Optimization Theory and Algorithm	**Homework 3 - 11/05/2021**

Homework 3

Lecturer: Xiangyu Chang — *Scribe:* 赵敬业

HW 1

Problem restatment

We consider the following optimization problem:

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

And assume that f is β-smooth and α-strong convex. Using mini-bath *SGD* with fixed learning rate to solve it as

$$\mathbf{x}^{t+1}=\mathbf{x}^{t}-\frac{s}{n_b}\sum_{i_t\in D_t}\nabla f_{i_t}\left(\mathbf{x}^{t}\right)$$

where $D_t\subset\{1,2,\ldots,m\}$ are drawn randomly and $|D_t|=n_b$ is the size of D_t. We further suppose that (1) The index D_t does not depended from the previous $D_0,D_1,\ldots,D_{t-1}$. (2) $\mathbb{E}_{i_t\in D_t}\left[\nabla f_{i_t}\left(\mathbf{x}^t\right)\right]=\nabla f\left(\mathbf{x}^t\right)$ (Unbiased Estimation). (3) $\mathbb{E}_{i_t\in D_t}\left[\left\|\nabla f_{i_t}\left(\mathbf{x}^t\right)\right\|^2\right]=\sigma^2+\left\|\nabla f\left(\mathbf{x}^t\right)\right\|^2$ (control the variance). Prove

(i) $\mathbb{E}_{D_t}\left\|g^t\right\|^2=\frac{\sigma^2}{n_b}+\left\|\nabla f\left(\mathbf{x}^t\right)\right\|^2$, where $\mathbf{g}^t=\frac{1}{n_b}\sum_{i\in D_t}\nabla f_i\left(\mathbf{x}^t\right)$.

(ii)

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

(iii)

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

(iv) Then

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

Answer

(i):

Consider that:

$$Var_{it}\left(\nabla f_{it}\left(x^t\right)_j\right)=E_{it}\left(\nabla f_{it}\left(x^t\right)_j-E\nabla f_{it}\left(x^t\right)_j\right)^2=E_{it}\left(\nabla f_{it}\left(x^t\right)_j^2\right)-\left(E_{it}\nabla f_{it}\left(x^t\right)_j\right)^2$$

so, the

$$E_{it}\left\|\nabla f_{it}\left(x^t\right)-E\nabla f_{it}\left(x^t\right)\right\|^2=E_{it}\left\|\nabla f_{it}\left(x^t\right)\right\|^2-\left\|E_{it}\nabla f_{it}\left(x^t\right)\right\|^2=\sigma^2$$

The Var_{Dt} can be written as :

$$\begin{aligned} Var_{Dt}\left(\sum_{i\in D_t}\nabla f_{it}\left(x^t\right)_j\right) &= E_{Dt}\left(\sum_{i\in D_t}\nabla f_{it}\left(x^t\right)_j - E\sum_{i\in D_t}\nabla f_{it}\left(x^t\right)_j\right)^2 \\ &= E_{Dt}\left(\sum_{i\in D_t}\nabla f_{it}\left(x^t\right)_j\right)^2 - \left(E_{Dt}\sum_{i\in D_t}\nabla f_{it}\left(x^t\right)_j\right)^2 \end{aligned}$$

Because $Var(\sum_{i=1} g(X_i)) = \sum_i (Varg(X_i))$,so Var_{Dt} can also be written as:

$$Var_{Dt}\left(\sum_{i\in D_t}\nabla f_{it}\left(x^t\right)_j\right) = n_b Var\left(\nabla f_{it}\left(x^t\right)_j\right)$$

so

$$E_{Dt}\left\|\sum\nabla f_{it}\left(x^t\right) - E\sum\nabla f_{it}\left(x^t\right)\right\|^2 = n_b\sigma^2$$

