最优化理论基础（中）

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

2021 年 7 月 11 日

基础数学理论请参考http://hliangzhao.me/math/math.pdf，《最优化理论基础（上)》参见http://hliangzhao.me/math/opt2.pdf。

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 $0 \leq \theta \leq 1$ 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

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

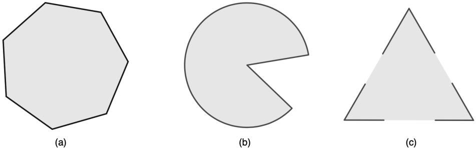
凸集

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

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

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 $x_1, \ldots, x_k$， $\forall \theta_i \geq 0, \sum_i \theta_i = 1$，形如

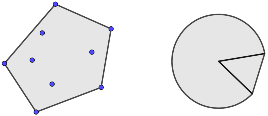
$$x = \sum_{i=1}^{k} \theta_i x_i$$

的点称为 $x_1, \ldots, x_k$ 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

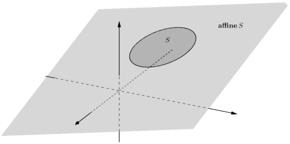
去掉凸组合的定义中的 $\{\theta_i\}_{\forall i} \geq 0$ 的限制，即可得到仿射包的概念。

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**。

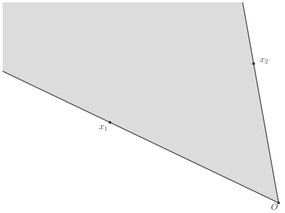
锥组合与凸锥

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。



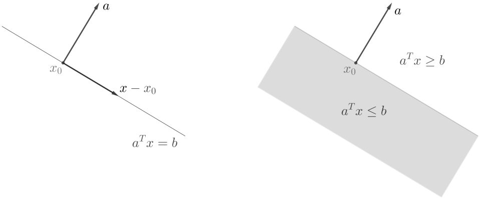
凸锥。

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

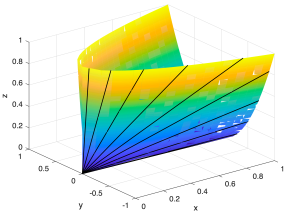
多面体和半正定锥:

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

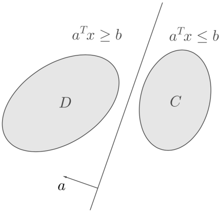
超平面可以分离不相交的凸集。

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

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

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，我们有严格分类定理。

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。

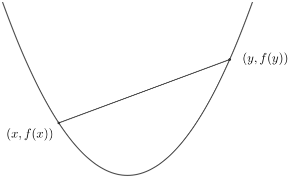
凸函数

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

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

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。 □

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 $x \in \textbf{dom} f, v \in \mathbb{R}^n, g : \mathbb{R} \to \mathbb{R}$,

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

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 $t_1, t_2 \in \textbf{dom} g$ 以及 $\theta \in (0, 1)$，则

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

由 $\textbf{dom} f$ 是凸集可以得到

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

这说明 $\theta t_1 + (1 - \theta) t_2 \in \textbf{dom} g$，因此 $\textbf{dom} g$ 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

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

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）：

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

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（重述）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

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

可得到

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

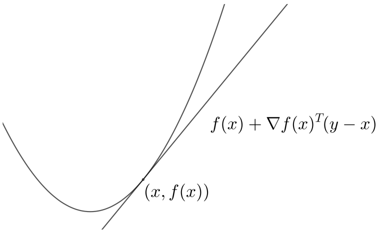
$\square$

利用导数判定凸性

定理

一阶条件（重述）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

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

相加即可。 □

梯度单调性

定理

梯度单调性（重述）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.

（2）**充分性**。构造一元辅助函数 $g(t) = f(x + t(y-x))$，则 $g'(t) = \nabla f\left(x + t(y-x)\right)^{\mathrm{T}}(y-x)$。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（重述）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。 $\square$

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 $\textbf{epi}f$ 是凸集。

证明.

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

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

所以 $\textbf{epi}f$ 为凸集。 □

利用上方图判定凸性

定理

上方图和凸函数的关系（重述）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。 □