbb{R}^{p\times n}$，$B\in\mathbb{R}^{m\times p}$，$X\in\mathbb{R}^{m\times n}$，则

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

所以 $\nabla f(X)=BA$。

示例 2：考虑二次函数 $f(X,Y)=\frac{1}{2}\|XY-A\|_F^2$，其中 $(X,Y)\in\mathbb{R}^{m\times p}\times\mathbb{R}^{p\times n}$，则

$$
f(X,Y+tV,Y)-f(X,Y)=t\langle V,(XY-A)Y^{\mathrm{T}}\rangle+\mathcal{O}(t^2),
$$

所以 $\nabla_Y f(X,Y)=X^{\mathrm{T}}(XY-A)$。同理 $\nabla_X f(X,Y)=(XY-A)Y^{\mathrm{T}}$。

利用 Gateaux 可微的定义计算梯度

示例 3：考虑 *ln-det* 函数 $f(X)=\ln(\det(X))$，其中 $X\in\mathcal{S}^n_{++}$，则给定 $X\succ 0$，对任意方向 $V\in\mathcal{S}^n$ 以及 $t\in\mathbb{R}$，有

$$
\begin{aligned}
& f(X+tV)-f(X) \\
&= \ln\left(\det\left(X^{1/2}(I+tX^{-1/2}VX^{-1/2})X^{1/2}\right)\right)-\ln\left(\det(X)\right) \\
&= \ln\left(\det\left(I+tX^{-1/2}VX^{-1/2}\right)\right). \qquad \triangleright \det(AB)=\det(A)\det(B)
\end{aligned}
$$

对正定阵 $X^{-1/2}VX^{-1/2}$ 正交对角化[1]，设其特征值为 $\lambda_1,\ldots,\lambda_n$，则

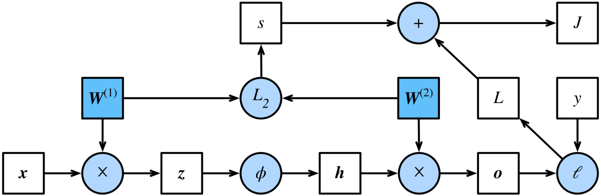
$$
\begin{aligned}
\ln\left(\det(I+tX^{-1/2}VX^{-1/2})\right) &= \ln\prod_{i=1}^n(1+t\lambda_i) \\
&= \sum_{i=1}^n t\lambda_i+\mathcal{O}(t^2) \qquad \triangleright x\to 0: \ln(1+x)=x+\mathcal{O}(x^2) \\
&= t\mathrm{Tr}(X^{-1/2}VX^{-1/2})+\mathcal{O}(t^2) \qquad \triangleright \text{迹是特征值之和} \\
&= t\langle (X^{-1})^{\mathrm{T}},V\rangle+\mathcal{O}(t^2) \qquad \triangleright \nabla f(X)=(X^{-1})^{\mathrm{T}}
\end{aligned}
$$

[1] 更多细节见`http://hliangzhao.me/math/math.pdf`。

自动微分

依据微积分中的链式法则，沿着从计算图中输出层到输入层的顺序，依次计算并存储目标函数关于计算图中各层的中间变量以及参数的梯度。

第 l 层的误差可由第 $l+1$ 层的误差得到。



正向传播的计算图示例。

细节请参考文档`http://hliangzhao.me/math/cheatsheet.pdf`的章节 5.2。

广义实值函数

定义

广义实值函数 令 $\bar{\mathbb{R}} \triangleq \mathbb{R} \cup \{\pm\infty\}$ 为广义实数空间，则 $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ 被称为广义实值函数。

定义

适当函数 给定广义实值函数 f 和非空集合 $\mathcal{X}$，若

$$\begin{cases} \exists x \in \mathcal{X}, f(x) < +\infty \\ \forall x \in \mathcal{X}, f(x) > -\infty, \end{cases}$$

则称 f 是关于 $\mathcal{X}$ 的适当函数（**否则求 minimum 没有意义**）。

若无特别指示，接下来及后续文档中提及的 f **均为适当函数**，且默认 $\mathbb{R} = \bar{\mathbb{R}}$。

闭函数

定义

α-下水平集 对于函数 $f:\mathbb{R}^n \to \mathbb{R}$，称

$$C_\alpha \triangleq \{x \mid f(x) \leq \alpha\}$$

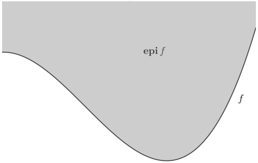
为 f 的 α-下水平集。α-下水平集是不超过某个阈值的自变量的集合。

定义

上方图（用于反推函数 f 的性质）对于函数 $f:\mathbb{R}^n \to \mathbb{R}$，称

$$\mathbf{epi} f \triangleq \{(x,t) \in \mathbb{R}^{n+1} \mid f(x) \leq t\}$$

为 f 的上方图。



闭函数

定义

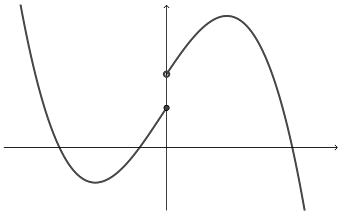
闭函数 对于函数 $f:\mathbb{R}^n \to \mathbb{R}$，若 $\textbf{epi}f$ 为闭集（集合所有的极限点都是这个集合中的点——集合包含它的边界)，则称 f 为闭函数。

定义

下半连续函数 对于函数 $f:\mathbb{R}^n \to \mathbb{R}$，若 $\forall x \in \mathbb{R}^n$，

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

则称 f 为下半连续函数。



闭函数

定理

闭函数和下半连续函数的等价性 对于函数 $f : \mathbb{R}^n \to \mathbb{R}$，以下命题等价：

1. f 的任意 α-下水平集都是闭集；
2. f 是下半连续的；
3. f 是闭函数。

证明.

