第十章 优化基础

第 30 讲 凸函数

黄定江

DaSE @ ECNU

djhuang@dase.ecnu.edu.cn

30.1.1 凸函数

在给出凸函数的定义前，我们需要对广义实值函数和适当函数进行界定：

定义 1

（广义实值函数）令 $\overline{\mathbb{R}} \stackrel{\text{def}}{=} \mathbb{R} \cup \{\pm\infty\}$ 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.

适当函数是一类很重要的广义实值函数，很多最优化理论都是建立在适当函数之上的.

定义 2

（适当函数）给定广义实值函数 f 和非空集合 $\mathcal{X}$. 如果存在 $\boldsymbol{x} \in \mathcal{X}$ 使得 $f(\boldsymbol{x}) < +\infty$，并且对任意的 $\boldsymbol{x} \in \mathcal{X}$，都有 $f(\boldsymbol{x}) > -\infty$，那么称函数 f 关于集合 $\mathcal{X}$ 是适当的.

注：对最优化问题 $\min f(x)$，适当函数可以帮助去掉一些我们不感兴趣的函数，从

凸函数

定义 3

设函数 $f: \mathbb{R}^n \to \mathbb{R}$ 为适当函数，如果 $\mathbf{dom}\, f$ 是凸集，且

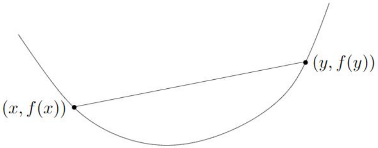
$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})$$

对所有 $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{dom}\, f$, $0 \leq \theta \leq 1$ 都成立，则称 f 是**凸函数**。

注： 如果对所有的 $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{dom}\, f, \boldsymbol{x} \neq \boldsymbol{y}, 0 < \theta < 1$，上述不等式严格成立，则称函数 f 是**严格凸函数**。

凸函数的几何意义

从几何意义上看，上述不等式意味着连接凸函数图像上任意两点的线段都在函数图像上方。



强凸函数

除严格凸函数以外，还可以定义如下强凸函数：

定义 4

(强凸函数) 若存在常数 $\sigma > 0$，使得

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

为凸函数，则称 $f(x)$ 为强凸函数，其中 σ 为强凸参数. 为了方便我们也称 $f(x)$ 为 σ-强凸函数.

从上述定义可以看出，强凸函数减去一个正定二次函数仍然是凸的。

强凸函数

通过直接对 $g(x)=f(x)-\frac{\sigma}{2}\|x\|^2$ 应用凸函数的定义, 我们可得到另一个常用的强凸函数定义.

定义 5

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

$$f(\theta x+(1-\theta y))\leqslant\theta f(x)+(1-\theta)f(y)-\frac{\sigma}{2}\theta(1-\theta)\|x-y\|^2,$$

则称 $f(x)$ 为强凸函数, 其中 σ 为强凸参数.

从上述定义可以看出强凸函数一定是严格凸函数，当 $\sigma=0$ 时退化成凸函数．无论从哪个定义出发，容易看出和凸函数相比，强凸函数有更好的性质。

强凸函数的性质

定理 1

设 f 为强凸函数且存在最小值, 则 f 的最小值点唯一.

Proof.

采用反证法. 设 $x \neq y$ 均为 f 的最小值点, 根据强凸函数的等价定义, 取 $\theta \in (0,1)$, 则有

$$\begin{aligned} f(\theta x + (1-\theta)y) &\leqslant \theta f(x) + (1-\theta)f(y) - \frac{\sigma}{2}\theta(1-\theta)\|x-y\|^2 \\ &= f(x) - \frac{\sigma}{2}\theta(1-\theta)\|x-y\|^2 \\ &< f(x), \end{aligned}$$

其中严格不等号成立是因为 $x \neq y$. 这显然和 $f(x)$ 为最小值矛盾, 得证. □

30.1.2 凹函数和凸 (凹) 函数示例

- 如果函数 $-f$ 是凸函数，则称函数 f 是**凹函数**。

通过定义，可以很容易判断出许多凸函数和凹函数的例子。首先考虑 $\mathbb{R}$ 上的一些函数，其自变量为 x。

- 指数函数。对任意 $a \in R$, 函数 e^{ax} 在 $\mathbb{R}$ 上是凸的。
- 幂函数。当 $a \geq 1$ 或 $a \leq 0$ 时，x^a 是在 $\mathbb{R}_{++}$ 上的凸函数，当 $0 \leq a \leq 1$ 时 x^a 是在 $\mathbb{R}_{++}$ 上的凹函数。
- 绝对值幂函数。当 $p \geq 1$ 时，函数 $|x|^p$ 在 $\mathbb{R}$ 上是凸函数。
- 对数函数。函数 $\log x$ 在 $\mathbb{R}_{++}$ 上的凹函数。
- 负熵。函数 $x \log x$ 在其定义域上是凸函数。

凸 (凹) 函数示例

下面我们给出 $\mathbb{R}^n$ 上的一些例子。

- 仿射函数：$\boldsymbol{a}^T\boldsymbol{x}+b$, 其中 $\boldsymbol{a},\boldsymbol{x}\in\mathbb{R}^n$ 是向量。它是 $\mathbb{R}^n$ 上的凸（凹）函数。
- 几何平均是其定义域上的凹函数。
- 所有范数都是凸函数（向量和矩阵版本），这是由于范数满足三角不等式。

30.1.3 凸性判定

凸函数的一个最基本的判定方式是：先将其限制在任意直线上，然后判断对应的一维函数是否是凸的。即如下判定定理：

定理 2

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

$$g(t) = f(\boldsymbol{x} + t\boldsymbol{v}), \textbf{dom}\ g = \{t | \boldsymbol{x} + t\boldsymbol{v} \in \textbf{dom}\ f\}$$

