邻近算子

文再文

北京大学北京国际数学研究中心

教材《最优化：建模、算法与理论》配套电子教案

http://bicmr.pku.edu.cn/~wenzw/optbook.html

致谢：本教案由陈乐恒、朱桢源协助准备

提纲

1 闭函数

2 共轭函数

3 邻近算子

4 投影算子

5 性质和推广

闭集

一个包含其边界的集合 C 被称为**闭集**：

$$x^k \in C, \quad x^k \to \bar{x} \qquad \Longrightarrow \qquad \bar{x} \in C$$

保持闭性的操作：

- （有限或无限个）闭集的交集仍是闭集
- 有限个闭集的并集仍是闭集
- 线性映射的原象集: 在 C 闭的情形下，$\{x \mid Ax \in C\}$ 是闭集，

线性映射的像

一个闭集在线性映射下的像不一定是闭的

例: (C 闭, $AC = \{Ax \mid x \in C\}$ 开)

$C = \{(x_1, x_2) \in \mathbf{R}^2_+ \mid x_1 x_2 \geq 1\}, \quad A = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad AC = \mathbf{R}_{++}$

（**充分条件**） 以下条件成立时，AC 为闭集：

- C 是闭凸集
- A 的零空间中不包含 C 的衰退方向 (recession direction)，即

$$Ay = 0, \quad \hat{x} \in C, \quad \hat{x} + \alpha y \in C \quad \forall \alpha \geq 0 \qquad \Longrightarrow \qquad y = 0$$

特别地，若 C 有界，则 AC 为闭集

闭函数

定义

一个函数被称为闭函数，如果它的上方图是闭集

闭函数的例子：

- $f(x) = -\log(1 - x^2)$, $\mathbf{dom}\, f = \{x \mid |x| < 1\}$
- $f(x) = x \log x$, $\mathbf{dom}\, f = \mathbf{R}_+$ 且$f(0) = 0$
- 闭集 C 的示性函数

不是闭函数的例子：

- $f(x) = x \log x$, $\mathbf{dom}\, f = \mathbf{R}_{++}$ 或 $\mathbf{dom}\, f = \mathbf{R}_+$, $f(0) = 1$
- 不是闭集的集合 C 的示性函数

性质

下水平集: f 是闭函数当且仅当 f 的所有 α-下水平集都是闭集

最小值: 如果 f 是闭函数且存在有界的下水平集，则 f 有最小值点

常见的保闭性的操作（凸函数）

- 加法: $f+g$ 是闭函数，如果 f 和 g 都是闭的 $(\mathbf{dom}\, f \cap \mathbf{dom}\, g \neq \varnothing)$
- 复合线性映射: $f(Ax+b)$是闭函数，如果 f 是闭的
- 取上确界: $\sup_\alpha f_\alpha(x)$ 是闭函数，如果任意函数 f_α是闭的

提纲

1 闭函数

2 共轭函数

3 邻近算子

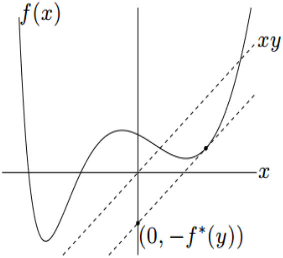
4 投影算子

5 性质和推广

共轭函数

函数 f 的**共轭函数**定义为：

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



f^* 恒为闭凸函数

Fenchel 不等式：

$$f(x) + f^*(y) \geq x^T y \quad \forall x, y$$

Proof.

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} \{y^\mathrm{T} x - f(x)\} \geq y^\mathrm{T} x - f(x), \quad \forall x \in \mathbf{dom}\, f$$

□

考察二次函数f的共轭函数：

$$f(x) = \frac{1}{2}x^T Ax + b^T x + c$$

- **强凸情形**（$A \succ 0$）

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

- **一般凸情形**（$A \succeq 0$）

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

这里$\mathcal{R}(A)$为A的像空间.