$2 \Rightarrow 3$：设 $(x_k, y_k) \in \textbf{epi} f$ 且 $\lim_{k\to\infty}(x_k, y_k) = (\bar{x}, \bar{y})$，则

$$f(\bar{x}) \le \liminf_{k\to\infty} f(x_k) \le \lim_{k\to\infty} y_k = \bar{y},$$

这说明 $(\bar{x}, \bar{y}) \in \textbf{epi} f$，所以 $\textbf{epi} f$ 是闭集。 □

闭函数

定理

闭函数和下半连续函数的等价性 对于函数 $f : \mathbb{R}^n \to \mathbb{R}$，以下命题等价：

1. f 的任意 α-下水平集都是闭集；
2. f 是下半连续的；
3. f 是闭函数。

证明.

$3 \Rightarrow 1$：取 α-下水平集的元素 $x_k \to \bar{x}$，有 $(x_k, \alpha) \in \textbf{epi} f$ 且 $(x_k, \alpha) \to (\bar{x}, \alpha)$，因为 f 是闭函数，所以必然有 $(\bar{x}, \alpha) \in \textbf{epi} f$，即 $f(\bar{x}) \le \alpha$。这说明 f 的任意 α-下水平集都是闭集。 $\square$

闭函数

定理

闭函数和下半连续函数的等价性 对于函数 $f:\mathbb{R}^n\to\mathbb{R}$，以下命题等价：

1. f 的任意 α-下水平集都是闭集；
2. f 是下半连续的；
3. f 是闭函数。

证明.

$1\Rightarrow 2$：反证法。假设存在序列 $\{x_k\}\to\bar{x}(k\to\infty)$ 但 $f(\bar{x})>\liminf_{k\to\infty}f(x_k)$，取 t 使得

$$f(\bar{x})>t>\liminf_{k\to\infty}f(x_k).$$

根据下极限的定义，$\{x_k\mid f(x_k)\le t\}$ 中必然含有无穷多个 x_k，因此 $\{x_k\}$ 中存在子列 $\{x_{k_l}\}$ 使得 $f(x_{x_l})\le t$ 且 $\lim_{x_{k_l}}=\bar{x}$，则 t-下水平集不是闭集，这与命题 1 矛盾。 □

闭函数

以上关于闭函数的相关命题将为后面的定理证明提供极大的方便。
闭函数间的简单运算会保持原有性质：

1. **加法**：若 f 和 g 均为闭函数，且 $\mathbf{dom} f \cap \mathbf{dom} g \neq \varnothing$，则 $f+g$ 也是闭函数。
2. **仿射映射的复合**：若 f 为闭函数，则 $f(Ax+b)$ 也为闭函数。
3. **取上确界**：若每一个 f_α 均为闭函数，则 $\sup_\alpha f_\alpha(x)$ 也为闭函数。

直线与仿射集

定义

直线 对于 $\mathbb{R}^n$ 中的两个点 x_1 和 x_2，形如

$$y = \theta x_1 + (1-\theta)x_2$$

的点组成了过点 x_1 和 x_2 的直线。当 $0 \le \theta \le 1$ 时，这样的点形成了连接点 x_1 和 x_2 的线段。

定义

仿射集 若过集合 C 中任意两点的**直线**都在 C 内，则称 C 为仿射集，即

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

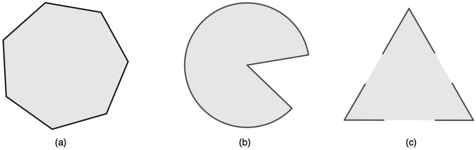
凸集

定义

凸集 若过集合 C 中任意两点的**线段**都在 C 内，则称 C 为凸集，即

$$x_1, x_2 \in C \quad \Longrightarrow \quad \theta x_1 + (1-\theta)x_2 \in C, \forall \theta \in [0,1].$$

显然，仿射集都是凸集。



（a）为凸集，（b）、（c）为非凸集。

凸组合与凸包

定义

凸组合（凸集定义的扩展——从两个点扩展到多个点来定义）对于给定的点 $x_1, ..., x_k$，$\forall \theta_i \geq 0, \sum_i \theta_i = 1$，形如

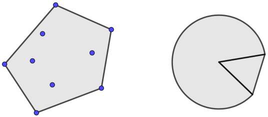
$$x = \sum_{i=1}^{k} \theta_i x_i$$

的点称为 $x_1, ..., x_k$ 的凸组合。

定义

凸包 集合 $\mathcal{S}$ 中的点所有可能的凸组合构成的集合称作 $\mathcal{S}$ 的凸包，记为 $\mathbf{conv}\mathcal{S}$。

$\mathbf{conv}\mathcal{S}$ **是包含** $\mathcal{S}$ **的最小的凸集**。



离散点集和扇形的凸包。

仿射包

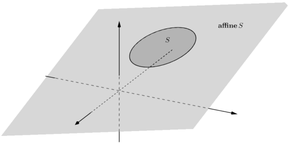
去掉凸组合的定义中的 $\{\theta_i\}_{\forall i} \geq 0$ 的限制，即可得到仿射包的概念。

定义

仿射包 设 $\mathcal{S}$ 为 $\mathbb{R}^n$ 的子集，则称

$$\left\{x \mid x = \sum_{i=1}^{k} \theta_i x_i, \quad x_1, \ldots, x_k \in \mathcal{S}, \sum_{\theta_i} = 1\right\}$$

为集合 $\mathcal{S}$ 的仿射包，记为 **affine**$\mathcal{S}$。



$\mathbb{R}^n$ 中的圆盘 $\mathcal{S}$ 的仿射包是一个平面。

affine$\mathcal{S}$ **是包含** $\mathcal{S}$ **的最小的仿射集**。

锥组合与凸锥

定义

锥组合 形如 $x = \theta_1 x_1 + \theta_2 x_2, \theta_1 > 0, \theta_2 > 0$ 的点称为点 x_1, x_2 的锥组合。

定义

凸锥 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。

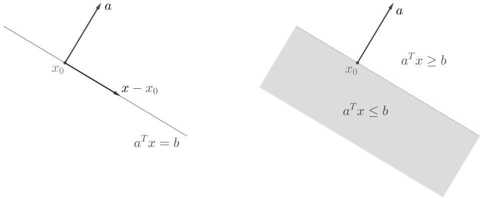
凸锥。

重要的凸集

超平面与半空间：

1. **超平面**： 集合 $\{x|a^{\mathrm{T}}x = b, a \neq 0\}$
2. **半空间**： 集合 $\{x|a^{\mathrm{T}}x \leq b, a \neq 0\}$

