近似点梯度法

宋晓良

大连理工大学数学科学学院

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

提纲

1. 近似点梯度法

2. 应用
   - LASSO问题

3. 收敛性分析

复合优化问题

我们将考虑如下复合优化问题：

$$\min_{x\in\mathbb{R}^n} \quad \psi(x) = f(x) + h(x)$$

- 函数f为可微函数，其定义域$\mathbf{dom}\, f = \mathbb{R}^n$
- 函数h为凸函数，可以是非光滑的，并且邻近算子容易计算
- LASSO问题：$f(x) = \frac{1}{2}\|Ax - b\|^2$，$h(x) = \mu\|x\|_1$
- 次梯度法计算的复杂度：$\mathcal{O}(1/\sqrt{k})$

是否可以设计复杂度为$\mathcal{O}(1/k)$的算法？

邻近算子：回顾

定义

对于一个凸函数h，定义它的**邻近算子**为

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

例子

- ℓ_1 范数: $h(x) = \|x\|_1$, $\quad \mathrm{prox}_{th}(x) = \mathrm{sign}(x)\max\{|x|-t, 0\}$
- ℓ_2 范数:$h(x) = \|x\|_2$, $\quad \mathrm{prox}_{th}(x) = \begin{cases} \left(1 - \frac{t}{\|x\|_2}\right)x, & \|x\|_2 \geqslant t, \\ 0, & \text{其他}. \end{cases}$
- 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:
  $h(x) = -\sum_{i=1}^n \ln x_i$, $\quad \mathrm{prox}_{th}(x)_i = \frac{x_i + \sqrt{x_i^2+4t}}{2}$, $\quad i = 1, 2, \cdots, n$

对于光滑部分f做梯度下降，对于非光滑部分h使用邻近算子，则近似点梯度法的迭代格式为

$$x^{k+1} = \mathrm{prox}_{t_k h}\left(x^k - t_k \nabla f\left(x^k\right)\right) \tag{1}$$

其中 $t_k > 0$ 为每次迭代的步长，它可以是一个常数或者由线搜索得出.

Algorithm 1 近似点梯度法

1: 输入：函数 $f(x), h(x)$, 初始点 x^0.
2: **while** 未达到收敛准则 **do**
3: $\quad x^{k+1} = \mathrm{prox}_{t_k h}\left(x^k - t_k \nabla f\left(x^k\right)\right)$.
4: **end while**

对近似点梯度法的理解

根据定义，(1)式等价于

$$
\begin{aligned}
x^{k+1} &= \arg\min_u \left\{ h(u) + \frac{1}{2t_k} \left\| u - x^k + t_k \nabla f\left(x^k\right) \right\|^2 \right\} \\
&= \arg\min_u \left\{ h(u) + f\left(x^k\right) + \nabla f\left(x^k\right)^{\mathrm{T}} \left(u - x^k\right) + \frac{1}{2t_k} \left\| u - x^k \right\|^2 \right\}
\end{aligned}
$$

根据邻近算子与次梯度的关系，又可以形式地写成

$$
x^{k+1} = x^k - t_k \nabla f\left(x^k\right) - t_k g^k, \quad g^k \in \partial h\left(x^{k+1}\right)
$$

即对光滑部分做显式的梯度下降，关于非光滑部分做隐式的梯度下降.

步长选取

- 当 f 为梯度L -利普希茨连续函数时，可取固定步长 $t_k = t \leqslant \frac{1}{L}$.
当L 未知时可使用线搜索准则

$$f\left(x^{k+1}\right) \leqslant f\left(x^k\right) + \nabla f\left(x^k\right)^{\mathrm{T}}\left(x^{k+1} - x^k\right) + \frac{1}{2t_k}\left\|x^{k+1} - x^k\right\|^2$$

- 利用BB 步长作为 t_k 的初始估计并用非单调线搜索进行校正：

$$\alpha_{\mathrm{BB1}}^k \stackrel{\text{def}}{=} \frac{\left(s^{k-1}\right)^{\mathrm{T}} y^{k-1}}{\left(y^{k-1}\right)^{\mathrm{T}} y^{k-1}} \quad \text{或} \quad \alpha_{\mathrm{BB2}}^k \stackrel{\text{def}}{=} \frac{\left(s^{k-1}\right)^{\mathrm{T}} s^{k-1}}{\left(s^{k-1}\right)^{\mathrm{T}} y^{k-1}},$$

其中$s^{k-1} = x^k - x^{k-1}$ 以及 $y^{k-1} = \nabla f\left(x^k\right) - \nabla f\left(x^{k-1}\right)$.

