

最优化方法 第十讲

Weiwen Wang(王伟文)

暨南大学

2025 年 5 月 20 日

目录

- 1 动机：投影梯度法回顾
- 2 Bregman 距离
 - Bregman 距离性质
- 3 镜像下降法 (Mirror Descent)
 - 迭代格式
 - 例子
 - 收敛性分析
 - 投影梯度法与镜像下降法对比
- 4 复合模型的镜像下降
 - 临近次梯度法
 - 收敛性分析

考虑优化问题

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}) \\ \text{s.t. } \mathbf{x} \in C \end{aligned}$$

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

假设 1

- (1) f 是适定的闭凸函数;
- (2) C 为闭合凸集且 $C \subseteq \text{int}(\text{dom } f)$;
- (3) $X^* \triangleq \arg\min_{\mathbf{x} \in C} f(\mathbf{x})$ 非空, 记最优值为 f^* ;

投影梯度法

梯度投影法迭代格式

$$\begin{aligned}
 \mathbf{x}_{k+1} &= \Pi_{\mathbf{C}}(\mathbf{x}_k - \gamma_k \mathbf{g}_k) \quad \mathbf{g}_k \in \partial f(\mathbf{x}_k) \text{ 或 } \mathbf{g}_k = \nabla f(\mathbf{x}_k). \\
 &= \arg \min_{\mathbf{x} \in \mathbf{C}} \|\mathbf{x} - (\mathbf{x}_k - \gamma_k \mathbf{g}_k)\|_2^2 \\
 &= \arg \min_{\mathbf{x} \in \mathbf{C}} \frac{1}{2\gamma_k} \|\mathbf{x} - (\mathbf{x}_k - \gamma_k \mathbf{g}_k)\|_2^2 \\
 &= \arg \min_{\mathbf{x} \in \mathbf{C}} \underbrace{f(\mathbf{x}_k) + \langle \mathbf{g}_k, \mathbf{x} - \mathbf{x}_k \rangle}_{f(\mathbf{x}) \text{ 线性化}} + \underbrace{\frac{1}{2\gamma_k} \|\mathbf{x} - \mathbf{x}_k\|_2^2}_{\text{距离的平方}} + \text{Const} \\
 &= \arg \min_{\mathbf{x} \in \mathbf{C}} \left\{ \langle \mathbf{g}_k, \mathbf{x} \rangle + \frac{1}{2\gamma_k} \|\mathbf{x} - \mathbf{x}_k\|_2^2 \right\}
 \end{aligned}$$

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

投影梯度法一般化

$$\mathbf{x}_{k+1} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{C}} \left\{ \langle \mathbf{g}_k, \mathbf{x} \rangle + \frac{1}{2\gamma_k} \operatorname{dist}_{\mathcal{C}}^2(\mathbf{x}, \mathbf{x}_k) \right\}.$$

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

定义 2

设 φ 为严格凸函数, $C \subseteq \text{dom}(\varphi)$, 且 φ 在闭凸集 C 上连续可微, 则 Bregman 距离定义为

$$B_{\varphi}(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{x}) - \varphi(\mathbf{y}) - \langle \nabla \varphi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \quad \forall \mathbf{x}, \mathbf{y} \in C.$$

Bregman 距离不是严格意义上的距离

- $B_{\varphi}(\mathbf{x}, \mathbf{y}) = B_{\varphi}(\mathbf{y}, \mathbf{x})$ 不一定成立;
- 不满足三角不等式.

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

命题 3

设 φ 为 $\|\cdot\|$ 下的 μ -强凸函数, 则

(a) $B_{\varphi}(\mathbf{x}, \mathbf{y}) \geq \frac{\mu}{2} \|\mathbf{x} - \mathbf{y}\|^2;$

(b) $B_{\varphi}(\mathbf{x}, \mathbf{y}) \geq 0;$

(c) $B_{\varphi}(\mathbf{x}, \mathbf{y}) = 0$ iff. $\mathbf{x} = \mathbf{y}.$

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror
Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

证明.