其中 a 是对应超平面和半空间的法向量。一个超平面将 $\mathbb{R}^n$ 分为两个半空间。显然，超平面是仿射集和凸集，半空间是凸集但不是仿射集（因为一端封闭）。



超平面和半空间。

重要的凸集

球、椭球和锥：

1. **球**：空间中到某个点距离不大于某个常数的点的结合。我们将

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

称为中心为 x_c、半径为 r 的欧几里得球。

2. **椭球**：我们将形如

$$\left\{x \mid (x - x_c)^{\mathrm{T}} P^{-1}(x - x_c) \le 1, P \in \mathcal{S}_{++}^n\right\}$$

的集合称为椭球。椭球的另一种表示为 $\{x_c + Au \mid \|u\| + 2 \le 1\}$，其中 A 为非奇异方阵。

以上定义中使用的是 l_2 范数，因此得到的都是欧几里得空间的距离。如果不限制范数的类型，即 $\|\cdot\|$ 是任意一个范数，则称 $\{x \mid \|x - x_c\| \le r\}$ 是中心为 x_c，半径为 r 的**范数球**，称 $\{(x, t) \mid \|x\| \le t\}$ 为范数锥。欧几里得范数锥也称为**二次锥**。

重要的凸集

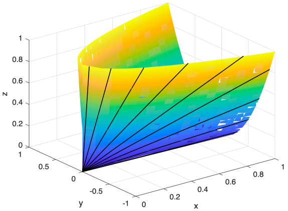
多面体和半正定锥:

1. **多面体**: 满足线性等式和不等式组的点集

$$\{x \mid Ax \leq b, Cx = d\}$$

被称为多面体。多面体是有限个半空间和超平面的交集，因此是凸集。

2. **(半) 正定锥**: $\mathcal{S}_+^n$ ($n \times n$ 半正定阵的集合) 是凸锥。因此又称之为半正定锥。同理 $\mathcal{S}_{++}^n$ 是正定锥。



二维半正定锥 $\mathcal{S}_+^2$。

保凸运算

证明一个集合是凸集的方法：（1）定义；（2）简单凸集的保凸运算。常用的保凸元算是**取交集**和**仿射变换**（缩放、平移、投影）。

定理

取交集 任意多个凸集的交为凸集。即，若 $C_i, i \in \mathcal{I}$ 是凸集，则

$$\bigcap_{i \in \mathcal{I}} C_i$$

为凸集。这里不要求指标集 $\mathcal{I}$ 可列。

定理

仿射变换 设 $f : \mathbb{R}^n \to \mathbb{R}$ 是仿射变换 $f(x) = Ax + b$，则

1. 凸集在 f 下的像是凸集：
   $S \subseteq \mathbb{R}^n$ 是凸集 $\Longrightarrow f(S) \triangleq \{f(x) \mid x \in S\}$ 是凸集；
2. 凸集在 f 下的原像是凸集：
   $C \subseteq \mathbb{R}^m$ 是凸集 $\Longrightarrow f^{-1}(C) \triangleq \{x \in \mathbb{R}^n \mid f(x) \in C\}$ 是凸集。

分离超平面

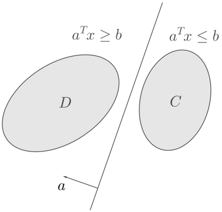
超平面可以分离不相交的凸集。

定理

分离超平面定理 如果 C 和 D 是两个不相交的凸集，则存在非零向量 a 和常数 b，使得

$$\begin{cases} a^{\mathrm{T}}x \le b & \forall x \in C \\ a^{\mathrm{T}}x \ge b & \forall x \in D. \end{cases}$$

即超平面 $\{x \mid a^{\mathrm{T}}x = b\}$ 分离了 C 和 D。



分离超平面。

支撑超平面

当 C 是闭凸集，D 是单点集时，我们有严格分离定理。

定理

严格分离定理 设 C 是闭凸集，点 $x_0 \notin C$，则存在非零向量 a 和常数 b，使得

$$\begin{cases} a^{\mathrm{T}} x < b & \forall x \in C \\ a^{\mathrm{T}} x_0 > b. \end{cases}$$

当点 x_0 恰好在凸集 C 的边界上时，即可构造支撑超平面。

定义

支撑超平面 给定集合 C 及其边界上的一点 x_0，如果 $a \neq 0$ 满足 $\forall x \in C, a^{\mathrm{T}} x \leq a^{\mathrm{T}} x_0$，那么集合

$$\{x \mid a^{\mathrm{T}} x = a^{\mathrm{T}} x_0\}$$

为 C 在边界点 x_0 处的支撑超平面。

几何上，$\{x \mid a^{\mathrm{T}} x = a^{\mathrm{T}} x_0\}$ 与 C 在点 x_0 处相切且半空间 $a^{\mathrm{T}} x \leq a^{\mathrm{T}} x_0$ 包含 C。

支撑超平面

定义

支撑超平面（重述）给定集合 C 及其边界上的一点 x_0，如果 $a \neq 0$ 满足 $\forall x \in C, a^{\mathrm{T}}x \leq a^{\mathrm{T}}x_0$，那么集合

$$\left\{x \mid a^{\mathrm{T}}x = a^{\mathrm{T}}x_0\right\}$$

为 C 在边界点 x_0 处的支撑超平面。

定理

支撑超平面定理 如果 C 是凸集，则在 C 的任意边界点处都存在支撑超平面。

支撑超平面的几何直观：给定一个平面后，可把凸集边界上的任意一点当成支撑点将凸集放置在该平面上。从几何直观来理解凸优化的性质是非常重要的。

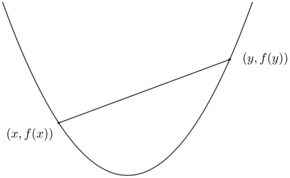
凸函数

定义

凸函数 若函数 f 的定义域 $\mathbf{dom} f$ 是凸集，且

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

对所有 $x, y \in \mathbf{dom} f, 0 \le 1$ 都成立，则称 f 是凸函数。如果对所有 $x, y \in \mathbf{dom} f, 0 \le 1$ 都有上式严格成立（处处小于），则称 f 是严格凸函数。