是凸函数。

凸函数判定示例

下面的例子说明如何利用上述定理判断函数的凸性。

例 1

$f(X) = -\ln\det \boldsymbol{X}$ 是凸函数，其中 $\mathbf{dom}\, f = \mathcal{S}^n_{++}$. 任取 $\boldsymbol{X} \succ 0$ 以及方向 $\boldsymbol{V} \in \mathcal{S}^n$，将 f 限制在直线 $\boldsymbol{X} + t\boldsymbol{V}$（$t$ 满足 $\boldsymbol{X} + t\boldsymbol{V} \succ 0$）上，考虑函数 $g(t) = -\ln\det(\boldsymbol{X} + t\boldsymbol{V})$. 那么

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

其中 λ_i 是 $\boldsymbol{X}^{-1/2}\boldsymbol{V}\boldsymbol{X}^{-1/2}$ 的第 i 个特征值。对每个 $\boldsymbol{X} \succ 0$ 以及方向 $\boldsymbol{V}$，易知 g 关于 t 是凸的，因此 f 是凸的。

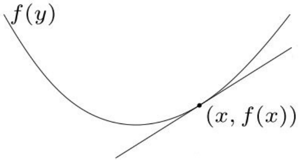
30.1.4 一阶条件

定理 3

对于定义在凸集上的可微函数 f, f 是凸函数当且仅当

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

该定理说明可微凸函数 f 的图形始终在其任一点处切线的上方，下图描述了上述不等式：



一阶条件

定理 4

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

$$(\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}))^T(\boldsymbol{x}-\boldsymbol{y}) \geq 0,\ \forall \boldsymbol{x}, \boldsymbol{y} \in \textbf{dom}\, f.$$

推论 1

f 是严格凸函数当且仅当 $(\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}))^T(\boldsymbol{x}-\boldsymbol{y}) > 0,\ \forall \boldsymbol{x}, \boldsymbol{y} \in \textbf{dom}\, f.$

f 是凹函数当且仅当 $(\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}))^T(\boldsymbol{x}-\boldsymbol{y}) \leq 0,\ \forall \boldsymbol{x}, \boldsymbol{y} \in \textbf{dom}\, f.$

30.1.5 二阶条件

进一步，如果函数二阶连续可微，可以得到下面的二阶条件：

定理 5

(二阶条件) 设 f 为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

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

如果 $\nabla^2 f(\boldsymbol{x}) \succ 0, \ \forall \boldsymbol{x} \in \textbf{dom}\ f$，则 f 是严格凸函数。

条件 $\nabla^2 f(\boldsymbol{x}) \succeq 0$ 从几何上可以理解为函数图像在点 $\boldsymbol{x}$ 处有具有正 (向上) 的曲率。

二阶条件

例 2

(1) 考虑二次函数 $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}+\boldsymbol{q}^T\boldsymbol{x}+r(\boldsymbol{P}\in S^n)$，容易计算出其梯度与海瑟矩阵分别为

$$\nabla f(\boldsymbol{x})=\boldsymbol{P}\boldsymbol{x}+\boldsymbol{q},\nabla^2 f(\boldsymbol{x})=\boldsymbol{P}.$$

那么，f 是凸函数当且仅当 $\boldsymbol{P}\succeq 0$.

(2) 考虑最小二乘函数 $f(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2$，其梯度与海瑟矩阵分别为

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

注意到 $\boldsymbol{A}^T\boldsymbol{A}$ 恒为半正定矩阵，因此对任意 $\boldsymbol{A}$，$\boldsymbol{f}$ 都是凸函数。

30.1.6 上镜图

上镜图是从集合的角度来描述一个函数的具体性质. 我们有上镜图定义如下:

定义 6

(上镜图) 对于广义实值函数 $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$,

$$\textit{epi}\ f = \left\{(x, t) \in \mathbb{R}^{n+1} \mid f(x) \leqslant t\right\}$$

称为 f 的上镜图。

上镜图与凸函数判定

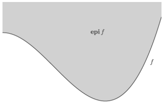


图 1: 函数 f 和其上镜图 epi f

上镜图将函数和集合建立了联系, f 的很多性质都可以在 epi f 上得到体现. 实际上，可以使用上镜图 epi f 来判断 f 的凸性。我们有如下定理：

定理 6

函数 $f(x)$ 为凸函数当且仅当其上镜图 epi f 是凸集.

函数的凸性可以由简单的凸函数通过一些保凸的运算得到。了解这些运算，为复杂函数凸性的判定提供了极大的便利，包括：

- 非负加权求和
- 复合仿射映射
- 逐点最大和逐点上确界
- 复合
- 最小化
- 透视函数

30.2.1 非负加权求和 & 复合仿射映射

定理 7

- **非负乘积**：若 f 是凸函数，$\alpha \geq 0$，则 αf 是凸函数。
- **求和**：若 f_1, f_2 都是凸函数，则 $f_1 + f_2$ 是凸函数。
- **复合仿射映射**：若 f 是凸函数，则 $f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$ 是凸函数。

非负加权求和 & 复合仿射映射

例 3

利用与仿射函数的复合函数保凸，可以证明：

- 线性不等式的对数障碍函数：

$$f(\boldsymbol{x}) = -\sum_{i=1}^{m} \log(b_i - \boldsymbol{a}_i^T \boldsymbol{x}), \qquad \mathbf{dom}\, f = \{\boldsymbol{x} | \boldsymbol{a}_i^T \boldsymbol{x} \leq b_i, i = 1, \cdots, m\}$$

  是凸函数。

- 仿射函数的 (任意) 范数 :$f(x) = \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|$ 都是凸函数。

30.2.2 逐点最大

定理 8

- **逐点取最大值：** 若 $f_1, f_2, \ldots, f_m$ 是凸函数，则 $f(\boldsymbol{x}) = \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})\}$ 是凸函数。