a. 由强凸函数定义

$$\varphi(\mathbf{x}) \geq \varphi(\mathbf{y}) + \langle \nabla \varphi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \frac{\mu}{2} \|\mathbf{x} - \mathbf{y}\|^2$$

故

$$\varphi(\mathbf{x}) - \varphi(\mathbf{y}) - \langle \nabla \varphi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq \frac{\mu}{2} \|\mathbf{x} - \mathbf{y}\|^2$$

即 $B_{\varphi}(\mathbf{x}, \mathbf{y}) \geq \frac{\mu}{2} \|\mathbf{x} - \mathbf{y}\|^2$.**b** 和 **c** 的结论可立即有 **a** 得到. □

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror
Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

定理 4 (三点等式)

$\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \text{dom } \varphi,$

$$B_{\varphi}(\mathbf{x}, \mathbf{z}) = B_{\varphi}(\mathbf{x}, \mathbf{y}) + B_{\varphi}(\mathbf{y}, \mathbf{z}) - \langle \nabla \varphi(\mathbf{z}) - \nabla \varphi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$$

证明.

有 Bregman 定义

$$B_{\varphi}(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{x}) - \varphi(\mathbf{y}) - \langle \nabla \varphi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$$

$$B_{\varphi}(\mathbf{y}, \mathbf{z}) = \varphi(\mathbf{y}) - \varphi(\mathbf{z}) - \langle \nabla \varphi(\mathbf{z}), \mathbf{y} - \mathbf{z} \rangle$$

由 $B_{\varphi}(\mathbf{x}, \mathbf{y}) + B_{\varphi}(\mathbf{y}, \mathbf{z})$

$$\begin{aligned} B_{\varphi}(\mathbf{x}, \mathbf{y}) + B_{\varphi}(\mathbf{y}, \mathbf{z}) &= \varphi(\mathbf{x}) - \varphi(\mathbf{z}) - \langle \nabla \varphi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle - \langle \nabla \varphi(\mathbf{z}), \mathbf{y} - \mathbf{z} \rangle \\ &= \varphi(\mathbf{x}) - \varphi(\mathbf{z}) - \langle \nabla \varphi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle - \langle \nabla \varphi(\mathbf{z}), \mathbf{x} - \mathbf{z} - (\mathbf{x} - \mathbf{y}) \rangle \\ &= B_{\varphi}(\mathbf{x}, \mathbf{z}) + \langle \nabla \varphi(\mathbf{z}) - \nabla \varphi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle. \end{aligned}$$

移项得到

$$B_{\varphi}(\mathbf{x}, \mathbf{z}) = B_{\varphi}(\mathbf{x}, \mathbf{y}) + B_{\varphi}(\mathbf{y}, \mathbf{z}) - \langle \nabla \varphi(\mathbf{z}) - \nabla \varphi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$$

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

镜像下降法

$$\mathbf{x}_{k+1} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{C}} \left\{ \langle \mathbf{g}_k, \mathbf{x} \rangle + \frac{1}{2\gamma_k} \operatorname{dist}_{\mathcal{C}}^2(\mathbf{x}, \mathbf{x}_k) \right\}.$$

取 $\frac{1}{2} \operatorname{dist}_{\mathcal{C}}^2(\mathbf{x}, \mathbf{x}_k) \triangleq B_{\varphi}(\mathbf{x}, \mathbf{x}_k)$

镜像下降法迭代格式

$$\begin{aligned} \mathbf{x}_{k+1} &= \operatorname{argmin}_{\mathbf{x} \in \mathcal{C}} \left\{ \langle \mathbf{g}_k, \mathbf{x} \rangle + \frac{1}{\gamma_k} B_{\varphi}(\mathbf{x}, \mathbf{x}_k) \right\}. \\ &= \operatorname{argmin}_{\mathbf{x} \in \mathcal{C}} \left\{ \langle \gamma_k \mathbf{g}_k, \mathbf{x} \rangle + B_{\varphi}(\mathbf{x}, \mathbf{x}_k) \right\} \\ &= \operatorname{argmin}_{\mathbf{x} \in \mathcal{C}} \left\{ \langle \gamma_k \mathbf{g}_k - \nabla \varphi(\mathbf{x}_k), \mathbf{x} \rangle + \varphi(\mathbf{x}) \right\} \quad \mathbf{g}_k \in \partial f(\mathbf{x}_k) \text{ 或 } \mathbf{g}_k = \nabla f(\mathbf{x}_k) \end{aligned}$$