To sum up

$$\begin{aligned} &E_{Dt}\left\|\sum\nabla f_{it}\left(x^t\right) - E\sum\nabla f_{it}\left(x^t\right)\right\|^2 = n_b\sigma^2 \\ &= E_{Dt}\left\|\sum\nabla f_{it}\left(x^t\right)\right\|^2 - \left\|E_{Dt}\sum\nabla f_{it}\left(x^t\right)\right\|^2 \\ &= E_{Dt}\left\|\sum\nabla f_{it}\left(x^t\right)\right\|^2 - n_b^2\left\|\nabla f\left(x^t\right)\right\|^2 \end{aligned}$$

so :

$$E_{Dt}\left\|\frac{1}{n_b}\sum\nabla f_{it}(x^t)\right\|^2 = \frac{\sigma^2}{n_b} + \left\|\nabla f(x^t)\right\|^2$$

(ii):

$$\begin{aligned} E_{Dt}(f(x^{t+1})) &\leq E_{Dt}\left(f(x^t) + < \nabla f(x^t), x^{t+1} - x^t > + \frac{\beta}{2}\|x^{t+1} - x^t\|^2\right) \\ &= f\left(x^t\right) + E_{Dt}\left(\left\langle\nabla f\left(x^t\right), x^{t+1} - x^t\right\rangle\right) + \frac{\beta}{2}E_{Dt}\|x^{t+1} - x^t\|^2 \\ &= f\left(x^t\right) + \left\langle\nabla f\left(x^t\right), E_{Dt}\left(-\frac{s}{n_b}\sum_{i_t\in Dt}\nabla f_{it}(x^t)\right)\right\rangle + \frac{\beta}{2}E_{Dt}\|\frac{s}{n_b}\sum_{i_t\in Dt}\nabla f_{it}(x^t)\|^2 \\ &= f\left(x^t\right) - s\nabla f\left(x^t\right)^T E_{Dt}\left(\frac{1}{n_b}\sum_{i_t\in Dt}\nabla f_{it}(x^t)\right) + \frac{\beta s^2}{2}E_{Dt}\|\frac{1}{n_b}\sum_{i_t\in Dt}\nabla f_{it}(x^t)\|^2 \\ &= f(x^t) - s\nabla f(x)^T E(g^t) + \frac{\beta s^2}{2}E_Dt\|g^t\|^2 \end{aligned}$$

(iii):

$$E_{Dt}\left(f\left(x^{t+1}\right)\right) - f\left(x^t\right) = -s\nabla f\left(x^t\right)^T E_{Dt}\left(g^t\right) + \frac{\beta s^2}{2}E_{Dt}\left\|g^t\right\|^2$$

so:

$$
\begin{aligned}
E_{Dt}(f\left(x^{t+1}\right)-f\left(x^{t}\right)) &\leq -s\nabla f\left(x^{t}\right)^{T} E_{Dt}\left(\frac{1}{n_b}\sum_{i\in D_t}\nabla f_{it}\left(x^{t}\right)\right)+\frac{\beta s^{2}}{2}E_{Dt}\left\|g^{t}\right\|^{2} \\
&= -s\nabla f\left(x^{t}\right)^{T}\frac{1}{n_b}E_{Dt}\left(\sum_{i\in D_t}\nabla f_{it}\left(x^{t}\right)\right)+\frac{\beta s^{2}}{2}E_{Dt}\left\|g^{t}\right\|^{2} \\
&= -s\nabla f\left(x^{t}\right)^{T}\frac{1}{n_b}n_b\nabla f\left(x^{t}\right)+\frac{\beta s^{2}}{2}\left[\frac{\sigma^{2}}{n_b}+\left\|\nabla f\left(x^{t}\right)\right\|^{2}\right] \\
&= \left(\frac{\beta s^{2}}{2}-s\right)\|\nabla f(x^{t})s\|^{2}+\frac{\beta s^{2}\sigma^{2}}{2n_b}
\end{aligned}
$$

$\Rightarrow$

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

(iv):