例 4

- 分段线性函数：$f(\boldsymbol{x}) = \max_{i=1,\cdots,m}\{\boldsymbol{a}_i^T\boldsymbol{x} + b_i\}$ 是凸函数。
- $\boldsymbol{x} \in \mathbb{R}^n$ 最大的 r 个分量之和：$f(\boldsymbol{x}) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$ 是凸函数 ($x_{[i]}$ 是 $\boldsymbol{x}$ 的从大到小排列的第 i 个分量)。

  证明：

$$f(\boldsymbol{x}) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} | 1 \le i_1 < i_2 \cdots < i_r \le n\}$$

30.2.3 逐点上确界

定理 9

如果对于任意 $\boldsymbol{y} \in \mathcal{A}$，函数 $f(\boldsymbol{x}, \boldsymbol{y})$ 关于 $\boldsymbol{x}$ 都是凸函数，则

$$g(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in \mathcal{A}} f(\boldsymbol{x}, \boldsymbol{y})$$

关于 $\boldsymbol{x}$ 亦是凸函数。

例 5

- 集合 C 的支撑函数:$S_C(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in C} \boldsymbol{y}^T \boldsymbol{x}$ 是凸函数。
- 集合 C 的最远点的距离：$f(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in C} \|\boldsymbol{x} - \boldsymbol{y}\|$ 是凸函数。
- n 阶对称矩阵的最大特征值: $\lambda_{max}(\boldsymbol{X}) = \sup_{\|\boldsymbol{y}\|_2 = 1} \boldsymbol{y}^T \boldsymbol{X} \boldsymbol{y}, \boldsymbol{X} \in S^n$ 是凸函数。

30.2.4 标量函数复合

定理 10

$g:\mathbb{R}^n\to\mathbb{R}$，$h:\mathbb{R}\to\mathbb{R}$，$f(\boldsymbol{x})=h(g(\boldsymbol{x}))$。

若 $\begin{cases} g \text{ 是凸函数，} h \text{ 是凸函数且非减} \\ g \text{ 是凹函数，} h \text{ 是凸函数且非增} \end{cases}$，则 f 是凸函数。

例 6

- 若 g 是凸函数，则 $\exp g(\boldsymbol{x})$ 是凸函数；
- 若 g 是正值凹函数，则 $1/g(\boldsymbol{x})$ 是凸函数；
- 若 g_i 是正值凹函数，则 $\sum_{i=1}^m \ln(g_i(\boldsymbol{x}))$ 是凹函数；

30.2.5 向量函数复合

定理 11

$g:\mathbb{R}^n\to\mathbb{R}^k, h:\mathbb{R}^k\to\mathbb{R}$，$f(\boldsymbol{x})=h(g(\boldsymbol{x}))=h(g_1(\boldsymbol{x}),\cdots,g_k(\boldsymbol{x}))$。

若 $\begin{cases} g_i \text{ 是凸的，} & h \text{ 是凸的且对每个分量 } h \text{ 非减} \\ g_i \text{ 是凹的，} & h \text{ 是凸的且对每个分量 } h \text{ 非增} \end{cases}$，则 f 是凸函数。

例 7

函数 $h(\boldsymbol{z})=\log(\sum_{i=1}^k e^{z_i})$ 是凸函数且在每一维分量上非减，因此只要 g_i 是凸函数，$\log(\sum_{i=1}^k e^{g_i})$ 就是凸函数。

30.2.6 最小化

定理 12

若 $f(x,y)$ 关于 (x,y) 整体是凸函数，C 是凸集，则

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

是凸函数.

例 8

点 $\boldsymbol{x}$ 到凸集 S 的距离：$\mathrm{dist}(\boldsymbol{x}, S) = \inf_{\boldsymbol{y} \in S} \|\boldsymbol{x} - \boldsymbol{y}\|$ 是凸函数.

30.2.7 透视函数

定理 13

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

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

若 f 是凸函数, 则 g 是凸函数.

透视函数

例 9

设 $f:\mathbb{R}^m\to\mathbb{R}$ 是凸函数，$\boldsymbol{A}\in\mathbb{R}^{m\times n}, \boldsymbol{b}\in\mathbb{R}^m, \boldsymbol{c}\in\mathbb{R}^n, d\in\mathbb{R}$。定义

$$g(\boldsymbol{x})=(\boldsymbol{c}^T\boldsymbol{x}+d)f((\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})/(\boldsymbol{c}^T\boldsymbol{x}+d))$$

其定义域为

$$\mathbf{dom}\, g=\{\boldsymbol{x}|\boldsymbol{c}^T\boldsymbol{x}+d>0,(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})/(\boldsymbol{c}^T\boldsymbol{x}+d)\in\mathrm{dom}\, f\}$$

则 g 是凸函数。

30.3.1 连续性

凸函数不一定是连续函数，但下面这个定理说明凸函数在定义域中内点处是连续的.

定理 14

设 $f: \mathbb{R}^n \to (-\infty, +\infty]$ 为凸函数. 对任意点 $x_0 \in \operatorname{int}\operatorname{dom} f$, 有 f 在点 x_0 处连续. 这里 $\operatorname{int}\operatorname{dom} f$ 表示定义域 $\operatorname{dom} f$ 的内点.

连续性

定理14表明凸函数“差不多”是连续的，它的一个直接推论为：

推论 2

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

Proof.

由于开集中所有的点都为内点，利用定理 14 可直接得到结论. □

注：凸函数在定义域的边界上可能不连续. 一个例子为：

$$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x = 0 \end{cases}$$

其中 $\mathrm{dom}\, f = (-\infty, 0]$. 容易证明 $f(x)$ 是凸函数，但其在点 $x = 0$ 处不连续。

30.3.2 凸下水平集

下水平集是描述实值函数取值情况的一个重要概念. 为此有如下定义:

定义 7

(α-下水平集) 对于广义实值函数 $f:\mathbb{R}^n \to \overline{\mathbb{R}}$,

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

称为 f 的 α-下水平集.