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

[动机：投影梯度法回顾](#)[Bregman 距离](#)[Bregman 距离性质](#)[镜像下降法 \(Mirror Descent\)](#)[迭代格式](#)[例子](#)[收敛性分析](#)[投影梯度法与镜像下降法对比](#)[复合模型的镜像下降](#)[临近次梯度法](#)[收敛性分析](#)

命题 5

若 φ 为 μ -强凸函数, 则镜像下降迭代格式

$$\mathbf{x}_{k+1} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{C}} \{ \langle \gamma_k \mathbf{g}_k - \nabla \varphi(\mathbf{x}_k), \mathbf{x} \rangle + \varphi(\mathbf{x}) \} \quad \mathbf{g}_k \in \partial f(\mathbf{x}_k) \text{ 或 } \mathbf{g}_k = \nabla f(\mathbf{x}_k)$$

是良态的, 即 $\mathbf{x}_{k+1}$ 存在且唯一.

取所有范数为 ℓ_2 -范数, $\varphi(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}\|_2^2$, $\varphi(\mathbf{x})$ 为 1-强凸函数,
 $\nabla \varphi(\mathbf{x}) = \mathbf{x}$.

$$\begin{aligned} B_{\varphi}(\mathbf{x}, \mathbf{y}) &= \frac{1}{2} \|\mathbf{x}\|_2^2 - \frac{1}{2} \|\mathbf{y}\|_2^2 - \langle \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle \\ &= \frac{1}{2} \|\mathbf{x}\|_2^2 + \frac{1}{2} \|\mathbf{y}\|_2^2 - \langle \mathbf{x}, \mathbf{y} \rangle \\ &= \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 \end{aligned}$$

此时镜像下降法迭代格式为

$$\begin{aligned} \mathbf{x}_{k+1} &= \arg \min_{\mathbf{x} \in \mathbf{C}} \left\{ \langle \gamma_k \mathbf{g}_k - \mathbf{x}_k, \mathbf{x} \rangle + \frac{1}{2} \|\mathbf{x}\|_2^2 \right\} \\ &= \arg \min_{\mathbf{x} \in \mathbf{C}} \left\{ \frac{1}{2} \|\mathbf{x} - (\mathbf{x}_k - \gamma_k \mathbf{g}_k)\|_2^2 + \text{Const} \right\} \\ &= \Pi_{\mathbf{C}}(\mathbf{x} - (\mathbf{x}_k - \gamma_k \mathbf{g}_k)) \end{aligned}$$

动机: 投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

取所有范数为 ℓ_1 范数, $C \triangleq \Delta_d$, $\varphi(\mathbf{x}) = \sum_{i=1}^d x_i \ln x_i$, 取 $0 \cdot \ln 0 = 0$, $\varphi(\mathbf{x})$ 为 ℓ_1 范数下的 1-强凸函数. 定义

$$\Delta_d^+ = \left\{ \mathbf{x} \in \mathbb{R}^d : \sum_{i=1}^d x_i = 1, x_i > 0, i \in [d] \right\}$$

$$\nabla \varphi(\mathbf{y}) = (1 + \ln y_i)_{i \in [d]} \quad \forall \mathbf{y} \in \Delta_n^+$$

$$\begin{aligned} B_\varphi(\mathbf{x}, \mathbf{y}) &= \sum_{i=1}^d x_i \ln x_i - \sum_{i=1}^d y_i \ln y_i - \sum_{i=1}^d (x_i - y_i)(1 + \ln y_i) \\ &= \sum_{i=1}^d x_i \ln \frac{x_i}{y_i} \quad \forall \mathbf{x} \in \Delta_d, \forall \mathbf{y} \in \Delta_d^+. \end{aligned}$$