负熵与负对数

- **负熵**

$$f(x)=\sum_{i=1}^{n} x_i \log x_i \qquad f^*(y)=\sum_{i=1}^{n} e^{y_i-1}$$

- **负对数**

$$f(x)=-\sum_{i=1}^{n} \log x_i \qquad f^*(y)=-\sum_{i=1}^{n} \log(-y_i)-n$$

- **矩阵对数**

$$f(X)=-\log\det X \quad (\mathbf{dom}\, f=\mathbf{S}_{++}^n) \quad f^*(Y)=-\log\det(-Y)-n$$

示性函数与范数

凸集 C 的**示性函数**：共轭为 C 的支撑函数

$$f(x) = \begin{cases} 0, & x \in C \\ +\infty, & x \notin C \end{cases} \qquad f^*(y) = \sup_{x \in C} y^T x$$

范数：共轭为单位对偶范数球的示性函数

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

Proof.

回忆对偶范数的定义：$||y||_* = \sup_{||x|| \leq 1} x^T y$

分两类讨论计算$f^*(y) = \sup_x (y^T x - ||x||)$

- 若$||y||_* \leq 1$，则 $y^T x \leq ||x|| \quad \forall x$ (对偶范数的定义)
  $x = 0$时等式成立，因此 $\sup_x (y^T x - ||x||) = 0$
- 若$||y||_* > 1$，则存在一个 x，满足$||x|| \leq 1, x^T y > 1$，因此有

$$f^*(y) \geq y^T (tx) - ||tx|| = t(y^T x - ||x||) \to \infty \quad (t \to \infty)$$

□

二次共轭函数

任一函数f的二次共轭函数定义为

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

- $f^{**}(x)$ 为闭凸函数
- 由Fenchel不等式，$x^T y - f^*(y) \leq f(x)$ 对所有x, y 都成立，推出：

  $$f^{**}(x) \leq f(x) \quad \forall x$$

  等价地，$\textbf{epi}\, f \subseteq \textbf{epi}\, f^{**}$ (对任意函数f 成立)
- 若f 是闭凸函数，则

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

  等价地，$\textbf{epi}\, f = \textbf{epi}\, f^{**}$ (若f 是闭凸函数); 证明在下一面

Proof.

假设$(x, f^{**}(x)) \notin \mathbf{epi}\, f$，则存在严格的分割超平面

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

其中$a \in \mathbb{R}^n, b, c \in \mathbb{R}$ 且$b \leqslant 0$ (若$b > 0$, 则取 $s \to +\infty$ 可推出矛盾).

- 若$b < 0$,取$s = f(z)$,有$a^{\mathrm{T}}z + bf(z) - a^{\mathrm{T}}x - bf^{**}(x) \leqslant c$
  记$y = a/(-b)$,两边除以$-b$，并将上式左边关于 z 极大化得到
  $$f^*(y) - y^{\mathrm{T}}x + f^{**}(x) \leqslant -\frac{c}{b} < 0$$
  与Fenchel 不等式矛盾.
- 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$\varepsilon$倍，则
  $$\begin{bmatrix} a+\varepsilon\hat{y} \\ -\varepsilon \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} z-x \\ s-f^{**}(x) \end{bmatrix} \leqslant c + \varepsilon\left(f^*(\hat{y}) - x^{\mathrm{T}}\hat{y} + f^{**}(x)\right) < 0$$
  即化为$b < 0$ 的情况，矛盾.

$\square$

共轭函数与次梯度

定理

如果 f 是闭凸函数，则

$$y \in \partial f(x) \quad \Leftrightarrow \quad x \in \partial f^*(y) \quad \Leftrightarrow \quad x^T y = f(x) + f^*(y)$$

Proof.