In (iii), I know that

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

and

$$
f(t)-f(x^{*})\leq\frac{1}{2\alpha}\|\nabla f(x^{t})\|^{2}\Rightarrow\left\|\nabla f(x^{t})\right\|^{2}\geq2\alpha\left(f(t)-f(x^{*})\right)
$$

then

$$
E_{Dt}(f(x^{t+1})-f(x^{t}))\leq-(s-\frac{\beta s^{2}}{2})2\alpha(f(x^{t})-f^{*})+\frac{\beta s^{2}\sigma^{2}}{2n_b}
$$

so

$$
\begin{aligned}
E_{Dt}(f(x^{t+1})-f(x^{*}))+f^{*}-f(x^{t}) &\leq -2\alpha(s-\frac{\beta s^{2}}{2})(f(x^{t})-f^{*})+\frac{\beta s^{2}\sigma^{2}}{2n_b} \\
E_{Dt}(f(x^{t+1})-f(x^{*})) &\leq \left(1-2\alpha(s-\frac{\beta s^{2}}{2})\right)(f(x^{t})-f^{*})+\frac{\beta s^{2}\sigma^{2}}{2n_b} \\
E_{Dt}(f(x^{t+1})-f(x^{*}))-\frac{\beta s^{2}\sigma^{2}}{2n_b\alpha s(2-\beta s)} &\leq (1-\alpha s(2-\beta s))(f(x^{t})-f^{*}-\frac{\beta s^{2}\sigma^{2}}{2n_b\alpha s(2-\beta s)})
\end{aligned}
$$

which means

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

HW 2

Problem restatment

Derive BCD algorithm for LASSO problem

Answer

Define the problem:

$$min f(x) = \frac{1}{2}\left\| Ax - b \right\|^2 + \lambda \left\| x \right\|_1$$

Look at the kth in the t step x_k^t ,Let $A = [a_1, a_2, ..., a_n]$

$$f(x) = \frac{1}{2}\left\| \sum_{i \neq k} a_i x_i^t + a_k x_k^t - b \right\|^2 + \lambda \left(\sum_{i \neq k} |x_i^t| + |x_k^t| \right)$$

Let

$$g(x_k) = \frac{1}{2}\left\| \sum_{i \neq k} a_i x_i^t + a_k x_k - b \right\|^2 + \lambda |x_k|$$

Derive the $g(x_k^t)$,I can get

$$\partial g = a_k^T \left(a_k x_k + \sum_{i \neq k} a_i x_i^t - b \right) + \lambda \partial |x_k|$$

If $x_k > 0$,then

$$\partial g = a_k^T \left(a_k x_k + \sum_{i \neq k} a_i x_i^t - b \right) + \lambda = 0$$

$$x_k^{t+1} = \frac{a_k^T b - a_k^T \sum_{i \neq k} a_i x_i^t - \lambda}{a_k^T a_k}, \lambda < a_k^T b - a_k^T \sum_{i \neq k} a_i x_i^t$$

If $x_k < 0$,then

$$\partial g = a_k^T \left(a_k x_k + \sum_{i \neq k} a_i x_i^t - b \right) - \lambda = 0$$

$$x_k^{t+1} = \frac{a_k^T b - a_k^T \sum_{i \neq k} a_i x_i^t + \lambda}{a_k^T a_k}, \lambda < -a_k^T b + a_k^T \sum_{i \neq k} a_i x_i^t$$

If $x_k = 0$,then

$$0 \in \partial g = a_k^T \left(\sum_{i \neq k} a_i x_i^t - b \right) - \lambda \partial |0|$$

$$| - a_k^T b + a_k^T \sum_{i \neq k} a_i x_i^t | \leq \lambda$$

To sum up

$$x_k^{t+1} = \frac{sign\left(a_k^T b - a_k^T \sum_{i \neq k} a_i x_i^t \right) \left(|a_k^T b - a_k^T \sum_{i \neq k} a_i x_i^t| - \lambda \right)_+}{a_k^T a_k}$$