凸下水平集

凸函数的所有下水平集都为凸集, 即有如下结果:

定理 15

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

Proof.

任取 $x_1, x_2 \in C_\alpha$, 对任意的 $\theta \in (0,1)$, 根据 $f(x)$ 的凸性我们有

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

这说明 C_α 是凸集. □

30.3.3 二次下界

强凸函数具有二次下界的性质

定理 16

（二次下界）定义在凸集 K 上的可微函数 $f: K \to \mathbb{R}$ 被称为关于范数 $\|\cdot\|$ 是 σ-强凸的，则如下不等式成立

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{\sigma}{2} \|\boldsymbol{y} - \boldsymbol{x}\|^2 .$$

二次下界

Proof.

由强凸函数的定义，$g(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{\sigma}{2}\|\boldsymbol{x}\|^2$ 是凸函数，根据凸函数的一阶条件可知

$$g(\boldsymbol{y}) \geq g(\boldsymbol{x}) + \nabla g(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}),$$

即

$$\begin{aligned} f(\boldsymbol{y}) &\geq f(\boldsymbol{x}) - \frac{\sigma}{2}\|\boldsymbol{x}\|^2 + \frac{\sigma}{2}\|\boldsymbol{y}\|^2 + (\nabla f(\boldsymbol{x}) - \sigma \boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) && (1) \\ &= f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{\sigma}{2}\|\boldsymbol{y} - \boldsymbol{x}\|^2. && (2) \end{aligned}$$

□

30.4.1 共轭函数

定义 8

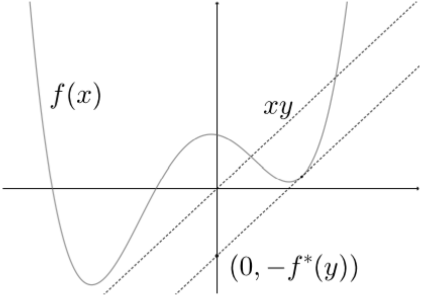
任一适当函数 f 的**共轭函数**定义为：

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

- 与 f 的性质无关，f^* 始终是凸函数

共轭函数

下图展示了对于固定的 y，$f^*(y)$ 的几何意义。



例 10

- 负对数函数 $f(x) = -\log x$

$$
\begin{aligned}
f^*(y) &= \sup_{x>0}(xy + \log x) \\
&= \begin{cases} -1 - \log(-y) & y < 0 \\ \infty & \mathtt{otherwise} \end{cases}
\end{aligned}
$$

30.4.2 共轭函数性质

- Fenchel 不等式：从共轭函数的定义可知，对于任意 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 有

$$f(\boldsymbol{x}) + f^*(\boldsymbol{y}) \geq \boldsymbol{x}^T \boldsymbol{y}$$

- 可微函数：设 $f(\boldsymbol{x})$ 为凸函数, 且可微, 对于 $\boldsymbol{z} \in \mathbb{R}^n$, 若

$$\boldsymbol{y} = \nabla f(\boldsymbol{z})$$

则 $f^*(\boldsymbol{y}) = \boldsymbol{z}^T \nabla f(\boldsymbol{z}) - f(\boldsymbol{z})$。

共轭函数性质

- 仿射变换：设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 非奇异，$\boldsymbol{b} \in \mathbb{R}^n$，则函数 $g(\boldsymbol{x}) = f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$ 的共轭函数为

$$g^*(\boldsymbol{y}) = f^*(\boldsymbol{A}^{-T}\boldsymbol{y}) - \boldsymbol{b}^T\boldsymbol{A}^{-T}\boldsymbol{y},$$

  其定义域为 $\mathbf{dom}\, g^* = \boldsymbol{A}^T \mathbf{dom}\, f^*$。

- 独立函数（各个函数具有不同的变量）的和：如果函数 $f(\boldsymbol{u}, \boldsymbol{v}) = f_1(\boldsymbol{u}) + f_2(\boldsymbol{v})$，其中 f_1 和 f_2 是凸函数，且共轭函数分别为 f_1^* 和 f_2^*，则

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

  换言之，独立函数的和的共轭函数是各个凸函数的共轭函数的和。

30.4.3 机器学习中常见共轭函数

机器学习中常涉及优化问题。在优化问题的求解时，常需要将优化问题转化为其对偶问题。这便涉及到对偶函数的计算。实际上对偶函数的计算可转化为共轭函数的计算，下面列出其中一些常见的例子。

机器学习中常见共轭函数

例 11

（二次函数） 考虑二次函数 $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}^T\boldsymbol{x}+c.$

(1) 强凸情形（$\boldsymbol{A}\succ 0$）：

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

(2) 一般凸情形（$\boldsymbol{A}\succeq 0$）：

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

这里 $\mathcal{R}(\boldsymbol{A})$ 为 $\boldsymbol{A}$ 的像空间.

机器学习中常见共轭函数

例 12

（任意范数） 给定任意范数，则 $f = \|\boldsymbol{x}\|$ 的共轭函数为

$$f_0^*(\boldsymbol{y}) = \begin{cases} 0 & \|\boldsymbol{y}\|_* \le 1 \\ \infty & \text{其他情况} \end{cases}$$

可以看出此函数是对偶范数单位球的示性函数。

机器学习中常见共轭函数

例 13

（负熵函数） 负熵函数 $f_0(\boldsymbol{x}) = \sum_{i=1}^n x_i \log x_i$ 的共轭函数为

$$f_0^*(\boldsymbol{y}) = \sum_{i=1}^n e^{y_i - 1}$$

这可以由标量负熵函数的共轭函数和共轭函数的性质得出。

30.5.1 次梯度

定义 9

（次梯度）设 f 为适当凸函数，x 为定义域 $\mathrm{dom}\, f$ 中的一点. 若向量 $g \in \mathbb{R}^n$ 满足

$$f(y) \geqslant f(x) + g^{\mathrm{T}}(y - x), \quad \forall y \in \mathrm{dom}\, f,$$

则称 g 为函数 f 在点 x 处的一个次梯度. 进一步地，称集合

$$\partial f(x) = \{g \mid g \in \mathbb{R}^n, f(y) \geqslant f(x) + g^{\mathrm{T}}(y - x), \forall y \in \mathrm{dom}\, f\}$$

为 f 在点 x 处的次微分.

