1.5em 0pt

Homework 3

Lecturer: Xiangyu Chang *Scribe:* 吴博成

1 HW 1

1.1 问题重述

考虑以下优化问题:

$$\min_{\mathbf{x}} f(\mathrm{x})=\frac{1}{m} \sum_{i=1}^{m} f_{i}(\mathrm{x})$$

并假设 f 是 β-smooth 并 α-strong convex 的。使用固定步长 mini-bath *SGD* 来解决问题:

$$\mathrm{x}^{t+1}=\mathrm{x}^{t}-\frac{s}{n_{b}} \sum_{i_{t} \in D_{t}} \nabla f_{i_{t}}\left(\mathrm{x}^{t}\right)$$

其中 $D_t \subset \{1,2,\ldots,m\}$ 是随机取样，$|D_t| = n_b$ 是 D_t 。
我们进一步假设:
(1) D_t 不依赖于之前的 $D_0, D_1, \ldots, D_{t-1}$.
(2) $\mathbb{E}_{i_t \in D_t}\left[\nabla f_{i_t}\left(\mathrm{x}^t\right)\right]=\nabla f\left(\mathrm{x}^t\right)$ (无偏估计).
(3) $\mathbb{E}_{i_t \in D_t}\left[\left\|\nabla f_{i_t}\left(\mathrm{x}^t\right)\right\|^2\right]=\sigma^2+\left\|\nabla f\left(\mathrm{x}^t\right)\right\|^2$ (方差控制).

1.2 问题求解

1.2.1 Lemma1:

$$\mathbb{E}_{D_t}\left\|g^t\right\|^2=\frac{\sigma^2}{n_b}+\left\|\nabla f\left(\mathbf{x}^t\right)\right\|^2$$

其中

$$\mathbf{g}^t=\frac{1}{n_b} \sum_{i \in D_t} \nabla f_i\left(\mathbf{x}^t\right)$$

证明:
因为:

$$E_{i_t \in D_t}\left[\left\|\nabla f_{i_t}\left(\mathbf{x}^t\right)\right\|^2\right]=\sigma^2+\left\|\nabla f\left(\mathbf{x}^t\right)\right\|^2$$

所以:

$$E_{i_t \in D_t}\left[\left\|\nabla f_{i_t}\left(\mathbf{x}^t\right)\right\|^2-\left\|\nabla f\left(\mathbf{x}^t\right)\right\|^2\right]=\sigma^2$$

上述抽样过程可以被看作两步，分别为 (1): 从 1,2,...,m 中抽取 D_t 集合；(2): 从 D_t 集合中抽取 i_t. 因为第二步抽样有 n_b 种方法，用求均值代替该抽样则消除了该项随机性，则有:

$$E_{D_t}\left[n_b\left(\left\|\frac{1}{n_b} \sum_{i \in D_t} \nabla f_i\left(x^t\right)\right\|^2-\left\|\nabla f\left(\mathbf{x}^t\right)\right\|^2\right)\right]=\sigma^2$$

1.2.2 Lemma2:

$$\mathbb{E}_{D_t}\left[f\left(\mathbf{x}^{t+1}\right)\right] \leq f\left(\mathbf{x}^t\right)-s \nabla f\left(\mathbf{x}^t\right)^{\top} \mathbb{E}_{D_t}\left[\mathbf{g}^t\right]+\frac{\beta s^2}{2} \mathbb{E}_{D_t}\left[\left\|\mathbf{g}^t\right\|^2\right]$$

证明:

因为 f 是 β-smooth 的，所以由定义知:

$$\|\nabla f(x)-\nabla f(y)\| \leq \beta\|x-y\|$$

先由 β-smooth 证明:

$$f(y) \leq f(x)+\langle\nabla f(x), y-x\rangle+\beta\|y-x\|^2$$

构建 $h(t)=f(x+t(y-x))$，则有:

$$f(y)-f(x)=f(1)-f(0)=\int_0^1 h^{'}(t) d t=\int_0^1\langle\nabla f(x+t(y-x)), y-x\rangle d t$$