动机: 投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

记 $\mathbf{y} = \mathbf{x}_k, \mathbf{g}_k \in \partial f(\mathbf{x}_k)$ 或 $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$, $\mathbf{g} = \mathbf{g}_k$. 对应的镜像下降法为

$$\begin{aligned}\mathbf{x}_{k+1} &= \arg \min_{\mathbf{x} \in \Delta_d} \left\{ \sum_{i=1}^d \gamma_k g_i x_i + \sum_{i=1}^d x_i \ln \frac{x_i}{y_i} \right\} \\ &= \arg \min_{\mathbf{x} \in \Delta_d} \left\{ \sum_{i=1}^d \gamma_k g_i x_i + \sum_{i=1}^d x_i \ln x_i - \sum_{i=1}^d x_i \ln y_i \right\}\end{aligned}$$

解得

$$\mathbf{x}_{k+1,i} = \frac{\mathbf{x}_{k,i} e^{-\gamma_k \mathbf{g}_{k,i}}}{\sum_{j=1}^d \mathbf{x}_{k,j} e^{-\gamma_k \mathbf{g}_{k,j}}}.$$

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

引理 6 (镜像下降基本不等式)

设 φ 为 $\|\cdot\|$ 下的 μ -强凸函数, $\mathbf{x}^*$ 为约束优化问题的最优解, 则镜像下降法的迭代序列 $\{\mathbf{x}_k\}$ 满足

$$\gamma_k(f(\mathbf{x}_k) - f^*) \leq B_\varphi(\mathbf{x}^*, \mathbf{x}_k) - B_\varphi(\mathbf{x}^*, \mathbf{x}_{k+1}) + \frac{\gamma_k^2}{2\mu} \|\mathbf{g}_k\|_*^2 \quad \forall k \geq 0,$$

其中 $\|\cdot\|_*$ 表示对偶范数.

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

证明.

由三点等式

$$B_{\varphi}(\mathbf{x}^*, \mathbf{x}_k) = B_{\varphi}(\mathbf{x}^*, \mathbf{x}_{k+1}) + B_{\varphi}(\mathbf{x}_{k+1}, \mathbf{x}_k) - \langle \nabla \varphi(\mathbf{x}_k) - \nabla \varphi(\mathbf{x}_{k+1}), \mathbf{x}^* - \mathbf{x}_{k+1} \rangle$$

同时由最优性条件

$$\langle \gamma_k \mathbf{g}_k - \nabla \varphi(\mathbf{x}_k) + \nabla \varphi(\mathbf{x}_{k+1}), \mathbf{x}^* - \mathbf{x}_{k+1} \rangle \geq 0.$$

即

$$\langle \gamma_k \mathbf{g}_k, \mathbf{x}^* - \mathbf{x}_{k+1} \rangle \geq \langle \varphi(\mathbf{x}_k) - \varphi(\mathbf{x}_{k+1}), \mathbf{x}^* - \mathbf{x}_{k+1} \rangle$$



证明.

因此

$$B_{\varphi}(\mathbf{x}_{k+1}, \mathbf{x}_k) - \langle \gamma_k \mathbf{g}_k, \mathbf{x}^* - \mathbf{x}_{k+1} \rangle \leq B_{\varphi}(\mathbf{x}^*, \mathbf{x}_k) - B_{\varphi}(\mathbf{x}^*, \mathbf{x}_{k+1})$$

又因为 $B_{\varphi}(\mathbf{x}_k, \mathbf{x}_{k+1}) \geq \frac{\mu}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2$

$$B_{\varphi}(\mathbf{x}_{k+1}, \mathbf{x}_k) - \langle \gamma_k \mathbf{g}_k, \mathbf{x}^* - \mathbf{x}_{k+1} \rangle = B_{\varphi}(\mathbf{x}_{k+1}, \mathbf{x}_k) - \langle \gamma_k \mathbf{g}_k, \mathbf{x}_k - \mathbf{x}_{k+1} \rangle + \langle \gamma_k \mathbf{g}_k, \mathbf{x}_k - \mathbf{x}^* \rangle$$

$$\begin{aligned} &\geq \frac{\mu}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 - \frac{\gamma^2}{2\mu} \|\mathbf{g}_k\|_*^2 - \frac{\mu}{2} + \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 + \text{投影梯度法与镜像下降法对比} \\ &\quad - \frac{\gamma^2}{2\mu} \|\mathbf{g}_k\|_*^2 + \gamma_k (f(\mathbf{x}_k) - f(\mathbf{x}^*)) \end{aligned}$$