次梯度几何意义

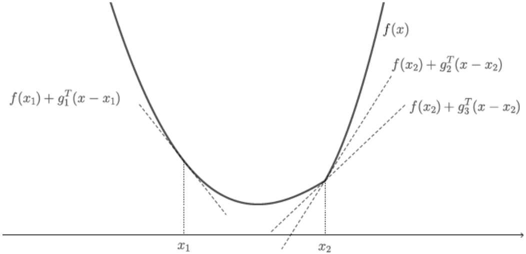


图 2: 虚线表示次梯度对应的切线（一维）

30.5.2 次梯度的存在性

一个问题自然就是：次梯度在什么条件下是存在的? 实际上对一般凸函数 f 而言, f 未必在所有的点处都存在次梯度. 但对于定义域中的内点, f 在其上的次梯度总是存在的.

定理 17

(次梯度存在性) 设 f 为凸函数, $\mathrm{dom}\, f$ 为其定义域. 如果 $x \in \mathrm{int\,dom}\, f$, 则 $\partial f(x)$ 是非空的, 其中 $\mathrm{int\,dom}\, f$ 的含义是集合 $\mathrm{dom}\, f$ 的所有内点.

Proof.

考虑 $f(x)$ 的上方图 epi f. 由于 $(x, f(x))$ 是 epi f 边界上的点, 且 epi f 为凸集, 根据支撑超平面定理, 存在 $a \in \mathbb{R}^n, b \in \mathbb{R}$ 使得:

$$\begin{bmatrix} a \\ b \end{bmatrix}^{\mathrm{T}} \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leqslant 0, \quad \forall (y, t) \in \mathbf{epi} f.$$

即

$$a^{\mathrm{T}}(y - x) \leqslant b(f(x) - t). \tag{3}$$

我们断言 $b < 0$. 这是因为根据 t 的任意性, 在式(3)中令 $t \to +\infty$, 可以得知式(3)成立的必要条件是 $b \leqslant 0$; □

Proof.

（续）同时由于 x 是内点，因此当取 $y = x + \varepsilon a \in \operatorname{dom} f, \varepsilon > 0$ 时，$b = 0$ 不能使得式(3)成立. 于是令 $g = -\frac{a}{b}$，则对任意 $y \in \operatorname{dom} f$，我们有

$$g^{\mathrm{T}}(y-x) = \frac{a^{\mathrm{T}}(y-x)}{-b} \leqslant -(f(x) - f(y)),$$

整理得

$$f(y) \geqslant f(x) + g^{\mathrm{T}}(y-x).$$

这说明 g 是 f 在点 x 处的次梯度。 □

例 14

(ℓ_2 范数的次微分) 设 $f(x)=\|x\|_2$, 则 $f(x)$ 在点 $x=0$ 处不可微, 我们求其在该点处的次梯度. 注意到对任意的 g 且 $\|g\|_2\leqslant 1$, 根据柯西不等式,

$$g^{\mathrm{T}}(x-0)\leqslant\|g\|_2\|x\|_2\leqslant\|x\|_2-0$$

因此 $\{g \mid \|g\|_2\leqslant 1\}\subseteq\partial f(0)$. 接下来说明若 $\|g\|_2>1$, 则 $g\notin\partial f(0)$. 取 $x=g$, 若 g 为次梯度, 则

$$\|g\|_2-0\geqslant g^{\mathrm{T}}(g-0)=\|g\|_2^2>\|g\|_2$$

这显然是矛盾的. 综上, 我们有

$$\partial f(0)=\{g \mid \|g\|_2\leqslant 1\}$$

30.5.3 有界性

凸函数 $f(x)$ 的次梯度和次微分有许多有用的性质. 下面的定理说明次微分 $\partial f(x)$ 在一定条件下分别为闭凸集和非空有界集.

定理 18

设 f 是凸函数, 则 $\partial f(x)$ 有如下性质:

1. 对任何 $x \in \mathrm{dom}\, f, \partial f(x)$ 是一个闭凸集 (可能为空集);
2. 如果 $x \in \mathrm{int}\,\mathrm{dom}\, f$, 则 $\partial f(x)$ 非空有界集.

Proof.

设 $g_1, g_2 \in \partial f(x)$，并设 $\lambda \in (0,1)$，由次梯度的定义我们有

$$f(y) \geqslant f(x) + g_1^{\mathrm{T}}(y-x), \quad \forall y \in \operatorname{dom} f$$
$$f(y) \geqslant f(x) + g_2^{\mathrm{T}}(y-x), \quad \forall y \in \operatorname{dom} f$$

由上面第一式的 λ 倍加上第二式的 $(1-\lambda)$ 倍，我们得到 $\lambda g_1 + (1-\lambda) g_2 \in \partial f(x)$，从而 $\partial f(x)$ 是凸集．此外令 $g_k \in \partial f(x)$ 为次梯度且 $g_k \to g$，则

$$f(y) \geqslant f(x) + g_k^{\mathrm{T}}(y-x), \quad \forall y \in \operatorname{dom} f,$$

在上述不等式中取极限，并注意到极限的保号性，最终我们有

$$f(y) \geqslant f(x) + g^{\mathrm{T}}(y-x), \quad \forall y \in \operatorname{dom} f.$$

这说明 $\partial f(x)$ 为闭集．

Proof.