所以:

$$\begin{aligned}
f(y)-f(x)-\langle\nabla f(x), y-x\rangle & =\int_0^1\langle\nabla f(x+t(y-x))-\nabla f(x), y-x\rangle d t \\
& \leq \int_0^1\|\nabla f(x+t(y-x))-\nabla f(x)\|\|\nabla y-x\| d t \\
& \leq \int_0^1 t \beta\|y-x\|^2 d t \\
& =\frac{\beta}{2}\|y-x\|^2
\end{aligned}$$

证得:

$$f(y) \leq f(x)+\langle\nabla f(x), y-x\rangle+\beta\|y-x\|^2$$

由 mini-bath SGD 的迭代方程: $\mathrm{x}^{t+1}=\mathrm{x}^t-\frac{s}{n_b} \sum_{i_t \in D_t} \nabla f_{i_t}\left(\mathrm{x}^t\right)$ 可知:

$$\begin{aligned}
& f(x^{t+1}) \leq f(x^t)+<\nabla f(x^t), x^{t+1}-x^t>+\frac{\beta}{2}\|x^{t+1}-x^t\|^2 \\
\Longleftrightarrow & f(x^{t+1}) \leq f(x^t)-s<\nabla f(x^t), \frac{1}{n_b} \sum_{i_t \in D_t} \nabla f_{i_t}\left(\mathrm{x}^t\right)>+\frac{\beta s^2}{2}\|\frac{1}{n_b} \sum_{i_t \in D_t} \nabla f_{i_t}\left(\mathrm{x}^t\right)\|^2
\end{aligned}$$

令 $\mathbf{g}^t=\frac{1}{n_b} \sum_{i \in D_t} \nabla f_i\left(\mathbf{x}^t\right)$, 对两边同时做 D_t 上的期望:

$$\mathbb{E}_{D_t}\left[f\left(\mathbf{x}^{t+1}\right)\right] \leq f\left(\mathbf{x}^t\right)-s \nabla f\left(\mathbf{x}^t\right)^{\top} \mathbb{E}_{D_t}\left[\mathbf{g}^t\right]+\frac{\beta s^2}{2} \mathbb{E}_{D_t}\left[\left\|\mathbf{g}^t\right\|^2\right]$$

证毕

1.2.3 Lemma3:

$$\mathbb{E}_{D_t}\left[f\left(\mathrm{x}^{t+1}\right)-f\left(\mathrm{x}^{t}\right)\right] \leq -\left(s-\frac{\beta s^2}{2}\right)\left\|\nabla f\left(\mathrm{x}^{t}\right)\right\|^2+\frac{\beta s^2}{2n_b}\sigma^2$$

证明:

由 Lemma2 可知:

$$\begin{aligned}\mathbb{E}_{D_t}\left[f\left(\mathrm{x}^{t+1}\right)-f\left(\mathrm{x}^{t}\right)\right] &= \mathbb{E}_{D_t}\left[f\left(\mathbf{x}^{t+1}\right)\right]-f\left(\mathbf{x}^{t}\right)\\ &\leq -s\nabla f\left(\mathbf{x}^{t}\right)^{\top}\mathbb{E}_{D_t}\left[\mathbf{g}^{t}\right]+\frac{\beta s^2}{2}\mathbb{E}_{D_t}\left[\left\|\mathbf{g}^{t}\right\|^2\right]\end{aligned}$$

由 Lemma1 可知:

$$\mathbb{E}_{D_t}\left\|g^{t}\right\|^2=\frac{\sigma^2}{n_b}+\left\|\nabla f\left(\mathbf{x}^{t}\right)\right\|^2$$

因为:

$$\mathbb{E}_{D_t}\left[\mathbf{g}^{t}\right]=\mathbb{E}_{i_t}\left[\nabla f_{i_t}(x^t)\right]=\nabla f(x^t)$$

故有:

$$\mathbb{E}_{D_t}\left[f\left(\mathrm{x}^{t+1}\right)-f\left(\mathrm{x}^{t}\right)\right] \leq -\left(s-\frac{\beta s^2}{2}\right)\left\|\nabla f\left(\mathrm{x}^{t}\right)\right\|^2+\frac{\beta s^2}{2n_b}\sigma^2$$