相应地，若 $-f$ 凸函数，则称 f 为凹函数。很多凸函数的性质都可以应用在凹函数上，因此我们主要讨论凸函数。

强凸函数

定义

强凸函数 若存在常数 $m>0$ 使得

$$g(x)=f(x)-\frac{m}{2}\|x\|^2$$

为凸函数，则称 f 是强凸函数。其中 m 为强凸参数，也称 f 为 m-强凸函数（曲线弯曲的更加厉害了，减去一个正定二次函数之后仍然是凸的）。

定义

强凸函数的等价定义 若存在常数 $m>0$ 使得对于任意 $x,y\in\mathbf{dom}f$ 以及 $\theta\in(0,1)$，有

$$f\big(\theta x+(1-\theta)y\big)\le\theta f(x)+(1-\theta)f(y)-\frac{m}{2}\theta(1-\theta)\|x-y\|^2,$$

则称 f 是强凸函数（可以看到 f 一定是严格凸函数）。

显然，强凸函数比一般凸函数具有更快的收敛速度。

强凸函数

定理

强凸函数解的唯一性 *若 f 为强凸函数且存在最小值，则 f 的最小值点唯一。*

证明.

反证法。设 $x \neq y$ 均为 f 的最小值点，取 $\theta \in (0,1)$，则

$$\begin{aligned} f\big(\theta x + (1-\theta)y\big) &\le \theta f(x) + (1-\theta) f(y) - \frac{m}{2}\theta(1-\theta)\|x-y\|^2 \\ &= f(x) - \frac{m}{2}\theta(1-\theta)\|x-y\|^2 \qquad \triangleright f(x) = f(y) \\ &< f(x), \end{aligned}$$

其中严格不等号成立是因为 $x \neq y$。这与 $f(x)$ 是最小值矛盾。 $\square$

凸函数判定定理

定理

凸函数判定定理 $f(x)$ 是凸函数当且仅当对任意的 $x \in \mathbf{dom} f, v \in \mathbb{R}^n, g : \mathbb{R} \to \mathbb{R}$,

$$g(t) = f(x + tv), \quad \mathbf{dom} g = \{t \mid x + tv \in \mathbf{dom} f\}$$

是凸函数（将其限制在直线上，然后判定对应的一维函数是否是凸的)。

证明.

（1）**必要性**。任取 $t_1, t_2 \in \mathbf{dom} g$ 以及 $\theta \in (0, 1)$，则

$$x + t_1 v \in \mathbf{dom} f, x + t_2 v \in \mathbf{dom} f,$$

由 $\mathbf{dom} f$ 是凸集可以得到

$$x + \big(\theta t_1 + (1 - \theta) t_2\big) v \in \mathbf{dom} f,$$

这说明 $\theta t_1 + (1 - \theta) t_2 \in \mathbf{dom} g$，因此 $\mathbf{dom} g$ 是凸集。同样地，带入 g 的定义可以证得 g 是凸函数。 □

凸函数判定定理

证明.

(2) **充分性**。对任意的 $x, y \in \mathbf{dom} f$ 以及 $\theta \in (0,1)$，取 $v = y - x, t_1 = 0, t_2 = 1$，则 $\theta \cdot 0 + (1-\theta) \cdot 1 \in \mathbf{dom} g$，所以 $x + (1-\theta)(y-x) = \theta x + (1-\theta) y \in \mathbf{dom} f$，这说明 $\mathbf{dom} f$ 是凸集。根据 $g(t)$ 的凸性，进一步有

$$\begin{aligned} g(1-\theta) &= g(\theta t_1 + (1-\theta) t_2) \\ &\leq \theta g(t_1) + (1-\theta) g(t_2) \\ &= \theta f(x) + (1-\theta) f(y). \end{aligned}$$

而等式左边有

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

因此 $f(x)$ 是凸函数。 □

有了凸函数判定定理，我们就可以结合定义和该定理来证明一个一般函数是否为凸函数（还有更多判定方法待介绍）。

凸函数示例

- **仿射函数**：$a^{\mathrm{T}}x+b, a, x \in \mathbb{R}^n$，$\langle A, X\rangle, A, X \in \mathbb{R}^{m\times n}$
- **指数函数**：$e^{ax}, a, x \in \mathbb{R}$
- **幂函数**：$x^\alpha, x>0$，当 $\alpha \geq 1$ 或 $\alpha \leq 0$ 是为凸函数
- **负熵**：$x\ln x, x>0$
- **范数**：所有向量与矩阵的范数都是凸函数（三角不等式）

利用凸函数判定定理来判定函数是否凸（示例）：

$$f(X)=-\ln\det(X) \text{ 是凸函数，} \mathbf{dom} f=\mathcal{S}^n_{++}\text{。}$$

任取 $X\succ 0$ 以及方向 $V\in\mathcal{S}^n$，则

$$\begin{aligned} g(t) = -\ln\det(X+tV) &= -\ln\det X - \ln\det(I+tX^{-1/2}VX^{1/2}) \\ &= -\ln\det X - \sum_{i=1}^{n}\ln\left(1+t\lambda_i\right), \end{aligned}$$

其中 $\lambda_i\geq 0$ 是 $X^{-1/2}VX^{1/2}$ 的第 i 个特征值。显然 $g(t)$ 是关于 t 的凸函数，因此 $f(X)$ 是凸函数。

利用导数判定凸性

定理

一阶条件 对于定义在凸集上的可微函数 f，f 是凸函数当且仅当

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

证明.

(1) **必要性**。对于任意的 $x, y \in \mathbf{dom} f$ 以及 $t \in (0, 1)$，

$$\begin{aligned} tf(y) + (1-t)f(x) \geq f(x + t(y-x)) \Rightarrow \\ f(y) - f(x) \geq \frac{f(x + t(y-x)) - f(x)}{t} \Rightarrow \\ t \to 0: f(y) - f(x) \geq \nabla f(x)^{\mathrm{T}}(y - x). \end{aligned}$$

$\square$

利用导数判定凸性

定理

一阶条件 对于定义在凸集上的可微函数 f，f 是凸函数当且仅当

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

证明.