若$y \in \partial f(x)$，则$f^*(y) = \sup_u (y^T u - f(u)) = y^T x - f(x)$

$$\begin{aligned} f^*(v) &= \sup_u (v^T u - f(u)) \\ &\geq v^T x - f(x) \\ &= x^T (v - y) - f(x) + y^T x \\ &= f^*(y) + x^T (v - y) \end{aligned}$$

对所有的v成立; 由此根据次梯度的定义推出$x \in \partial f^*(y)$
另一方面，$x \in \partial f^*(y) \Rightarrow y \in \partial f(x)$ 可以由 $f^{**} = f$ 得到 □

计算规则

- 可分解的和：

$$f(x_1, x_2) = g(x_1) + h(x_2) \qquad f^*(y_1, y_2) = g^*(y_1) + h^*(y_2)$$

- 数乘：($\alpha > 0$)

$$f(x) = \alpha g(x) \qquad f^*(y) = \alpha g^*(y/\alpha)$$

- 添加线性函数：

$$f(x) = g(x) + a^T x + b \qquad f^*(y) = g^*(y - a) - b$$

- 卷积下确界：

$$f(x) = \inf_{u+v=x} (g(u) + h(v)) \qquad f^*(y) = g^*(y) + h^*(y)$$

提纲

1 闭函数

2 共轭函数

3 邻近算子

4 投影算子

5 性质和推广

邻近算子

定义**邻近算子**：

$$\mathrm{prox}_h(x) = \underset{u}{\mathrm{argmin}}\left(h(u) + \frac{1}{2}\|u - x\|_2^2\right)$$

直观理解：求解一个距x不算太远的点u，并使函数值$h(u)$也相对较小

定理（邻近算子是良定义的）

如果 h 为闭凸函数，则对任意 x，$\mathrm{prox}_h(x)$**存在且唯一**

Proof.

首先注意到$h(u) + \frac{1}{2}\|u - x\|_2^2$ 是强凸函数，则

- 存在性：强凸函数的所有α-下水平集有界，故由Weierstrass定理知最小值存在
- 唯一性：强凸函数最小值唯一

□

邻近算子与次梯度的关系

定理

若h 是适当的闭凸函数，则 $u = \mathrm{prox}_h(x) \iff x - u \in \partial h(u)$

Proof.

若$u = \mathrm{prox}_h(x)$, 则由最优性条件得$0 \in \partial h(u) + (u - x)$, 因此
有$x - u \in \partial h(u)$. 反之，若$x - u \in \partial h(u)$ 则由次梯度的定义可得到

$$h(v) \geqslant h(u) + (x-u)^{\mathrm{T}}(v-u), \quad \forall v \in \mathbf{dom}\ h$$

两边同时加$\frac{1}{2}\|v-x\|^2$, 即有

$$\begin{aligned} h(v) + \frac{1}{2}\|v-x\|^2 &\geqslant h(u) + (x-u)^{\mathrm{T}}(v-u) + \frac{1}{2}\|(v-u)-(x-u)\|^2 \\ &\geqslant h(u) + \frac{1}{2}\|u-x\|^2, \quad \forall v \in \mathbf{dom}\ h \end{aligned}$$

根据定义可得$u = \mathrm{prox}_h(x)$. □

邻近算子的例子

在近似点梯度法中，我们关心那些邻近算子prox_{th}容易计算的函数 h

例：ℓ_1 范数

$$h(x) = \|x\|_1, \quad \mathrm{prox}_{th}(x) = \mathrm{sign}(x)\max\{|x| - t, 0\}$$

Proof.