证毕

1.2.4 Lemma4:

$$\mathbb{E}\left[f\left(\mathrm{x}^{t+1}\right)-f^{*}\right]-\frac{\beta s}{2n_b\alpha(2-\beta s)}\sigma^2 \leq (1-\alpha s(2-\beta s))\left[\mathbb{E}\left[f\left(\mathrm{x}^{t}\right)-f^{*}\right]-\frac{\beta s}{2n_b\alpha(2-\beta s)}\sigma^2\right]$$

证明:

因为 f 是 α-strong convex 的，由定义知:

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

所以:

$$\begin{aligned}f(x^*) &\geq \min_y\{f(x)+\langle\nabla f(x), y-x\rangle+\frac{\alpha}{2}\|x-y\|^2\}\\ &= f(x)-\frac{1}{2\alpha}\|\nabla f(x)\|^2\end{aligned}$$

所以:

$$f(x)-f^{*} \leq \frac{1}{2\alpha}\|\nabla f(x)\|^2$$

由 Lemma3 可知：可知：

$$
\begin{aligned}
\mathbb{E}_{D_t}\left[f\left(\mathbf{x}^{t+1}\right)-f\left(\mathbf{x}^t\right)\right] &\leq -\left(s-\frac{\beta s^2}{2}\right)\left\|\nabla f\left(\mathbf{x}^t\right)\right\|^2+\frac{\beta s^2}{2n_b}\sigma^2 \\
&= -\frac{s}{2}\left(2-\beta s\right)\left\|\nabla f\left(\mathbf{x}^t\right)\right\|^2+\frac{\beta s^2}{2n_b}\sigma^2 \\
&\leq \frac{\beta s^2}{2n_b}\sigma^2-\alpha s\left(2-\beta s\right)\left(f(x^t)-f^*\right)
\end{aligned}
$$

所以：

$$
\mathbb{E}_{D_t}\left[f\left(\mathbf{x}^{t+1}\right)-f^*\right] \leq \frac{\beta s^2}{2n_b}\sigma^2+(1-\alpha s\left(2-\beta s\right))(f(x^t)-f^*)
$$

两边同时加上 $-\frac{\beta s}{2n_b\alpha(2-\beta s)}\sigma^2$, 则有：

$$
\mathbb{E}_{D_t}\left[f\left(\mathbf{x}^{t+1}\right)-f^*\right]-\frac{\beta s}{2n_b\alpha(2-\beta s)}\sigma^2 \leq (1-\alpha s\left(2-\beta s\right))(f(x^t)-f^*-\frac{\beta s}{2n_b\alpha(2-\beta s)}\sigma^2)
$$

两边同时对 x^t 求期望：

$$
\mathbb{E}\left[f\left(\mathrm{x}^{t+1}\right)-f^*\right]-\frac{\beta s}{2n_b\alpha(2-\beta s)}\sigma^2 \leq (1-\alpha s(2-\beta s))\left[\mathbb{E}\left[f\left(\mathrm{x}^t\right)-f^*\right]-\frac{\beta s}{2n_b\alpha(2-\beta s)}\sigma^2\right]
$$

证毕

2 HW 2

2.1 问题重述

写出 LASSO 问题的 *BCD* 算法.

2.2 问题求解

假设 $\mathbf{x} = (x_1, x_2, ..., x_n)^\top \in \mathbb{R}^n$， $\mathbf{A} = [\mathbf{A_1}, \mathbf{A_2}, ..., \mathbf{A_n}] \subset \mathbb{R}^{m\times n}$
LASSO Problem:

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

先观察 x_1^{t+1}，固定 $x_2^t, x_3^t, ..., x_n^t$，则 $x_1^{t+1} = arg\min_{\mathbf{x}}\{\frac{1}{2}\|A_1x_1 - b + \sum_{i=2}^n x_i\|^2 + \lambda\|x_1\|_1\}$
定义 $b^t = b - \sum_{i=2}^n x_i^t$，则 $x_1^{t+1} = arg\min_{\mathbf{x}}\{\frac{1}{2}\|A_1x_1 - b^t\|^2 + \lambda\|x_1\|_1\}$
对 x_1^t 分类讨论：
当 $x_1 > 0$，$0 \in (\lambda + A_1^\top(A_1x_1 - b^t))$，则 $x^{t+1} = (A_1^\top A_1)^{-1}(A_1^\top b^t - \lambda), A_1^\top b^t > \lambda$
当 $x_1 = 0$，$0 \in \partial(\frac{1}{2}\|b^t\|^2)$ 恒成立，则 $x^{t+1} = 0, -\lambda \leq A_1^\top b^t \leq \lambda$
当 $x_1 < 0$，$0 \in (-\lambda + A_1^\top(A_1x_1 - b^t))$，则 $x^{t+1} = (A_1^T A_1)^{-1}(A_1^T b^t + \lambda), A_1^\top b^t < -\lambda$