综上,

$$\gamma_k (f(\mathbf{x}_k) - f(\mathbf{x}^*)) \leq B_{\varphi}(\mathbf{x}^*, \mathbf{x}_k) - B_{\varphi}(\mathbf{x}^*, \mathbf{x}_{k+1}) + \frac{\gamma_k^2}{2\mu} \|\mathbf{g}_k\|_*^2$$

动机: 投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

由镜像下降基本不等式可直接得到

定理 7

设 φ 为 $\|\cdot\|$ 下的 μ -强凸函数, $\mathbf{x}^*$ 为约束优化问题的最优解, 则镜像下降法的迭代序列 $\{\mathbf{x}_k\}$ 满足

$$\min_{k=0,1,\dots,K} f(\mathbf{x}_k) - f^* \leq \frac{B_\varphi(\mathbf{x}^*, \mathbf{x}_0) + \sum_{k=0}^K \frac{\gamma_k^2}{2\mu} \|\mathbf{g}_k\|_*^2}{\sum_{k=0}^K \gamma_k} \quad \forall K \geq 0.$$

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

若 $D = \max_{\mathbf{x} \in \mathcal{C}} B_\varphi(\mathbf{x}, \mathbf{x}_0)$, $\forall \mathbf{g}_k$, $\|\mathbf{g}_k\|_* \leq L_f$, 取

$$\gamma_k = \frac{\sqrt{2D\mu}}{L_f\sqrt{K+1}}$$

有

$$\min_{k=0,1,\dots,K} f(\mathbf{x}_k) - f^* \leq \frac{\sqrt{2DL_f}}{\sqrt{\mu(K+1)}}.$$

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

考虑约束优化问题

$$\min_{\mathbf{x} \in \Delta_d} f(\mathbf{x})$$

其中 f 为适定闭凸函数, 且 $\Delta_d \subseteq \text{int}(\text{dom } f)$, 取初始点

$$\mathbf{x}_0 = \frac{1}{d} \mathbf{1} = (1/d, \dots, 1/d)$$

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

欧氏空间

取 $\|\cdot\| = \|\cdot\|_2$, 此时 $\|\cdot\|_* = \|\cdot\|_2$, 令 $\varphi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$, φ 为 1-强凸函数,
 $B_\varphi(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2$.

镜像下降法即投影梯度法, 迭代格式

$$\mathbf{x}_{k+1} = \Pi_{\Delta_d}(\mathbf{x}_k - \gamma_k \mathbf{g}_k) \quad \mathbf{g}_k \in \partial f(\mathbf{x}_k) \text{ 或 } \mathbf{g}_k = \nabla f(\mathbf{x}_k).$$

$$D = \max_{\mathbf{x} \in \Delta_d} B_\varphi(\mathbf{x}, \mathbf{x}_0) = \max_{\mathbf{x} \in \Delta_d} \frac{1}{2}\|\mathbf{x} - \frac{1}{d}\mathbf{1}\|_2^2 = \frac{1}{2}\left(1 - \frac{1}{d}\right) \leq \frac{1}{2},$$

若

$$\|\mathbf{g}\|_2 \leq L_{f,2}, \forall \mathbf{x} \in \Delta_d, \forall \mathbf{g} \in \partial f(\mathbf{x})$$

$$\min_{k=0,1,\dots,K} f(\mathbf{x}_k) - f^* \leq \frac{L_{f,2}}{\sqrt{K+1}}.$$

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

非欧氏空间

取 $\|\cdot\| = \|\cdot\|_1$, $\|\cdot\|_* = \|\cdot\|_\infty$, $\varphi(\mathbf{x}) = \sum_{i=1}^d x_i \log x_i$, φ 为 $\|\cdot\|_1$ 下的 1-强凸函数. 镜像下降法迭代格式为