邻近算子$u = \mathrm{prox}_{th}(x)$ 的最优性条件为

$$x - u \in t\partial\|u\|_1 = \begin{cases} \{t\}, & u > 0 \\ [-t, t], & u = 0 \\ \{-t\}, & u < 0 \end{cases}$$

当$x > t$ 时, $u = x - t$; 当$x < -t$ 时, $u = x + t$; 当$x \in [-t, t]$ 时, $u = 0$,
即有$u = \mathrm{sign}(x)\max\{|x| - t, 0\}$. □

邻近算子的例子

例：ℓ_2 范数

$$h(x) = \|x\|_2, \quad \mathrm{prox}_{th}(x) = \begin{cases} \left(1 - \frac{t}{\|x\|_2}\right)x, & \|x\|_2 \geqslant t, \\ 0, & \text{其他}. \end{cases}$$

Proof.

邻近算子$u = \mathrm{prox}_{th}(x)$ 的最优性条件为

$$x - u \in t\partial\|u\|_2 = \begin{cases} \left\{\frac{tu}{\|u\|_2}\right\}, & u \neq 0, \\ \{w : \|w\|_2 \leqslant t\}, & u = 0, \end{cases}$$

因此，当 $\|x\|_2 > t$ 时，$u = x - \frac{tx}{\|x\|_2}$；当 $\|x\|_2 \leqslant t$ 时，$u = 0$. □

邻近算子的例子

- 二次函数(其中 A 对称正定)

$$h(x) = \frac{1}{2}x^{\mathrm{T}}Ax + b^{\mathrm{T}}x + c, \quad \mathrm{prox}_{th}(x) = (I + tA)^{-1}(x - tb)$$

- 负自然对数的和

$$h(x) = -\sum_{i=1}^{n} \ln x_i, \quad \mathrm{prox}_{th}(x)_i = \frac{x_i + \sqrt{x_i^2 + 4t}}{2}, \quad i = 1, 2, \cdots, n$$

邻近算子的计算规则

- 变量的常数倍放缩以及平移 $(\lambda \neq 0)$：

$$h(x) = g(\lambda x + a), \quad \text{prox}_h(x) = \frac{1}{\lambda}\left(\text{prox}_{\lambda^2 g}(\lambda x + a) - a\right)$$

- 函数（及变量）的常数倍放缩 $(\lambda > 0)$：

$$h(x) = \lambda g\left(\frac{x}{\lambda}\right), \quad \text{prox}_h(x) = \lambda \text{prox}_{\lambda^{-1} g}\left(\frac{x}{\lambda}\right)$$

- 加上线性函数：

$$h(x) = g(x) + a^{\mathrm{T}} x, \quad \text{prox}_h(x) = \text{prox}_g(x - a)$$

计算规则（续）

- 加上二次项 $(u > 0)$

$$h(x) = g(x) + \frac{u}{2}\|x - a\|_2^2, \quad \mathrm{prox}_h(x) = \mathrm{prox}_{\theta g}(\theta x + (1-\theta)a)$$

其中$\theta = \frac{1}{1+u}$

- 向量函数：

$$h\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \varphi_1(x) + \varphi_2(y), \quad \mathrm{prox}_h\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \mathrm{prox}_{\varphi_1}(x) \\ \mathrm{prox}_{\varphi_2}(y) \end{bmatrix}$$

- 已知函数$g(x)$和矩阵A，设$h(x)=g(Ax+b)$. 在通常情况下，我们不能使用g的邻近算子直接计算关于h的邻近算子.
- 然而，如果有$AA^\mathrm{T}=\frac{1}{\alpha}I$（其中$\alpha$为任意正常数），则

$$\mathrm{prox}_h(x)=(I-\alpha A^\mathrm{T}A)x+\alpha A^\mathrm{T}(\mathrm{prox}_{\alpha^{-1}g}(Ax+b)-b).$$

- 例如，$h(x_1,x_2,\cdots,x_m)=g(x_1+x_2+\cdots+x_m)$ 的邻近算子为

$$\mathrm{prox}_h(x_1,x_2,\cdots,x_m)_i=x_i-\frac{1}{m}\left(\sum_{j=1}^m x_j-\mathrm{prox}_{mg}\left(\sum_{j=1}^m x_j\right)\right).$$

Proof.