所以，

$$x_i^{t+1} = \begin{cases} \left(A_i^\top A_1\right)^{-1}\left(A_1^\top b^t - \lambda\right) & , A_i^\top b^t > \lambda \\ 0 & , -\lambda \leq A_i^\top b^t \leq \lambda \\ \left(A_i^\top A_i\right)^{-1}\left(A_i^\top b^t + \lambda\right) & , A_i^\top b^t < -\lambda \end{cases}$$

对 $i = 1, 2, ..., n$ 成立。
所以 BCD 算法如下：

Algorithm 1 Block Coodinate Descent for LASSO

Input: 给定的初始迭代点 $x^0 = (x_1^0, x_2^0, ..., x_n^0), t = 0$
  for $t = 0, 1, ..., T$ do **do**
    for $i = 0, 1, ..., n$ do **do**

$$x_i^{t+1} = \begin{cases} \left(A_i^\top A_1\right)^{-1}\left(A_1^\top b^t - \lambda\right) & , A_i^\top b^t > \lambda \\ 0 & , -\lambda \leq A_i^\top b^t \leq \lambda \\ \left(A_i^\top A_i\right)^{-1}\left(A_i^\top b^t + \lambda\right) & , A_i^\top b^t < -\lambda \end{cases}$$

    end for
  end for
  return x^T, z^T

3 HW 3

3.1 问题重述

写出 Fused-LASSO 问题的 $ADMM$ 算法.

3.2 问题求解

Fused LASSO $\underset{\mathbf{x}}{min}\frac{1}{2}\|\mathbf{Ax}-\mathbf{b}\|^2+\lambda\|\mathbf{Dx}\|_1$，该模型可以将较小系数压缩为零，不可以将部分系数的差分压缩为零，使相邻系数更平滑。

其中 $\mathbf{A}\in\mathbb{R}^{m\times n}$,$\mathbf{x}\in\mathbb{R}^n$,$\mathbf{D}=\begin{bmatrix}-1 & 1 & 0 & 0 & 0\\ 0 & -1 & 1 & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & -1 & 1 & 0\\ 0 & 0 & 0 & -1 & 1\end{bmatrix}\in\mathbb{R}^{(n-1)\times n}$

将原优化问题转化为 ADMM 形式：

$$\begin{aligned}\underset{\mathbf{x}}{min}\quad & \frac{1}{2}\left\|\mathbf{Ax}-\mathbf{b}\right\|^2+\lambda\left\|\mathbf{z}\right\|_1\\ s.t.\quad & \mathbf{Dx}-\mathbf{z}=\mathbf{0}\end{aligned}$$

构建增广拉格朗日函数：

$$L_\rho\left(\mathbf{x},\mathbf{z},\boldsymbol{\nu}\right)=\frac{1}{2}\left\|\mathbf{Ax}-\mathbf{b}\right\|^2+\lambda\left\|\mathbf{z}\right\|_1+\frac{\rho}{2}\|\mathbf{Dx}-\mathbf{z}\|^2+\boldsymbol{\nu}^\top(\mathbf{Dx}-\mathbf{z})$$

令 $\mathbf{u}=\frac{\boldsymbol{\nu}}{\rho}$，则 $\mathbf{u}^{t+1}=\mathbf{u}^t+\mathbf{Dx}^{t+1}-\mathbf{z}^{t+1}$

$$\mathbf{x}^{t+1}=arg\min_{\mathbf{x}}\left\{\frac{1}{2}\left\|\mathbf{Ax}-\mathbf{b}\right\|^2+\frac{\rho}{2}\left\|\mathbf{Dx}-\mathbf{z}^t+\mathbf{u}^t\right\|^2\right\}=arg\min_{\mathbf{x}}h(\mathbf{x})$$