$$\mathbf{x}_{k+1,i} = \frac{\mathbf{x}_{k,i} e^{-\gamma_k \mathbf{g}_{k,i}}}{\sum_{j=1}^d \mathbf{x}_{k,j} e^{-\gamma_k \mathbf{g}_{k,j}}} \quad \mathbf{g}_k \in \partial f(\mathbf{x}_k), i \in [d].$$

$$D = \max_{\mathbf{x} \in \Delta_d} B_\varphi(\mathbf{x}, \mathbf{x}_0) = \max_{\mathbf{x} \in \Delta_d} \sum_{i=1}^d x_i \ln(d \cdot x_i) = \ln d$$

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

若

$$\|\mathbf{g}\|_{\infty} \leq L_{f,2}, \forall \mathbf{x} \in \Delta_d, \forall \mathbf{g} \in \partial f(\mathbf{x})$$

$$\min_{k=0,1,\dots,K} f(\mathbf{x}_k) - f^* \leq \frac{\sqrt{2 \log d} L_{f,\infty}}{\sqrt{K+1}}.$$

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

"Worst-Case" 收敛性对比

$$\text{记 } C_2^f = \frac{L_{f,2}}{\sqrt{K+1}}, C_\infty^f = \frac{\sqrt{2\log d} L_{f,\infty}}{\sqrt{K+1}},$$

$$\rho^f = \frac{C_\infty^f}{C_2^f} = \sqrt{2\log d} \frac{L_{f,\infty}}{L_{f,2}}$$

因为 $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{d}\|\mathbf{x}\|_\infty, \forall \mathbf{x} \in \mathbb{R}^d$, 故

$$\frac{1}{\sqrt{d}} \leq \frac{L_{f,\infty}}{L_{f,2}} \leq 1$$

因此

$$\frac{\sqrt{2\log d}}{\sqrt{d}} \leq \rho^f \leq \sqrt{2\log d}.$$

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

$$\frac{\sqrt{2\log d}}{\sqrt{d}} \leq \rho^f \leq \sqrt{2\log d}.$$

这意味着存在最坏的情形, 镜像下降法比投影梯度法快 $\mathcal{O}(\sqrt{d/\ln d})$ 倍.

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

考虑复合优化问题

$$\min_{\mathbf{x} \in \text{dom } h} F(\mathbf{x}) \triangleq f(\mathbf{x}) + h(\mathbf{x})$$

假设 8

- $f, h: \mathbb{E} \rightarrow (-\infty, +\infty]$ 为适定闭凸函数
- $\text{dom } h \subseteq \text{int}(\text{dom } f)$
- $\|\mathbf{g}\|_* \leq L_f, \forall \mathbf{x} \in \text{dom } h, \forall \mathbf{g} \in \partial f(\mathbf{x})$
- $X^* \triangleq \argmin_{\mathbf{x} \in C} F(\mathbf{x})$ 非空, 记最优值为 F^*

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

为方便起见 f 和 h 的次微分简记为 f' 和 g' .

解法一：直接套用镜像下降法

$$\mathbf{x}_{k+1} = \arg \min_{\text{dom } h} \{ \langle \gamma_k(f'(\mathbf{x}_k) + h'(\mathbf{x}_k)) - \nabla \varphi(\mathbf{x}_k), \mathbf{x} \rangle + \varphi(\mathbf{x}) \}$$

- $\text{dom } h$ 不一定是闭集, 迭代格式不一定是适定的
- 收敛性依赖 $F(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$ 的 Lipschitz 常数, 首先没有假设 $g(\mathbf{x})$ 是 Lipschitz 连续; 其次, 即使 $g(\mathbf{x})$ 是 Lipschitz 连续, 两个函数相加得到更大的 Lipschitz 常数, 弱化收敛性.