（2）**充分性**。对于任意的 $x, y \in \mathbf{dom} f$ 以及 $t \in (0,1)$，定义 $z = tx + (1-t)y$，则

$$\begin{aligned} f(x) &\geq f(z) + \nabla f(z)^{\mathrm{T}}(x-z) \\ f(y) &\geq f(z) + \nabla f(z)^{\mathrm{T}}(y-z) \end{aligned}$$

可得到

$$tf(x) + (1-t)f(y) \geq f(z) + 0.$$

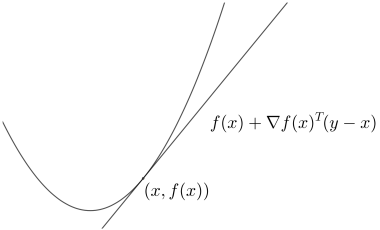
$\square$

利用导数判定凸性

定理

一阶条件 对于定义在凸集上的可微函数 f，f 是凸函数当且仅当

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



可微函数的任意一点处的一阶近似可以得到 f 的一个全局下界。

梯度单调性

定理

梯度单调性 设 f 是可微函数，则 f 为凸函数当且仅当 $\mathbf{dom} f$ 为凸集且 ∇f 为单调映射，即

$$\left(\nabla f(x) - \nabla f(y)\right)^{\mathrm{T}}(x-y) \geq 0, \quad \forall x, y \in \mathbf{dom} f.$$

证明.

（1）**必要性**。若 f 可微且为凸函数，则根据一阶条件有

$$f(y) \geq f(x) + \nabla f(x)^{\mathrm{T}}(y-x)$$
$$f(x) \geq f(y) + \nabla f(y)^{\mathrm{T}}(x-y),$$

相加即可。 □

梯度单调性

定理

梯度单调性 设 f 是可微函数，则 f 为凸函数当且仅当 $\mathbf{dom} f$ 为凸集且 ∇f 为单调映射，即

$$\big(\nabla f(x) - \nabla f(y)\big)^{\mathrm{T}}(x - y) \geq 0, \quad \forall x, y \in \mathbf{dom} f.$$

证明.

（2）**充分性**。构造一元辅助函数 $g(t) = f(x + t(y - x))$，则 $g'(t) = \nabla f\big(x + t(y - x)\big)^{\mathrm{T}}(y - x)$。根据 ∇f 的单调性可得 $g'(t) \geq g'(0), \forall t \geq 0$。因此

$$\begin{aligned} f(y) = g(1) &= g(0) + \int_0^1 g'(t)dt \\ &\geq g(0) + g'(0) = f(x) + \nabla f(x)^{\mathrm{T}}(y - x). \end{aligned}$$

$\square$

梯度单调性

严格凸函数和强凸函数也具有对应的**梯度单调性**。

定理

梯度单调性 设 f 是可微函数，$\mathbf{dom} f$ 为凸集，则

1. f 是 严格凸函数 当且仅当

$$\left(\nabla f(x) - \nabla f(y)\right)^{\mathrm{T}}(x - y) > 0, \quad \forall x, y \in \mathbf{dom} f.$$

2. f 是 m-强凸函数 当且仅当

$$\left(\nabla f(x) - \nabla f(y)\right)^{\mathrm{T}}(x - y) \geq m\|x - y\|^2, \quad \forall x, y \in \mathbf{dom} f.$$

利用二阶导数判定凸性

定理

二阶条件 设 f 是定义在凸集上的二阶连续可微函数，则 f 是凸函数当且仅当

$$\nabla^2 f(x) \succeq 0, \quad \forall x \in \mathbf{dom} f.$$

若正定，则 f 是严格凸函数。

证明.

(1) **必要性**。假设存在点 x 使得 $\nabla^2 f(x) \not\succeq 0$，即存在非零向量 $v \in \mathbb{R}^n$ 使得 $v^{\mathrm{T}} \nabla^2 f(x) v < 0$。在 $x + tv$ 处泰勒展开得

$$\frac{f(x+tv) - f(x) - t\nabla f(x)^{\mathrm{T}} v}{t^2} = \frac{1}{2} v^{\mathrm{T}} \nabla^2 f(x) v + o(1).$$

所以当 t 充分小的时候 $o(1) \to 0$，即
$f(x+tv) - f(x) - t\nabla f(x)^{\mathrm{T}} v < 0$。这与一阶条件的结论相矛盾。 □

利用二阶导数判定凸性

定理

二阶条件 设 f 是定义在凸集上的二阶连续可微函数，则 f 是凸函数当且仅当

$$\nabla^2 f(x) \succeq 0, \quad \forall x \in \mathbf{dom} f.$$

若正定，则 f 是严格凸函数。

证明.

（2）**充分性**。对任意的 $x, y \in \mathbf{dom} f$，根据泰勒展开可得

$$f(y) = f(x) + \nabla f(x)^{\mathrm{T}}(y-x) + \frac{1}{2}\underbrace{(y-x)^{\mathrm{T}}\nabla^2 f\big(x+t(y-x)\big)(y-x)}_{\text{半正定性可得本项}\geq 0},$$

其中 $t \in (0,1)$。这正是凸函数判定的一阶条件。 □

本证明稍加改造，即可得到严格凸函数的二阶条件的证明。

利用二阶导数判定凸性

示例：

- **二次函数**：$f(x)=\frac{1}{2}x^{\mathrm{T}}Px+q^{\mathrm{T}}x_r(P\in\mathcal{S}^n)$，显然

$$\nabla f(x)=Px+q,\quad \nabla^2 f(x)=P.$$

因此 f 是凸函数当且仅当 $P\succeq 0$。

- **最小二乘函数**：$f(x)=\frac{1}{2}\|Ax-b\|^2$，显然

$$\nabla f(x)=A^{\mathrm{T}}(Ax-b),\quad \nabla^2 f(x)=A^{\mathrm{T}}A.$$

所以对任意的 A 都有 f 是凸函数。

利用上方图判定凸性

定理

上方图和凸函数的关系 函数 $f(x)$ 是凸函数当且仅当其上方图 $\mathbf{epi}f$ 是凸集。

证明.