考虑如下优化问题：

$$\begin{aligned}\min_{u,y}\quad & g(y)+\frac{1}{2}\|u-x\|^2,\\ \text{s.t.}\quad & Au+b=y,\end{aligned}$$

则其解中的$u=\mathrm{prox}_h(x)$．固定y 对于u 求极小值，这是一个到仿射集的投影问题，其解为

$$\begin{aligned}u&=x+A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}(y-b-Ax)\\ &=(I-\alpha A^{\mathrm{T}}A)x+\alpha A^{\mathrm{T}}(y-b).\end{aligned}$$

将其代入优化问题，将目标函数化为

$$g(y)+\frac{\alpha^2}{2}\|A^{\mathrm{T}}(y-b-Ax)\|^2=g(y)+\frac{\alpha}{2}\|y-b-Ax\|^2.$$

由此得到$y=\mathrm{prox}_{\alpha^{-1}g}(Ax+b)$，再代入$u$的表达式中即可得到结果. □

提纲

1. 闭函数
2. 共轭函数
3. 邻近算子
4. 投影算子
5. 性质和推广

闭凸集上的投影

设 C 为闭凸集，则示性函数 I_C 的邻近算子为点 x 到 C 的投影$\mathcal{P}_C(x)$：

$$\begin{aligned}\mathrm{prox}_{I_C}(x) &= \arg\min_u \left\{ I_C(u) + \frac{1}{2}\|u-x\|^2 \right\} \\ &= \operatorname*{argmin}_{u\in C} \|u-x\|^2 = \mathcal{P}_C(x)\end{aligned}$$

这个等式具有几何意义：

$$u = \mathcal{P}_C(x) \Leftrightarrow (x-u)^{\mathrm{T}}(z-u) \leqslant 0, \quad \forall z \in C$$

投影到仿射集

超平面 $C = \{x | a^T x = b\}$ ($a \neq 0$)

$$P_C(x) = x + \frac{b - a^T x}{\|a\|_2^2} a$$

仿射集 $C = \{x | Ax = b\}$ ($A \in \mathbb{R}^{p \times n}$，且$\text{rank}(A) = p$)

$$P_C(x) = x + A^T (AA^T)^{-1} (b - Ax)$$

当$p \ll n$, 或$AA^T = I, \ldots$ 时，计算成本较低

投影到多面体集

半平面 $C=\{x|a^Tx\leq b\}$ $(a\neq 0)$

$$P_C(x)=x+\frac{b-a^Tx}{\|a\|_2^2}a \quad if \quad a^Tx>b,$$

$$P_C(x)=x \quad if \quad a^Tx\leq b$$

矩形：$C=[l,u]=\{x \mid l\leq x\leq u\}$

$$P_C(x)_i=\begin{cases} l_i & x_i\leq l_i \\ x_i & l_i\leq x_i\leq u_i \\ u_i & x_i\geq u_i \end{cases}$$

非负象限：$C=\mathbf{R}_+^n$

$$P_C(x)=x_+ \qquad (x_+ \text{表示各分量取} max\{0,x\})$$

概率单纯形：$C=\{x|1^Tx=1,x\geq 0\}$

$$P_C(x)=(x-\lambda 1)_+$$

其中，λ是下面方程的解：

$$1^T(x-\lambda 1)_+=\sum_{i=1}^{n}\max\{0,x_k-\lambda\}=1$$

（一般的）概率单纯形：$C=\{x|a^Tx=b,l\leq x\leq u\}$

$$P_c(x)=P_{[l,u]}(x-\lambda a)$$

其中, λ 是下面方程的解：

$$a^TP_{[l,u]}(x-\lambda a)=b$$

投影到范数球

Euclid 球：$C = \{x | \|x\|_2 \leq 1\}$

$$P_C(x) = \frac{1}{\|x\|_2} x \quad \textit{if} \quad \|x\|_2 > 1,$$

$$P_C(x) = x \quad \textit{if} \quad \|x\|_2 \leq 1$$

ℓ_1 **范数球**：$C = \{x | \|x\|_1 \leq 1\}$

$$P_c(x)_k = \begin{cases} x_k - \lambda & x_k > \lambda \\ 0 & -\lambda \leq x_k \leq \lambda \\ x_k + \lambda & x_k < -\lambda \end{cases}$$