Algorithm:

Algorithm 1 BCD algorithm for LASSO

1.Input: Given a initial starting point x^0, z^0, u^0, ϵ
and t=0

2. for $t = 0, 1, 2, ..., T$,do:

3. for $k = 1, 2, ..., N$,do

4.

$$x_k^{t+1} = \frac{sign\left(a_k^T b - a_k^T \sum_{i \neq k} a_i x_i^t\right)\left(|a_k^T b - a_k^T \sum_{i \neq k} a_i x_i^t| - \lambda\right)_+}{a_k^T a_k}$$

5. end for

6. end for

7. output x^T

HW 3

Problem restatment

Derive ADMM algorithm for Fused LASSO problem

Answer

Define the problem:

$$minf(x) = \frac{1}{2}\|Ax - b\|^2 + \lambda \|Fx\|_1$$

The augmented Lagrangian is

$$\begin{aligned} L(x, z, v) &= \frac{1}{2}\|Ax - b\|^2 + \lambda \|z\| + \frac{\rho}{2}\|Fx - z\|^2 + v^T(Fx - z) \\ &= \frac{1}{2}\|Ax - b\|^2 + \lambda \|z\| + \frac{\rho}{2}\left\|Fx - z + \frac{v}{\rho}\right\|^2 - \frac{\rho}{2}\left\|\frac{v}{\rho}\right\|^2 \end{aligned}$$

Let $u^t = \frac{v^t}{\rho}$

Look at the x , $x^{t+1} = arg\min_x \left\{\frac{1}{2}\|Ax - b\|^2 + \frac{\rho}{2}\left\|Fx - z + \frac{v}{\rho}\right\|^2\right\}$

so, I get:

$$\begin{aligned} &\nabla_x \left\{\frac{1}{2}\|Ax - b\|^2 + \frac{\rho}{2}\left\|Fx - z + \frac{v}{\rho}\right\|^2\right\} \\ &= A^T(Ax - b) + \rho F^T\left(Fx - z^t + u^t\right) \\ &= 0 \end{aligned}$$

so,

$$x^{t+1} = \left(A^T A + \rho F^T F\right)^{-1}\left(A^T b + \rho F^T z^t - u^t\right)$$

Look at z:

$$\begin{aligned} z^{t+1} &= arg\min_z \left\{\lambda \|z\|_1 + \frac{\rho}{2}\left\|Fx^{t+1} - z + u^t\right\|^2\right\} \\ &= prox_{\frac{1}{\rho}\lambda\|\|_1}\{Fx^{t+1} + u^t\} \end{aligned}$$

Let

$$h(z_i) = \frac{\rho}{2}\left(f_i^T x^{t+1} - z_i + u_i^t\right)^2 + \lambda|z_i|$$

It is obviously that

$$z_i^{t+1} = \begin{cases} \frac{\rho f_i^T x^{t+1} + \rho u_i^t - \lambda}{\rho} & \rho f_i^T x^{t+1} + \rho u_i^t - \lambda > 0 \\ 0 & -\lambda \le \rho f_i^T x^{t+1} + \rho u_i^t \le \lambda \\ \frac{\rho f_i^T x^{t+1} + \rho u_i^t + \lambda}{\rho} & \rho f_i^T x^{t+1} + \rho u_i^t + \lambda < 0 \end{cases}$$

To sum up:

$$z^{t+1} = z_1 \times z2 \times ... \times z_n, z_i = \frac{sign\left(\rho x_{i+1}^{t+1} - \rho x_i^{t+1} + \rho u_i^t\right)\left(|\rho x_{i+1}^{t+1} - \rho x_i^{t+1} + \rho u_i^t| - \lambda\right)_+}{\rho}$$

Algorithm:

Algorithm 2 ADMM algorithm for Fused LASSO

1.Input: Given a initial starting point
x^0, z^0, u^0, ϵ,and t=0

2. while $t \leq T, \|\rho F^T\left(z^t - z^{t+1}\right)\|_2 >$
$\epsilon, \|Fx^{t+1} - z^{t+1}\|_2 > \epsilon$,do:

3.

$$x^{t+1} = \left(A^T A + \rho F^T F\right)^{-1}\left(A^T b + \rho F^T z^t - u^t\right)$$

$$z^{t+1} = z_1 \times z2 \times ... \times z_n, z_i = \frac{sign\left(\rho x_{i+1}^{t+1} - \rho x_i^{t+1} + \rho u_i^t\right)\left(|\rho x_{i+1}^{t+1} - \rho x_i^{t+1} + \rho u_i^t| - \lambda\right)_+}{\rho}$$

$$u^{t+1} = u^t + Fx^{t+1} - z^{t+1}$$

$$t := t + 1$$

4. end

5. output x^T, z^T, u^T

HW 4

Problem restatment

Consider the following Basis Pursuit problem as:

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

Derive ADMM algorithm for it.
Hit: using the indicator function of $\Omega = \{\mathbf{x} \mid A\mathbf{x} = \mathbf{b}\}$.

Answer

The problem is equivalent to ;

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

The augmented Lagrangian is:

$$\begin{aligned} L(x, z, v) &= \|x\|_1 + \delta_\Omega(z) + \frac{\rho}{2}\|x - z\|^2 - v^T(x - z) \\ &= \|x\|_1 + \delta_\Omega(z) + \frac{\rho}{2}\left\|x - z + \frac{v}{\rho}\right\|^2 - \frac{\rho}{2}\left\|\frac{v}{\rho}\right\|^2 \\ &= \|x\|_1 + \delta_\Omega(z) + \frac{\rho}{2}\|x - z + u\|^2 - \frac{\rho}{2}\|u\|^2 \end{aligned}$$

Consider x,

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

so, $x^t = x_1 \times x_2 \times ... \times x_n, x_i = sign(z_i^t - u_i^t)(|z_i^t - u_i^t| - \frac{1}{\rho})_+$
Considder z,

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

the problem can be written as :

$$min \left\|x^{t+1} - z + u^t\right\|^2$$
$$s.t. Az = b$$

This is a convex problem,
The Lagrangian:

$$L(z, a) = \frac{\rho}{2}\left\|x^{t+1} + u^t - z\right\|^2 + \alpha^T(Az - b)$$

The dual problem is:

$$maxg(\alpha) = \inf_z L(z, \alpha)$$

apply KKT condition:

$$\begin{cases} \rho(z - x^{t+1} - u^t) + A^T\alpha = 0 \\ Az = b \end{cases}$$

get:

$$z^* = -A^T(AA^T)^{-1}\left(A\left(x^{t+1} + u^t\right) - b\right) + x^{t+1} + u^t$$

so, the algorithm:

Algorithm 3 ADMM algorithm for Basic Pursuit problem

1.Input: Given a initial starting point x^0, z^0, u^0, ϵ,and t=0

2. while $t \le T, \|\rho\left(z^t - z^{t+1}\right)\|_2 > \epsilon, \|x^{t+1} - z^{t+1}\|_2 > \epsilon$,do:

3.

$$\begin{aligned} x^{t+1} &= x_1 \times x_2 \times ... \times x_n, x_i = sign\left(z_i^t - u_i^t\right)\left(|z_i^t - u_i^t + | - \frac{1}{\rho}\right)_+ \\ z^{t+1} &= -A^T(AA^T)^{-1}\left(A\left(x^{t+1} + u^t\right) - b\right) + x^{t+1} + u^t \\ u^{t+1} &= u^t + x^{t+1} - z^{t+1} \\ t &:= t + 1 \end{aligned}$$

4. end while

5. output x^T, z^T, u^T