则 $h'(\mathbf{x})=\mathbf{A}^\top(\mathbf{Ax}-\mathbf{b})+\rho\mathbf{D}^\top(\mathbf{Dx}-\mathbf{z}^t+\mathbf{u}^t)=\left(\mathbf{A}^\top A+\rho\mathbf{D}^\top\mathbf{D}\right)\mathbf{x}-(\mathbf{A}^\top\mathbf{b}+\rho\mathbf{D}^\top\mathbf{z}^t-\rho\mathbf{D}^\top\mathbf{u}^t)=0$

因为 $\rho>0$，所以 $\mathbf{A}^\top\mathbf{A}+\rho\mathbf{D}^\top\mathbf{D}$ 可逆

$$\textbf{Step1：}\mathbf{x}^{t+1}=arg\min_{\mathbf{x}}h(\mathbf{x})=\left(\mathbf{A}^\top A+\rho\mathbf{D}^\top\mathbf{D}\right)^{-1}(\mathbf{A}^\top\mathbf{b}+\rho\mathbf{D}^\top\mathbf{z}^t-\rho\mathbf{D}^\top\mathbf{u}^t)$$

$$\mathbf{z}^{t+1}=arg\min_{\mathbf{z}}\left\{\lambda\left\|\mathbf{z}\right\|_1+\frac{\rho}{2}\left\|\mathbf{Dx}^{t+1}-\mathbf{z}+\mathbf{u}^t\right\|^2\right\}$$

若 $\mathbf{z}>0$，则 $0\in(\lambda-\rho(\mathbf{Dx}^{t+1}-\mathbf{z}+\mathbf{u}^t)),\mathbf{z}^{t+1}=\mathbf{Dx}^{t+1}+\mathbf{u}^t-\frac{\lambda}{\rho},\mathbf{Dx}^{t+1}+\mathbf{u}^t>\frac{\lambda}{\rho}$

若 $\mathbf{z}=0$，则 $0\in\partial(\frac{\rho}{2}\|\mathbf{Dx}^{t+1}+\mathbf{u}^t\|^2)$ 恒成立，$\mathbf{z}^{t+1}=0,-\frac{\lambda}{\rho}\leq\mathbf{Dx}^{t+1}+\mathbf{u}^t\leq\frac{\lambda}{\rho}$

若 $\mathbf{z}<0$，则 $0\in(-\lambda-\rho(\mathbf{Dx}^{t+1}-\mathbf{z}+\mathbf{u}^t)),\mathbf{z}^{t+1}=\mathbf{Dx}^{t+1}+\mathbf{u}^t+\frac{\lambda}{\rho},\mathbf{Dx}^{t+1}+\mathbf{u}^t<-\frac{\lambda}{\rho}$

所以，

$$\textbf{Step2：}\mathbf{z}^{t+1}=\begin{cases}\mathbf{Dx}^{t+1}+\mathbf{u}^t-\frac{\lambda}{\rho},\mathbf{Dx}^{t+1}+\mathbf{u}^t>\frac{\lambda}{\rho}\\ 0,-\frac{\lambda}{\rho}<\mathbf{Dx}^{t+1}+\mathbf{u}^t<\frac{\lambda}{\rho}\\ \mathbf{Dx}^{t+1}+\mathbf{u}^t+\frac{\lambda}{\rho},\mathbf{Dx}^{t+1}+\mathbf{u}^t<-\frac{\lambda}{\rho}\end{cases}$$

$$\textbf{Step3:}\ \mathbf{u}^{t+1} = \mathbf{u}^t + \mathbf{D}\mathbf{x}^{t+1} - \mathbf{z}^{t+1}$$

所以 ADMM 算法如下：

Algorithm 2 ADMM for Fused LASSO

Input: 给定的初始迭代点 $x^0, z^0, t = 0$

for $t = 0, 1, ..., T$ do **do**

Step1： $\mathbf{x}^{t+1} = arg\min_{\mathbf{x}} h(\mathbf{x}) = \left(\mathbf{A}^\top A + \rho \mathbf{D}^\top \mathbf{D}\right)^{-1} (\mathbf{A}^\top \mathbf{b} + \rho \mathbf{D}^\top \mathbf{z}^t - \rho \mathbf{D}^\top \mathbf{u}^t)$