若$\|x\|_1 \leq 1$，则$\lambda = 0$; 其他情形，λ 是下面方程的解

$$\sum_{k=1}^{n} \max\{|x_k| - \lambda, 0\} = 1$$

投影到简单锥

二阶锥： $C = \{(x,t) \in \mathbf{R}^{n\times 1} | \|x\|_2 \leq t\}$

$$P_C(x,t) = (x,t) \quad \textit{if} \quad \|x\|_2 \leq t,$$

$$P_C(x,t) = (0,0) \quad \textit{if} \quad \|x\|_2 \leq -t$$

且

$$P_C(x,t) = \frac{t + \|x\|_2}{2\|x\|_2} \begin{bmatrix} x \\ \|x\|_2 \end{bmatrix} \quad \textit{if} \ \|x\|_2 \geq |t|, \ x \neq 0$$

半正定锥： $C = \mathbf{S}_+^n$

$$P_C(X) = \sum_{i=1}^{n} \max\{0, \lambda_i\} q_i q_i^T$$

其中， $X = \sum_{i=1}^{n} \lambda_i q_i q_i^T$ 是X 的特征值分解

提纲

1 闭函数

2 共轭函数

3 邻近算子

4 投影算子

5 性质和推广

Moreau分解

Moreau分解描述了邻近算子与共轭函数之间的关系：

$$x = \operatorname{prox}_h(x) + \operatorname{prox}_{h^*}(x)$$

- 这来自于共轭函数和次梯度的性质:

$$\begin{aligned} u = \operatorname{prox}_h(x) &\Longleftrightarrow x - u \in \partial h(u) \\ &\Longleftrightarrow u \in \partial h^*(x - u) \\ &\Longleftrightarrow x - u = \operatorname{prox}_{h^*}(x) \end{aligned}$$

- 可以由此推出，子空间的正交投影的广义分解式:

$$x = P_L(x) + P_{L^\perp}(x)$$

L为一个子空间，$L^\perp$是它的正交补
（在Moreau分解中有$h = I_L$, $h^* = I_{L^\perp}$,其中I表示示性函数）

广义Moreau分解

定理

对任意的$\lambda > 0$，我们有广义的Moreau分解式：

$$x = \text{prox}_{\lambda f}(x) + \lambda \text{prox}_{\lambda^{-1} f^*}(x/\lambda)$$

Proof.

对 λf 应用Moreau分解

$$\begin{aligned} x &= \text{prox}_{\lambda f}(x) + \text{prox}_{(\lambda f)^*}(x) \\ &= \text{prox}_{\lambda f}(x) + \lambda \text{prox}_{\lambda^{-1} f^*}(x/\lambda) \end{aligned}$$

第二行运用了共轭函数的性质：$(\lambda f)^*(y) = \lambda f^*(y/\lambda)$ □

支撑函数

支撑函数的共轭（在闭凸集上）是示性函数：

$$f(x) = S_C(x) = \sup_{y \in C} x^T y, \qquad f^*(y) = I_C(y)$$

支撑函数的**邻近算子**：（应用Moreau 分解）

$$\begin{aligned} \mathrm{prox}_{tf}(x) &= x - t\,\mathrm{prox}_{t^{-1}f^*}(x/t) \\ &= x - tP_C(x/t) \end{aligned}$$

可以发现，支撑函数的邻近算子可以通过投影计算得到**例子**：$f(x)$ 是 x 最大的 r 个分量的和，则

$$f(x) = x_{[1]} + \cdots + x_{[r]} = S_C(x), \qquad C = \{y | 0 \le y \le 1, 1^T y = r\}$$