提纲

1 近似点梯度法

2 应用
- LASSO问题

3 收敛性分析

LASSO问题

考虑用近似点梯度法求解 LASSO 问题

$$\min_{x} \quad \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|^2$$

令 $f(x) = \frac{1}{2}\|Ax - b\|^2, h(x) = \mu\|x\|_1$,则

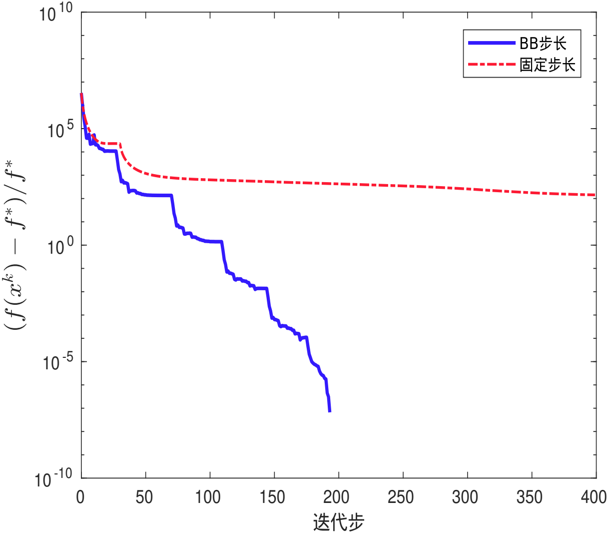
$$\begin{aligned} \nabla f(x) &= A^{\mathrm{T}}(Ax - b) \\ \operatorname{prox}_{t_k h}(x) &= \operatorname{sign}(x) \max\{|x| - t_k\mu, 0\} \end{aligned}$$

故相应的迭代格式为：

$$\begin{aligned} y^k &= x^k - t_k A^{\mathrm{T}}\left(Ax^k - b\right) \\ x^{k+1} &= \operatorname{sign}\left(y^k\right) \max\left\{\left|y^k\right| - t_k\mu, 0\right\} \end{aligned}$$

即第一步做梯度下降，第二步做收缩

我们还可以使用BB步长加速收敛



提纲

1. 近似点梯度法
2. 应用
   - LASSO问题
3. 收敛性分析

基本假设：

- f 在$\mathbb{R}^n$ 上是凸的; ∇f 为 L-利普希茨连续，即

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x - y\|, \quad \forall x, y$$

- h 是适当的闭凸函数 (因此 prox_{th} 的定义是合理的);
- 函数 $\psi(x) = f(x) + h(x)$ 的最小值 ψ^* 是有限的，并且在点x^* 处可取到(并不要求唯一).

收敛性分析

定理 1(固定步长近似点梯度法的收敛性)

取定步长为 $t_k = t \in \left(0, \frac{1}{L}\right]$，设 $\left\{x^k\right\}$ 由迭代格式(1)产生，则

$$\psi\left(x^k\right) - \psi^* \leqslant \frac{1}{2kt}\left\|x^0 - x^*\right\|^2$$

步长选取

定理**1**中要求 $t \leq \frac{1}{L}$，而根据定理证明的过程，也可以用线搜索准则：

- 从某个 $t = \hat{t} > 0$ 开始进行回溯$(t \leftarrow \beta t)$, 直到满足不等式

$$f\left(x^{k+1}\right) \leqslant f\left(x^{k}\right) + \nabla f\left(x^{k}\right)^{\mathrm{T}}\left(x^{k+1} - x^{k}\right) + \frac{1}{2t_k}\left\|x^{k+1} - x^{k}\right\|^{2} \tag{2}$$

定理 2 (非固定步长的近似点梯度法的收敛性)

从某个 $t = \hat{t} > 0$ 开始进行回溯($t \leftarrow \beta t$) 直到满足不等式(2), 设 $\{x^k\}$ 是由迭代格式(1) 产生的序列，则

$$\psi\left(x^k\right) - \psi^* \leqslant \frac{1}{2k\min\{\hat{t}, \beta/L\}} \left\|x^0 - x^*\right\|^2$$