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

解法二 (Mirror-C):

$$\begin{aligned}\mathbf{x}_{k+1} &= \arg \min_{\text{dom } h} \left\{ \langle \gamma_k f'(\mathbf{x}_k) - \nabla \varphi(\mathbf{x}_k), \mathbf{x} \rangle + \gamma_k h(\mathbf{x}) + \varphi(\mathbf{x}) \right\} \\ &= \arg \min_{\text{dom } h} \left\{ \langle \gamma_k f'(\mathbf{x}_k), \mathbf{x} \rangle + \gamma_k h(\mathbf{x}) + B_\varphi(\mathbf{x}, \mathbf{x}_k) \right\}\end{aligned}$$

[动机：投影梯度法回顾](#)[Bregman 距离](#)[Bregman 距离性质](#)[镜像下降法 \(Mirror Descent\)](#)[迭代格式](#)[例子](#)[收敛性分析](#)[投影梯度法与镜像下降法对比](#)[复合模型的镜像下降](#)[临近次梯度法](#)[收敛性分析](#)

引理 9

设 $f: \mathbb{E} \rightarrow (-\infty, +\infty]$ 为闭凸函数, 且为 μ -强凸函数, 则 f 有唯一最小值点.

该引理表明 Mirror-C 是良态的, 若 $\varphi(\mathbf{x})$ 为闭凸函数, 且为 μ -强凸函数.

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

取 $\varphi(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}\|_2^2$, $B_\varphi(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2$

Mirror-C

$$\begin{aligned}
 \mathbf{x}_{k+1} &= \operatorname{argmin}_{\mathbf{x} \in \operatorname{dom} h} \left\{ \langle \gamma_k f'(\mathbf{x}_k), \mathbf{x} \rangle + \gamma_k h(\mathbf{x}) + B_\varphi(\mathbf{x}, \mathbf{x}_k) \right\} \\
 &= \operatorname{argmin}_{\mathbf{x} \in \operatorname{dom} h} \left\{ \langle \gamma_k f'(\mathbf{x}_k), \mathbf{x} \rangle + \gamma_k h(\mathbf{x}) + \frac{1}{2} \|\mathbf{x} - \mathbf{x}_k\|_2^2 \right\} \\
 &= \operatorname{argmin}_{\mathbf{x} \in \operatorname{dom} h} \left\{ \gamma_k h(\mathbf{x}) + \frac{1}{2} \|\mathbf{x} - (\mathbf{x}_k - \gamma_k f'(\mathbf{x}_k))\|_2^2 \right\} \\
 &\triangleq \underbrace{\operatorname{prox}_{\gamma_k h(\mathbf{x})}(\mathbf{x}_k - \gamma_k f'(\mathbf{x}_k))}_{\text{Proximal Operator}}
 \end{aligned}$$

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

定理 10

若假设 8 成立且 $h(\mathbf{x})$ 为非负函数, φ 为 $\|\cdot\|$ 下的 μ -强凸函数. 设 $\mathbf{x}^*$ 为复合优化问题的最优解, 则 Mirror-C 迭代序列 $\{\mathbf{x}_k\}_{k=0}^K$ 满足

$$\min_{k=0,1,\dots,K} F(\mathbf{x}_k) - F^* \leq \frac{\gamma_0 h(\mathbf{x}_0) + B_\varphi(\mathbf{x}^*, \mathbf{x}_0) + \frac{1}{2\mu} \sum_{k=0}^K \gamma_k^2 \|f'(\mathbf{x}_k)\|_*^2}{\sum_{k=0}^K \gamma_k}$$

证明.

由三点等式

$$B_{\varphi}(\mathbf{x}^*, \mathbf{x}_k) = B_{\varphi}(\mathbf{x}^*, \mathbf{x}_{k+1}) + B_{\varphi}(\mathbf{x}_k, \mathbf{x}_{k+1}) - \langle \nabla \varphi(\mathbf{x}_k) - \nabla \varphi(\mathbf{x}_{k+1}), \mathbf{x}^* - \mathbf{x}_{k+1} \rangle$$

又因为

$$\mathbf{x}_{k+1} = \operatorname{argmin}_{\mathbf{x} \in \operatorname{dom} h} \{ \langle \gamma_k f'(\mathbf{x}_k) - \nabla \varphi(\mathbf{x}_k), \mathbf{x} \rangle + \gamma_k h(\mathbf{x}) + \varphi(\mathbf{x}) \}$$