(1) **必要性**。取 x_1, x_2, t_1, t_2 使得 $f(x_1) \leq t_1, f(x_2) \leq t_2$。则 $f(\theta x_1 + (1-\theta)x_2) \leq \theta t_1 + (1-\theta)t_2$。这说明

$$\big(\theta x_1 + (1-\theta)x_2, \theta t_1 + (1-\theta)t_2\big) \in \mathbf{epi}f, \theta \in (0,1).$$

所以 $\mathbf{epi}f$ 为凸集。 □

利用上方图判定凸性

定理

上方图和凸函数的关系 函数 $f(x)$ 是凸函数当且仅当其上方图 $\mathbf{epi}f$ 是凸集。

证明.

（2）**充分性**。显然 $(x_1, f(x_1)) \in \mathbf{epi}f, (x_2, f(x_2)) \in \mathbf{epi}f$，因为 $\mathbf{epi}f$ 是凸集，所以

$$\big(\theta x_1 + (1-\theta)x_2, \theta f(x_1) + (1-\theta)f(x_2)\big) \in \mathbf{epi}f, \theta \in (0,1).$$

所以 $f\big(\theta x_1 + (1-\theta)x_2\big) \leq \theta f(x_1) + (1-\theta)f(x_2)$。即 f 是凸函数。 $\square$

保凸运算

判定一个函数是否是凸函数的方法：（1）凸函数判定定理（定义）；（2）对可微函数利用一阶条件和二阶条件；（3）判断 $\mathbf{epi}f$ 是否为凸集；（4）**由简单凸函数的保凸运算得到**（非负加权和、与仿射函数复合、逐点取最大值等）。

定理

保凸运算

- f 凸 $\Longrightarrow \forall \alpha \geq 0, \alpha f$ 凸
- f_1 和 f_2 凸 $\Longrightarrow f_1 + f_2$ 凸
- f 凸 $\Longrightarrow f(Ax+b)$ 凸
- $f_1, f_2, ..., f_m$ 凸 $\Longrightarrow f(x) = \max\{f_1(x), f_2(x), ..., f_m(x)\}$ 凸
- $\forall y \in \mathcal{A}, f(x,y)$ 关于 x 凸 $\Longrightarrow g(x) \triangleq \sup_{y\in\mathcal{A}} f(x,y)$ 凸

  proof sketch. $\mathbf{epi}g = \bigcap_{y\in\mathcal{A}} \mathbf{epi}f(\cdot,y)$ 作为多个凸集的交集仍然是凸集。

保凸运算

定理

保凸运算（续）

- $g : \mathbb{R}^n \to \mathbb{R}$ 凸 & $h : \mathbb{R} \to \mathbb{R}$ 凸且单调不减 $\Longrightarrow f(x) \triangleq h(g(x))$ 凸
- $g : \mathbb{R}^n \to \mathbb{R}$ 凹 & $h : \mathbb{R} \to \mathbb{R}$ 凸且单调不增 $\Longrightarrow f(x) \triangleq h(g(x))$ 凸
- 给定 $g : \mathbb{R}^n \to \mathbb{R}^k, h : \mathbb{R}^k \to \mathbb{R}$,

$$f(x) = h(g(x)) = h(g_1(x), ..., g_k(x)).$$

  g_i 凸 & h 凸且关于每个分量单调不减 $\Longrightarrow f$ 凸
  g_i 凹 & h 凸且关于每个分量单调不增 $\Longrightarrow f$ 凸

保凸运算

定理

保凸运算（续）

- $f(x, y)$ 是关于 (x, y) 的凸函数且 C 是凸集
  $\Longrightarrow g(x) \triangleq \inf_{y \in C} f(x, y)$ 凸

  <u>*proof sketch.*</u> 根据 g 的定义可得

$$f(x_i, y_i) \leq g(x_i) + \varepsilon, \forall i = 1, 2.$$

因此

$$\begin{aligned} g(\theta x_1 + (1-\theta)x_2) &= \inf_{y \in C} f(\theta x_1 + (1-\theta)x_2, y) \\ &\leq f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \quad \triangleright C\text{是凸集} \\ &\leq \theta f(x_1, y_1) + (1-\theta) f(x_2, y_2) \quad \triangleright f\text{是凸的} \\ &\leq \theta g(x_1) + (1-\theta) g(x_2) + \varepsilon. \end{aligned}$$

最后令 $\varepsilon \to 0$ 即可。

保凸运算

定理

保凸运算（续）

- 定义函数 $f : \mathbb{R}^n \to \mathbb{R}$ 的透视函数 $g : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$，

$$g(x,t) = tf\left(\frac{x}{t}\right), \quad \mathbf{dom}g = \left\{(x,t) \mid \frac{x}{t} \in \mathbf{dom}f, t > 0\right\}$$

若 f 是凸函数，则 g 是凸函数。

带入凸函数的定义即可证明。

保凸运算示例

与仿射函数复合保凸:

- 线性不等式的对数障碍函数

$$f(x) = -\sum_{i=1}^{m} \ln\left(b_i - a_i^{\mathrm{T}} x\right), \quad \mathbf{dom} f = \{x \mid a_i^{\mathrm{T}} x < b_i, i = 1, \ldots, m\}$$

  是凸函数。

- 仿射函数的任意范数 $f(x) = \|Ax + b\|$ 都是凸函数。

逐点取最大值保凸:

- 分段线性函数 $f(x) = \max_{i=1,\ldots,m}\left(a_i^{\mathrm{T}} x + b\right)$ 是凸函数。
- $x \in \mathbb{R}^n$ 的前 r 个最大分量之和

$$f(x) = \max\left\{\sum_{i=i_1}^{i_r} x_i \mid 1 \le i_1 < \ldots < i_r \le n\right\}$$

  是凸函数。

保凸运算示例

逐点取上确界保凸:

- 集合 C 的支撑函数
$$S_C(x) = \sup_{y \in C} y^{\mathrm{T}} x$$
是凸函数。
- 集合 C 的点到给定点 x 的最远距离
$$f(x) = \sup_{y \in C} \|x - y\|$$
是凸函数。
- 求对称阵的最大特征值
$$\lambda_{\max}(X) = \sup_{\|y\|_2 = 1} y^{\mathrm{T}} X y, \quad X \in \mathcal{S}^n$$
是凸函数。

保凸运算示例

凸函数之间的复合保凸：

- $g(x)$ 凸 $\Longrightarrow \exp(g(x))$ 凸。
- $g(x)$ 是正值凹函数 $\Longrightarrow \frac{1}{g(x)}$ 是凸函数。
- $g(x)$ 是正值凹函数 $\Longrightarrow \sum_{i=1}^{m} \ln(g_i(x))$ 是凹函数。
- $g_i(x)$ 凸 $\Longrightarrow \ln \sum_{i=1}^{m} \exp(g_i(x))$ 凸。

保凸运算示例

取下确界保凸:

- 对于函数 $f(x,y)=x^{\mathrm{T}}Ax+2x^{\mathrm{T}}By+y^{\mathrm{T}}Cy$，其中 $A\in\mathcal{S}^m, C\in\mathcal{S}^n, B\in\mathbb{R}^{m\times n}$（这其实是一个二次型，$(x,y)\to x'$）。其海瑟矩阵若满足

$$\begin{bmatrix} A & B \\ B^{\mathrm{T}} & C \end{bmatrix}\succeq 0, \qquad C\succ 0,$$

  则 $f(x,y)$ 为凸函数。对 y 求最小值得（求导[2]得 $y^*=-C^{-1}B^{\mathrm{T}}x$）

$$g(x)=\inf_y f(x,y)=x^{\mathrm{T}}(A-BC^{-1}B^{\mathrm{T}})x,$$

  根据 $f(x,y)$ 的海瑟矩阵满足的条件可以得到 $(A-BC^{-1}B^{\mathrm{T}})\succeq 0$，因此 g 是凸函数。
- 点 x 到凸集 $\mathcal{C}$ 的最短距离 $\mathrm{dist}(x,\mathcal{C})=\inf_{y\in\mathcal{C}}\|x-y\|$ 是凸函数。

[2] 矩阵微积分可参考 *The Matrix Cookbook*。

凸函数的性质

定理

连续性 设 $f:\mathbb{R}^n \to (-\infty,+\infty]$ 为凸函数。对任意点 $x_0 \in \mathbf{intdom} f$，有 f 在点 x_0 处连续。这里 $\mathbf{intdom} f$ 表示定义域 $\mathbf{dom} f$ 的内点（一个点是某个集合的内点 *iff* 存在以该点为中心的开球/邻域被包含于本集合）。

这表明凸函数在定义域的内点处连续，可以认为凸函数 “差不多是连续的”。

这一定理可直接得到如下推论：

定理

连续性 设 $f(x)$ 是凸函数，且 $\mathbf{dom} f$ 是开集，则 $f(x)$ 在 $\mathbf{dom} f$ 上是连续的。

凸函数的性质

定理

凸下水平集 设 $f(x)$ 是凸函数，则 $f(x)$ 所有的 α-下水平集 C_α 为凸集。

证明.

任取 $x_1, x_2 \in C_\alpha$，对任意的 $\theta \in (0,1)$ 有

$$\begin{aligned} f\big(\theta x_1 + (1-\theta)x_2\big) &\le \theta f(x_1) + (1-\theta)f(x_2) \\ &\le \theta\alpha + (1-\theta)\alpha = \alpha. \end{aligned}$$

□

本命题的逆命题不成立。

凸函数的性质

定理

二次下界 设 $f(x)$ 是参数为 m 的可微强凸函数，则

$$f(y) \geq f(x) + \nabla f(x)^{\mathrm{T}}(y - x) + \frac{m}{2}\|y - x\|^2, \quad \forall x, y \in \mathbf{dom} f.$$

证明.

结合强凸函数的定义和一阶条件可证。令 $g(x) = f(x) - \frac{m}{2}\|x\|^2$，则对 $g(x)$ 运用一阶条件可得

$$f(y) - \frac{m}{2}\|y\|^2 \geq f(x) - \frac{m}{2}\|x\|^2 + (\nabla f(x) - mx)^{\mathrm{T}}(y - x).$$

□

定理

下水平集有界 设 $f(x)$ 是可微强凸函数，则 f 的所有 α-下水平集都有界。(<u>*proof sketch*</u>：取 $x \in C_\alpha, y \in C_\beta$ 带入上式则 β 有下界。)

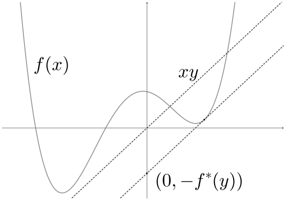
共轭函数

定义

共轭函数 任意（适当）函数 f 的共轭函数定义为

$$f^*(y) = \sup_{x \in \mathbf{dom} f} \{y^{\mathrm{T}} x - f(x)\}.$$

我们关心共轭函数是因为引入拉格朗日乘子法和对偶问题之后，一定会出现此式。$f^*(y)$ 是广义实值函数，可取到正无穷。因此我们规定 $\mathbf{dom} f^*$ 为使得 $f^*(y)$ 有限的 y 的集合。根据**保凸运算**第五条可知任意函数的共轭函数都是凸函数。



对固定的 y，$f^*(y)$ 的几何意义。

常见的共轭函数

定理

Fenchel **不等式** $f(x)+f^*(y)\geq x^{\mathrm{T}}y$（定义显然可得）。

常见函数的共轭函数：

- 对于二次函数 $f(x)=\frac{1}{2}x^{\mathrm{T}}Ax+b^{\mathrm{T}}x+c$，

  （1）若 f 强凸（$A\succ 0$）：

  $$f^*(y)=\frac{1}{2}(y-b)^{\mathrm{T}}A^{-1}(y-b)-c$$

  （带入 $x^*=-A^{-1}(y-b)$ 得到。）

  （2）若 f 一般凸（$A\succeq 0$）：

  $$f^*(y)=\frac{1}{2}(y-b)^{\mathrm{T}}A^{+}(y-b)-c,\quad \mathbf{dom}f^*=\mathcal{R}(A)+b.$$