（续）下设 $x \in \operatorname{int}\operatorname{dom} f$, 我们来证明 $\partial f(x)$ 是非空有界的. 首先, $\partial f(x)$ 非空是定理17的直接结果, 因此我们只需要证明有界性. 对 $i=1,2,\cdots,n$, 定义 $e_i=(0,\cdots,1,\cdots,0)$ (第 i 个分量为 1 , 其余分量均为 0), 易知 $\{e_i\}_{i=1}^n$ 为 $\mathbb{R}^n$ 的一组标准正交基. 取定充分小的正数 r, 使得

$$B=\{x \pm r e_i \mid i=1,2,\cdots,n\} \subset \operatorname{dom} f$$

对任意 $g \in \partial f(x)$, 不妨设 g 不为 0 . 存在 $y \in B$ 使得

$$f(y) \geqslant f(x)+g^{\mathrm{T}}(y-x)=f(x)+r\|g\|_{\infty}.$$

由此得到

$$\|g\|_{\infty} \leqslant \frac{\max_{y \in B} f(y)-f(x)}{r}<+\infty$$

即 $\partial f(x)$ 有界. □

定理 19

设 $f(x)$ 在 $x_0 \in \mathrm{int\ dom}\, f$ 处可微，则 $\partial f(x_0) = \{\nabla f(x_0)\}$

Proof.

根据可微凸函数的一阶条件可知梯度 $\nabla f(x_0)$ 为次梯度. 下证 $f(x)$ 在点 x_0 处不可能有其他次梯度. 设 $g \in \partial f(x_0)$，根据次梯度的定义，对任意的非零 $v \in \mathbb{R}^n$ 且 $x_0 + tv \in \mathrm{dom}\, f, t > 0$ 有

$$f(x_0 + tv) \geqslant f(x_0) + \mathrm{tg}^{\mathrm{T}} v.$$

若 $g \neq \nabla f(x_0)$，取 $v = g - \nabla f(x_0) \neq 0$，上式变形为

$$\frac{f(x_0 + tv) - f(x_0) - t\nabla f(x_0)^{\mathrm{T}} v}{t\|v\|} \geqslant \frac{(g - \nabla f(x_0))^{\mathrm{T}} v}{\|v\|} = \|v\|.$$

不等式两边令 $t \to 0$，根据 Fréchet 可微的定义，左边趋于 0 ，而右边是非零正数，可得到矛盾. □

30.5.4 单调性

和梯度类似，凸函数的次梯度也具有某种单调性．这一性质在很多和次梯度有关的算法的收敛性分析中起到了关键的作用．

定理 20

(次梯度的单调性) 设 $f: \mathbb{R}^n \to \mathbb{R}$ 为凸函数，$x, y \in \operatorname{dom} f$，则

$$(u-v)^{\mathrm{T}}(x-y) \geqslant 0$$

其中 $u \in \partial f(x), v \in \partial f(y)$．

Proof.

由次梯度的定义，

$$f(y) \geqslant f(x) + u^{\mathrm{T}}(y-x)$$

$$f(x) \geqslant f(y) + v^{\mathrm{T}}(x-y).$$

将以上两个不等式相加即得结论． □

30.5.5 下半连续性

对于闭凸函数（即凸下半连续函数），次梯度还具有某种连续性.

定理 21

设 $f(x)$ 是闭凸函数且 ∂f 在点 $\bar{x}$ 附近存在且非空. 若序列 $x^k \to \bar{x}, g^k \in \partial f\left(x^k\right)$ 为 $f(x)$ 在点 x^k 处的次梯度, 且 $g^k \to \bar{g}$, 则 $\bar{g} \in \partial f(\bar{x})$.

Proof.

对任意 $y \in \operatorname{dom} f$, 根据次梯度的定义,

$$f(y) \geqslant f\left(x^k\right) + \left\langle g^k, y - x^k \right\rangle .$$

对上述不等式两边取下极限, 我们有

$$\begin{aligned} f(y) &\geqslant \liminf_{k\to\infty} \left[f\left(x^k\right) + \left\langle g^k, y - x^k \right\rangle \right] \\ &\geqslant f(\bar{x}) + \langle \bar{g}, y - \bar{x} \rangle \end{aligned}$$

其中第二个不等式利用了 $f(x)$ 的下半连续性以及 $g^k \to \bar{g}$, 由此可推出 $\bar{g} \in \partial f(\bar{x})$. □

30.5.6 凸函数的方向导数

在微积分中我们接触过方向导数的概念. 设 f 为适当函数, 给定点 $\boldsymbol{x}_0$ 以及方向 $\boldsymbol{d} \in \mathbb{R}^n$, 方向导数 (若存在) 定义为

$$\lim_{t \downarrow 0} \phi(t) = \lim_{t \downarrow 0} \frac{f(\boldsymbol{x}_0 + t\boldsymbol{d}) - f(\boldsymbol{x}_0)}{t},$$

其中 $t \downarrow 0$ 表示 t 单调下降趋于 0. 对于凸函数 $f(\boldsymbol{x})$, 易知 $\phi(t)$ 在 $(0, +\infty)$ 上是单调不减的, 上式中的极限号 $\lim$ 可以替换为下确界 inf. 上述此时极限总是存在 (可以为无穷), 进而凸函数总是可以定义方向导数.

定义 10

（方向导数）对于凸函数 f, 给定点 $\boldsymbol{x}_0 \in \operatorname{dom} f$ 以及方向 $\boldsymbol{d} \in \mathbb{R}^n$, 其方向导数定义为

$$\partial f(\boldsymbol{x}_0; \boldsymbol{d}) = \inf_{t>0} \frac{f(\boldsymbol{x}_0 + t\boldsymbol{d}) - f(\boldsymbol{x}_0)}{t}.$$

方向导数可能是正负无穷，但在定义域的内点处方向导数 $\partial f(\boldsymbol{x}_0; \boldsymbol{d})$ 是有限的.