Step2： $\mathbf{z}^{t+1} = \begin{cases} \mathbf{D}\mathbf{x}^{t+1} + \mathbf{u}^t - \frac{\lambda}{\rho}, \mathbf{D}\mathbf{x}^{t+1} + \mathbf{u}^t > \frac{\lambda}{\rho} \\ 0, -\frac{\lambda}{\rho} < \mathbf{D}\mathbf{x}^{t+1} + \mathbf{u}^t < \frac{\lambda}{\rho} \\ \mathbf{D}\mathbf{x}^{t+1} + \mathbf{u}^t + \frac{\lambda}{\rho}, \mathbf{D}\mathbf{x}^{t+1} + \mathbf{u}^t < -\frac{\lambda}{\rho} \end{cases}$

Step3： $\mathbf{u}^{t+1} = \mathbf{u}^t + \mathbf{D}\mathbf{x}^{t+1} - \mathbf{z}^{t+1}$

end for

return x^T, z^T

4 HW 4

4.1 问题重述

请思考以下问题的 $ADMM$ 算法:

$$\min_{\mathbf{x}} \|\mathrm{x}\|_1$$
$$\text{s.t. } A\mathrm{x} = \mathrm{b}$$

要使用指示函数 $\Omega = \{\mathrm{x} \mid A\mathrm{x} = \mathrm{b}\}$

4.2 问题求解

原优化问题:

$$\min_{\mathbf{x}} \|\mathrm{x}\|_1$$
$$\text{s.t. } A\mathrm{x} = \mathrm{b}$$

通过引入指示函数: $\delta_\Omega(z) = \begin{cases} 0, z \in \Omega = \{\mathrm{z} \mid A\mathrm{z} = \mathrm{b}\} \\ \inf, otherwise \end{cases}$ 构建成 ADMM 形式:

$$\min_{\mathbf{x}} \|\mathrm{x}\|_1 + \delta_\Omega(z)$$
$$\text{s.t. } \mathrm{x} - \mathrm{z} = 0$$

构建增广拉格朗日函数:

$$L_\rho(x, z, v) = \|\mathrm{x}\|_1 + \delta_\Omega(z) + \frac{\rho}{2}\|x - z\|^2 + v^T(x - z)$$

令 $u = \frac{v}{\rho}$, 则 $u^{t+1} = u^t + x^{t+1} - z^{t+1}$

已知近端梯度算子定义为 $prox_{rg}(z) = \arg\min_x g(x) + \frac{1}{2}\|x - z\|^2$; 投影函数的定义为 $\pi_\Omega(z) = \arg\min_{x\in\Omega} \|x - z\|^2$

由 ADMM 算法知:

$$x^{t+1} = \arg\min_x(\|\mathrm{x}\|_1 + \frac{\rho}{2}\|x - z^t + u^t\|^2) = prox_{\frac{\|\cdot\|}{\rho}}(z^t - u^t) = \begin{cases} z^t - u^t - \frac{1}{\rho} & , z^t - u^t > \frac{1}{\rho} \\ 0 & , -\frac{1}{\rho} \le z^t - u^t \le \frac{1}{\rho} \\ z^t - u^t + \frac{1}{\rho} & , z^t - u^t < -\frac{1}{\rho} \end{cases}$$

$$z^{t+1} = \arg\min_z(\delta_\Omega(z) + \frac{\rho}{2}\|x^{t+1} - z + u^t\|^2) = \arg\min_{z\in\Omega}(\frac{\rho}{2}\|x^{t+1} - z + u^t\|^2) = \pi_\Omega(x^{t+1} + u^t)$$

所以 ADMM 算法如下:

Algorithm 3 ADMM for Basis Pursuit Problem

Input: 给定的初始迭代点 $x^0, z^0, t = 0$

for $t = 0, 1, ..., T$ do **do**

Step 1:$x^{t+1} = prox_{\frac{\|\cdot\|}{\rho}}(z^t - u^t)$

Step 2:$z^{t+1} = \pi_\Omega(x^{t+1} + u^t)$

Step 3:$u^{t+1} = u^t + x^{t+1} - z^{t+1}$

end for

return x^T, z^T