范数的共轭是对偶范数球的共轭函数：

$$f(x) = \|c\|, \qquad f^*(x) = I_B(y) \qquad (B = \{y \mid \|y\|_* \leq 1\})$$

范数的**邻近算子**：（应用Moreau 分解）

$$\begin{aligned} \mathrm{prox}_{tf}(x) &= x - t\mathrm{prox}_{t^{-1}f^*}(x/t) \\ &= x - tP_B(x/t) \\ &= x - P_{tB}(x) \end{aligned}$$

当$tB = \{x \mid \|x\| \leq t\}$ 容易计算时，可以用这个公式高效地计算$\mathrm{prox}_{t\|\cdot\|}$

单点距离

距离（一般范数意义下）

$$f(x) = \|x - a\|$$

邻近算子： 令$g(x) = \|x\|$

$$\begin{aligned}
\operatorname{prox}_{tf}(x) &= a + \operatorname{prox}_{tg}(x - a) \\
&= a + x - a - tP_B\left(\frac{x-a}{t}\right) \\
&= x - P_{tB}(x - a)
\end{aligned}$$

B 定义同上一页

Euclid 距离（对于闭凸集C）

$$d(x) = \inf_{y \in C} \|x - y\|_2$$

距离的邻近算子

$$\mathrm{prox}_{td}(x) = \theta P_C(x) + (1-\theta)x, \qquad \theta = \begin{cases} t/d(x) & d(x) \geq t \\ 1 & otherwise \end{cases}$$

平方距离的邻近算子: $f(x) = d(x)^2/2$

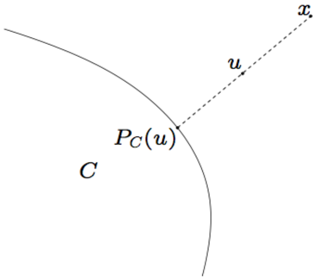
$$\mathrm{prox}_{tf}(x) = \frac{1}{1+t}x + \frac{t}{1+t}P_C(x)$$

证明（关于$\mathrm{prox}_{td}(x)$的表达式）

- 若 $u = \mathrm{prox}_{td}(x) \notin C$, 则有

$$x - u = \frac{t}{d(u)}(u - P_C(u))$$

由此可推出$P_C(u) = P_C(x)$, $d(x) \geq t$, 且 u 是 x 和 $P_C(x)$ 的加权平均



- 若 $u \in C$ 最小化 $d(u) + \frac{1}{2t}\|u - x\|_2^2$, 等价于最小化 $\frac{1}{2t}\|u - x\|_2^2$，即得 $u = P_C(x)$

证明（关于$\mathrm{prox}_{tf}(x)$的表达式，当$f(x)=d(x)^2/2$时）

$$\begin{aligned}\mathrm{prox}_{tf}(x) &= \arg\min_u \left(\frac{1}{2}d(u)^2 + \frac{1}{2t}\|u-x\|_2^2\right)\\ &= \arg\min_u \inf_{v\in C}\left(\frac{1}{2}\|u-v\|_2^2 + \frac{1}{2t}\|u-x\|_2^2\right)\end{aligned}$$

最优的 u 可以看成 v 的函数：

$$u = \frac{t}{t+1}v + \frac{1}{t+1}x$$

最优的 v 在集合 C 上极小化

$$\frac{1}{2}\left\|\frac{t}{t+1}v + \frac{1}{t+1}x - v\right\|_2^2 + \frac{1}{2t}\left\|\frac{t}{t+1}v + \frac{1}{t+1}x - x\right\|_2^2 = \frac{1}{2(1+t)}\|v-x\|_2^2$$