  这里 A^+ 是 A 的 Moore-Penrose 逆，$\mathcal{R}(A)=\{Ax\mid x\in\mathbf{dom}f\}$ 是 A 的像空间。

常见的共轭函数

- （凸集的示性函数）给定凸集 C，其示性函数为

$$I_C(x) = \begin{cases} 0 & x \in C \\ +\infty & x \notin C. \end{cases}$$

则

$$I_C^*(y) = \sup_x \{y^{\mathrm{T}}x - I_C(x)\} = \sup_{x \in C} y^{\mathrm{T}}x.$$

又称 $I_C^*(y)$ 为凸集 C 的支撑函数。

常见的共轭函数

- （范数）范数的共轭函数为其**单位对偶范数球**的示性函数，即若 $f(x)=\|x\|$，则

$$f^*(y)=\begin{cases}0 & \|y\|_*\leq 1\\ +\infty & \|y\|_*>1,\end{cases}$$

其中对偶范数 $\|y\|_*\triangleq\sup_{\|x\|\leq 1}x^{\mathrm{T}}y$。
分类讨论：
(*1*) 若 $\|y\|_*\leq 1$，则结合柯西不等式可得 $y^{\mathrm{T}}x\leq\|x\|$ 对任意 x 成立（在 $x=0$ 时取等）。从而 $\sup_{x\in\mathbf{dom}\,f}\{y^{\mathrm{T}}x=-\|x\|\}=0$。
(*2*) 若 $\|y\|_*>1$，则至少存在一个 x 使得 $\|x\|\leq 1$ 且 $x^{\mathrm{T}}y>1$。从而对任意 $t>0$ 有

$$f^*(y)\geq y^{\mathrm{T}}(tx)-\|tx\|=t(y^{\mathrm{T}}x-\|x\|).$$

显然 $t\to\infty$ 时右端 $\to\infty$。

二次共轭函数

定义

二次共轭函数 任意函数 f 的二次共轭函数定义为

$$f^{**}(x) = \sup_{y \in \mathbf{dom} f^*} \{x^{\mathrm{T}} y - f^*(y)\}.$$

根据闭函数之间简单运算的保闭性以及任意函数的共轭函数的凸性可知 $f^{**}(x)$ 恒为**闭凸函数**。由 Fenchel 不等式可知 $\forall x$,

$$f^{**}(x) \le \sup_{y \in \mathbf{dom} f^*} \{x^{\mathrm{T}} y + f(x) - x^{\mathrm{T}} y\} = f(x).$$

等价地，$\mathbf{epi} f \subseteq \mathbf{epi} f^{**}$。

二次共轭函数

定理

二次共轭函数与原函数的等价性 若 f 为闭凸函数，则

$$f^{**}(x) = f(x), \forall x.$$

等价地，$\mathbf{epi} f^{**} = \mathbf{epi} f$。

证明.

因为对任意 x 有 $f^{**}(x) \le f(x)$，所以只要同时有 $\forall x, f^{**}(x) \ge f(x)$ 即可得证。采用反证法：假设存在 x 使得 $(x, f^{**}(x)) \notin \mathbf{epi} f$，则必然存在严格分割超平面使得

$$\begin{bmatrix} a \\ b \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} z - x \\ s - f^{**}(x) \end{bmatrix} \le c < 0, \quad \forall (z, s) \in \mathbf{epi} f$$

且 $a \in \mathbb{R}^n, b, c \in \mathbb{R}, b \le 0$（$b > 0$ 则 $s \to \infty$ 时矛盾）。这个式子由 $\begin{cases} a^{\mathrm{T}} x < b & \forall x \in C \triangleq \mathbf{epi} f \\ a^{\mathrm{T}} x > b \end{cases}$ 相减得到。 $\square$

二次共轭函数

定理

二次共轭函数与原函数的等价性 若 f 为闭凸函数，则

$$f^{**}(x) = f(x), \forall x.$$

等价地，$\mathbf{epi} f^{**} = \mathbf{epi} f$。

证明.

接下来对 b 分类讨论。
若 $b < 0$，取 $s = f(z)$ 并用 y 替换 $-\frac{a}{b}$ 代入上述不等式可得

$$f^*(y) - y^{\mathrm{T}}x + f^{**}(x) \leq -\frac{c}{b} < 0$$

这与 Fenchel 不等式矛盾。 $\square$

二次共轭函数

定理

二次共轭函数与原函数的等价性 若 f 为闭凸函数，则

$$f^{**}(x) = f(x), \forall x.$$

等价地，$\mathbf{epi} f^{**} = \mathbf{epi} f$。

证明.

若 $b = 0$，取 $\hat{y} \in \mathbf{dom} f^*$ 并给 $\begin{bmatrix} a \\ b \end{bmatrix}$ 加上一个 $\begin{bmatrix} \hat{y} \\ -1 \end{bmatrix}$ 的 ε 倍（$\varepsilon > 0$），得

$$\begin{bmatrix} a + \varepsilon \hat{y} \\ -\varepsilon \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} z - x \\ s - f^{**}(x) \end{bmatrix} \le c + \varepsilon\big(f^*(\hat{y}) - \hat{y}^{\mathrm{T}} x + f^{**}(x)\big) < 0,$$

这化归成了 $b < 0$ 的情况，遵循相同的步骤，从而得出矛盾。 $\square$

总结

本 slide 主要讨论了以下内容：

- 向量范数和矩阵范数的定义与性质；
- 导数与二阶导数的定义，对应版本的泰勒展开；
- 梯度 *Lipschitz* 连续及其性质；
- 矩阵变量函数的导数的定义与计算；
- 广义实值函数与闭函数；
- 一些几何概念：凸集、凸组合、凸包、凸锥、超平面等；
- 凸函数以及强凸函数的定义与性质；
- 凸函数的四类判定方法；
- 共轭函数与二次共轭函数。