由最优性条件, 存在 $h'(\mathbf{x}_{k+1}) \in \partial h(\mathbf{x}_{k+1})$ 使得

$$\gamma_k f'(\mathbf{x}_k) - \nabla \varphi(\mathbf{x}_k) + \gamma_k h'(\mathbf{x}_{k+1}) + \nabla \varphi(\mathbf{x}_{k+1}) = 0$$

即

$$\nabla \varphi(\mathbf{x}_k) - \nabla \varphi(\mathbf{x}_{k+1}) = \gamma_k f'(\mathbf{x}_k) + \gamma_k h'(\mathbf{x}_{k+1})$$

动机: 投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror

Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

续.

故

$$\begin{aligned}
 B_{\varphi}(\mathbf{x}^*, \mathbf{x}_k) - B_{\varphi}(\mathbf{x}^*, \mathbf{x}_{k+1}) &= B_{\varphi}(\mathbf{x}_{k+1}, \mathbf{x}_k) - \langle \gamma_k f'(\mathbf{x}_k) + \gamma_k h'(\mathbf{x}_{k+1}), \mathbf{x}^* - \mathbf{x}_{k+1} \rangle \\
 &= \frac{\mu}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 - \langle \gamma_k f'(\mathbf{x}_k), \mathbf{x}^* - \mathbf{x}_k + \mathbf{x}_k - \mathbf{x}_{k+1} \rangle - \langle \gamma_k h'(\mathbf{x}_{k+1}), \mathbf{x}^* - \mathbf{x}_{k+1} \rangle \\
 &\geq \gamma_k (f(\mathbf{x}_k) - f(\mathbf{x}^*)) - \frac{\gamma_k^2}{2\mu} \|f'(\mathbf{x}_k)\|_*^2 + \gamma_k (h(\mathbf{x}_k) - h(\mathbf{x}^*)) - \gamma_k (h(\mathbf{x}_k) - h(\mathbf{x}_{k+1}))
 \end{aligned}$$

又因为 $\gamma_k \geq \gamma_{k+1}$, 移项并整理得到

$$\gamma_k (F(\mathbf{x}_k) - F(\mathbf{x}^*)) \leq B_{\varphi}(\mathbf{x}^*, \mathbf{x}_k) - B_{\varphi}(\mathbf{x}^*, \mathbf{x}_{k+1}) + \frac{\gamma_k^2}{2\mu} \|f'(\mathbf{x}_k)\|_*^2 + \gamma_k h(\mathbf{x}_k) - \gamma_{k+1} h(\mathbf{x}_{k+1})$$

□

动机：投影梯度法回顾

Bregman 距离

Bregman 距离性质

镜像下降法 (Mirror Descent)

迭代格式

例子

收敛性分析

投影梯度法与镜像下降法对比

复合模型的镜像下降

临近次梯度法

收敛性分析

稳健回归

考虑优化问题

$$\min_{\mathbf{x} \in \mathbb{R}^d} \left\{ F(\mathbf{x}) \triangleq \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1 + \lambda \|\mathbf{x}\|_1 \right\}$$

- 投影次梯度法, $\gamma_k = \frac{1}{\|F'(\mathbf{x}_k)\|_2 \sqrt{k+1}}$ (解法一)

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \gamma_k \left(\mathbf{A}^T \text{sgn}(\mathbf{A}\mathbf{x}_k - \mathbf{b}) + \lambda \text{sgn}(\mathbf{x}_k) \right)$$

- 临近次梯度法 (Mirror-C: $\varphi(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}\|_2^2$), $\gamma_k = \frac{1}{\|f'(\mathbf{x}_k)\|_2 \sqrt{k+1}}$ (解法二)

$$\mathbf{x}_{k+1} = \text{prox}_{\lambda \gamma_k \|\cdot\|_1} \left(\mathbf{x}_k - \gamma_k \left(\mathbf{A}^T \text{sgn}(\mathbf{A}\mathbf{x}_k - \mathbf{b}) \right) \right)$$