由此即得，$v = P_C(x)$

非凸函数的邻近算子

(适当闭函数的邻近算子) 设 h 是适当闭函数(可以非凸)，且具有有限的下界，即满足$\inf_{x\in\textbf{dom}h} h(x) > -\infty$, 定义 h 的邻近算子为

$$\operatorname{prox}_h(x) = \underset{u\in\textbf{dom}\ h}{\arg\min}\left\{h(u) + \frac{1}{2}\|u-x\|^2\right\}.$$

定理

设h 是适当闭函数且$\inf_{x\in\textbf{dom}h} h(x) > -\infty$, 则$\forall x \in \textbf{dom}\ h$, $\operatorname{prox}_h(x)$ 是$\mathbb{R}^n$ 上的非空紧集.

Proof.

定义 $g(u) = h(u) + \frac{1}{2}\|u-x\|^2$, 设 $\inf_{x\in\textbf{dom}h} h(x) = l$.
取$u_0 \in \textbf{dom}\ h$, 由于$\frac{1}{2}\|u-x\|^2$ 无上界，故$\exists R > 0$, 对$\forall$ 满足$\|u-x\| > R$ 的u, 成立$\frac{1}{2}\|u-x\|^2 > g(u_0) - l$，即$g(u) > g(u_0)$.
这说明下水平集$\{u \mid g(u) \leqslant g(u_0)\}$ 含于球$\|u-x\| \leqslant R$ 内, 即g 有一个非空有界下水平集. 显然$g(u)$ 是闭函数, 由Weierstrass 定理可知, $g(u)$ 的最小值点集合$\operatorname{prox}_h(x)$ 是非空紧集. □

非光滑非凸问题函数的次微分

前面介绍了闭凸函数的邻近算子与次梯度的关系，而对于非凸函数有类似的结论。首先回顾一下非光滑非凸函数的次微分。

次微分

设$f:\mathbb{R}^n \to (-\infty,+\infty]$ 是适当下半连续函数.

- 对给定的$x\in \textbf{dom}\, f$, 满足如下条件的所有向量$u\in\mathbb{R}^n$ 的集合定义为f 在点x 处的*Fréchet* 次微分:

$$\liminf_{y\to x, y\neq x} \frac{f(y)-f(x)-\langle u, y-x\rangle}{\|y-x\|} \geq 0,$$

  记为$\hat{\partial} f(x)$.当$x\notin \textbf{dom}\, f$ 时, 将$\hat{\partial} f(x)$ 定义为空集$\varnothing$.

- f 在点$x\in\mathbb{R}^n$ 处的极限次微分(或简称为次微分)定义为

$$\partial f(x) = \{u\in\mathbb{R}^n \ : \ \exists\, x^k \to x, f(x^k)\to f(x), u^k \in \hat{\partial} f(x^k) \to u\}.$$

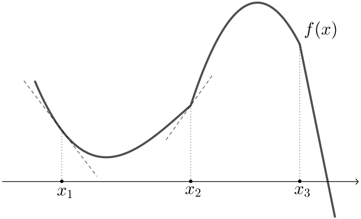
  极限次微分通过对x 附近的点处的*Fréchet* 次微分取极限得到.

- $\hat{\partial} f(x) \subseteq \partial f(x)$, 前者是闭凸集, 后者是闭集.并非在所有的$x \in \mathbf{dom}\, f$处都存在*Fréchet* 次微分.
- 凸函数的次梯度要求不等式

$$f(y) \geq f(x) + \langle g, y - x \rangle, \quad g \in \partial f(x)$$

在定义域内全局成立, 而非凸函数只要求在极限意义下成立.

- 当f 是可微函数时, *Fréchet* 次微分和次微分都退化成梯度.



如图，$f(x)$ 在 x_3 处不存在Fréchet 次微分, 但存在次微分

定理

设h 是适当闭函数(可非凸)且有下界, $u \in \mathrm{prox}_h(x)$, 则$x - u \in \partial h(u)$