定理 22

设 $f(\boldsymbol{x})$ 为凸函数，$\boldsymbol{x}_0 \in \operatorname{int} \operatorname{dom} f$, 则对任意 $\boldsymbol{d} \in \mathbb{R}^n, \partial f(\boldsymbol{x}_0; \boldsymbol{d})$ 有限.

Proof.

首先 $\partial f(\boldsymbol{x}_0; \boldsymbol{d})$ 不为正无穷是显然的. 由于 $\boldsymbol{x}_0 \in \operatorname{int} \operatorname{dom} f$, 根据存在定理可知 $f(\boldsymbol{x})$ 在点 $\boldsymbol{x}_0$ 处存在次梯度 $\boldsymbol{g}$. 根据方向导数的定义，我们有

$$
\begin{aligned}
\partial f(\boldsymbol{x}_0; \boldsymbol{d}) &= \inf_{t>0} \frac{f(\boldsymbol{x}_0 + t\boldsymbol{d}) - f(\boldsymbol{x}_0)}{t} \\
&\geqslant \inf_{t>0} \frac{t\boldsymbol{g}^{\mathrm{T}} \boldsymbol{d}}{t} = \boldsymbol{g}^{\mathrm{T}} \boldsymbol{d}.
\end{aligned}
$$

其中的不等式利用了次梯度的定义. 这说明 $\partial f(\boldsymbol{x}_0; \boldsymbol{d})$ 不为负无穷. □

30.5.7 方向导数和次梯度的联系

凸函数的方向导数和次梯度之间有很强的联系. 以下结果表明, 凸函数 $f(\boldsymbol{x})$ 关于 $\boldsymbol{d}$ 的方向导数 $\partial f(\boldsymbol{x}; \boldsymbol{d})$ 正是 f 在点 $\boldsymbol{x}$ 处的所有次梯度与 $\boldsymbol{d}$ 的内积的最大值.

定理 23

设 $f : \mathbb{R}^n \to (-\infty, +\infty]$ 为凸函数, 点 $\boldsymbol{x}_0 \in \operatorname{int} \operatorname{dom} f$, $\boldsymbol{d}$ 为 $\mathbb{R}^n$ 中任一方向, 则

$$\partial f(\boldsymbol{x}_0; \boldsymbol{d}) = \max_{\boldsymbol{g} \in \partial f(\boldsymbol{x}_0)} \boldsymbol{g}^{\mathrm{T}} \boldsymbol{d}.$$

上述定理可对一般的 $\boldsymbol{x} \in \operatorname{dom} f$ 作如下推广：

定理 24

设 f 为适当凸函数，且在 $\boldsymbol{x}_0$ 处次微分不为空集，则对任意 $\boldsymbol{d} \in \mathbb{R}^n$ 有

$$\partial f(\boldsymbol{x}_0; \boldsymbol{d}) = \sup_{\boldsymbol{g} \in \partial f(\boldsymbol{x}_0)} \boldsymbol{g}^{\mathrm{T}} \boldsymbol{d},$$

且当 $\partial f(\boldsymbol{x}_0; \boldsymbol{d})$ 不为无穷时，上确界可以取到.

30.5.8 次梯度的计算规则

如何计算一个不可微凸函数的次梯度在优化算法设计中是很重要的问题. 根据定义来计算次梯度一般来说比较繁琐, 我们来介绍一些次梯度的计算规则. 本小节讨论的计算规则都默认 $x \in \operatorname{int} \operatorname{dom} f$.

1. 基本规则

我们首先不加证明地给出一些计算次梯度 (次微分）的基本规则.

1. 可微凸函数: 设 f 为凸函数, 若 f 在点 x 处可微, 则 $\partial f(x) = \{\nabla f(x)\}$.
2. 凸函数的非负线性组合: 设 f_1, f_2 为凸函数, 且满足

$$\text{int dom } f_1 \cap \operatorname{dom} f_2 \neq \varnothing,$$

而 $x \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$. 若

$$f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x), \quad \alpha_1, \alpha_2 \geqslant 0,$$

则 $f(x)$ 的次微分

$$\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x).$$

基本规则

3. 线性变量替换: 设 h 为适当凸函数, 并且函数 f 满足

$$f(x) = h(Ax + b), \quad \forall x \in \mathbb{R}^m,$$

其中 $A \in \mathbb{R}^{n\times m}, b \in \mathbb{R}^n$. 若存在 $x^\sharp \in \mathbb{R}^m$, 使得 $Ax^\sharp + b \in$ int dom h, 则

$$\partial f(x) = A^{\mathrm{T}} \partial h(Ax + b), \quad \forall x \in \mathrm{int\,dom}\, f.$$

2. 两个函数之和的次梯度

以下的 Moreau-Rockafellar 定理给出两个凸函数之和的次微分的计算方法.

定理 25

设 $f_1, f_2 : \mathbb{R}^n \to (-\infty, +\infty]$ 是两个凸函数, 则对任意的 $x_0 \in \mathbb{R}^n$,

$$\partial f_1(x_0) + \partial f_2(x_0) \subseteq \partial (f_1 + f_2)(x_0).$$

进一步地, 若 $\operatorname{int}\operatorname{dom} f_1 \cap \operatorname{dom} f_2 \neq \varnothing$, 则对任意的 $x_0 \in \mathbb{R}^n$,

$$\partial (f_1 + f_2)(x_0) = \partial f_1(x_0) + \partial f_2(x_0).$$

3. 函数族的上确界

容易验证一族凸函数的上确界函数仍是凸函数. 我们有如下重要结果:

定理 26

设 $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to (-\infty, +\infty]$ 均为凸函数, 令

$$f(x) = \max\left\{f_1(x), f_2(x), \cdots, f_m(x)\right\}, \quad \forall x \in \mathbb{R}^n.$$

对 $x_0 \in \bigcap_{i=1}^m \operatorname{int} \operatorname{dom} f_i$, 定义 $I(x_0) = \{i \mid f_i(x_0) = f(x_0)\}$, 则

$$\partial f(x_0) = \operatorname{conv} \bigcup_{i \in I(x_0)} \partial f_i(x_0).$$

4. 固定分量的函数极小值

设 $h:\mathbb{R}^n\times\mathbb{R}^m\to(-\infty,+\infty]$ 是关于 (x,y) 的凸函数, 则 $f(x)\overset{\text{def}}{=}\inf_y h(x,y)$ 是关于 $x\in\mathbb{R}^n$ 的凸函数. 以下结果可以用于求解 f 在点 x 处的一个次梯度.

定理 27

考虑函数

$$f(x)=\inf_y h(x,y)$$

其中

$$h:\mathbb{R}^n\times\mathbb{R}^m\to(-\infty,+\infty]$$

是关于 (x,y) 的凸函数. 对 $\hat{x}\in\mathbb{R}^n$, 设 $\hat{y}\in\mathbb{R}^m$ 满足 $h(\hat{x},\hat{y})=f(\hat{x})$, 且存在 $g\in\mathbb{R}^n$ 使得 $(g,0)\in\partial h(\hat{x},\hat{y})$, 则 $g\in\partial f(\hat{x})$.

5. 复合函数

对于复合函数的次梯度，我们有如下链式法则（注意比较其与可微情形下链式法则的异同):

定理 28

设 $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to (-\infty, +\infty]$ 为 m 个凸函数，$h : \mathbb{R}^m \to (-\infty, +\infty]$ 为关于各分量单调递增的凸函数，令

$$f(x) = h\left(f_1(x), f_2(x), \cdots, f_m(x)\right)$$

设 $z = (z_1, z_2, \cdots, z_m) \in \partial h\left(f_1(\hat{x}), f_2(\hat{x}), \cdots, f_m(\hat{x})\right)$ 以及 $g_i \in \partial f_i(\hat{x})$，则

$$g \overset{\text{def}}{=} z_1 g_1 + z_2 g_2 + \cdots + z_m g_m \in \partial f(\hat{x})$$

一些常见的次梯度计算示例

例 15

（分段线性函数）令

$$f(x) = \max_{i=1,2,\cdots,m} \{a_i^{\mathrm{T}} x + b_i\}$$

其中 $x, a_i \in \mathbb{R}^n, b_i \in \mathbb{R}, i = 1, 2, \cdots, m,$ 则

$$\partial f(x) = \operatorname{conv}\{a_i \mid i \in I(x)\}$$

其中

$$I(x) = \{i \mid a_i^{\mathrm{T}} x + b_i = f(x)\}$$

一些常见的次梯度计算示例

例 16

（ℓ_1 范数）定义 $f:\mathbb{R}^n\to\mathbb{R}$ 为 ℓ_1 范数，则对 $x=(x_1,x_2,\cdots,x_n)\in\mathbb{R}^n$，有

$$f(x)=\|x\|_1=\max_{s\in\{-1,1\}^n} s^{\mathrm{T}}x.$$

于是

$$\partial f(x)=J_1\times J_2\times\cdots\times J_n,\quad J_k=\begin{cases}[-1,1], x_k=0\\ \{1\}, x_k>0\\ \{-1\}, x_k<0\end{cases}$$

一些常见的次梯度计算示例

例 17

设 C 是 $\mathbb{R}^n$ 中一闭凸集，令 $f(\boldsymbol{x}) = \inf_{\boldsymbol{y}\in C} \|\boldsymbol{x}-\boldsymbol{y}\|_2$. 令 $\hat{\boldsymbol{x}} \in \mathbb{R}^n$，我们来求 f 在 $\hat{\boldsymbol{x}}$ 处的一个次梯度. *(1)* 若 $f(\hat{\boldsymbol{x}}) = 0$，则容易验证 $\boldsymbol{g} = 0 \in \partial f(\hat{\boldsymbol{x}})$; *(2)* 若 $f(\hat{\boldsymbol{x}}) > 0$，由 C 是闭凸集，可取 $\hat{\boldsymbol{y}}$ 为 $\hat{\boldsymbol{x}}$ 在 C 上的投影，即

$$\hat{\boldsymbol{y}} = \mathcal{P}_c(\hat{\boldsymbol{x}}) \overset{def}{=} \underset{\boldsymbol{y}\in C}{\operatorname{argmin}} \|\hat{\boldsymbol{x}}-\boldsymbol{y}\|_2.$$

利用 $\hat{\boldsymbol{y}}$ 的定义可以验证

$$\boldsymbol{g} = \frac{1}{\|\hat{\boldsymbol{x}}-\hat{\boldsymbol{y}}\|_2}(\hat{\boldsymbol{x}}-\hat{\boldsymbol{y}}) = \frac{1}{\|\hat{\boldsymbol{x}}-\mathcal{P}_c(\hat{\boldsymbol{x}})\|_2}\left(\hat{\boldsymbol{x}}-\mathcal{P}_c(\hat{\boldsymbol{x}})\right),$$

满足固定分量的函数极小值情形的条件. 故 $\boldsymbol{g} \in \partial f(\hat{\boldsymbol{x}})$.

本讲小结

凸函数

- 凸函数的概念
- 凸函数的判定条件：一阶条件、二阶条件
- 凸函数的保凸运算：非负加权求和、复合仿射映射、逐点最大和逐点上确界、最小化、透视函数

共轭函数、次梯度

- 共轭函数的定义
- 共轭函数的性质
- 机器学习中的共轭函数
- 次梯度：次梯度的定义、性质及计算规则

凸函数是凸优化中研究的主要对象，了解保凸运算判定一个优化问题是否是凸优化有着重要的应用价值。共轭函数将在对偶函数的求解中发挥作用。