ELEN E6876 Sparse and Low-Dimensional Models for Hi
Homework #2

Chenye Yang cy2540@columbia.edu

March 25, 2020

P1

(1)

To show that for Lasso problem, ∇f is $\|\boldsymbol{A}\|^2$-Lipschitz, where $\|\boldsymbol{A}\|$ is the operator norm of $\boldsymbol{A}$, we need to show that:

$$\|\nabla f(\boldsymbol{x_1}) - \nabla f(\boldsymbol{x_2})\| \leq \|\boldsymbol{A}\|^2 \cdot \|\boldsymbol{x_1} - \boldsymbol{x_2}\|$$

where $f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{Ax} - \boldsymbol{y}\|_2^2$ and $\nabla f(\boldsymbol{x}) = \boldsymbol{A^*(Ax - y)}$.

Because

$$\begin{aligned}
\|\nabla f(\boldsymbol{x_1}) - \nabla f(\boldsymbol{x_2})\| &= \|\boldsymbol{A^*(Ax_1 - y)} - \boldsymbol{A^*(Ax_2 - y)}\| \\
&= \|\boldsymbol{A^*Ax_1} - \boldsymbol{A^*Ax_2}\| \\
&= \|\boldsymbol{A^*A}(\boldsymbol{x_1} - \boldsymbol{x_2})\| \\
&\leq \|\boldsymbol{A^*A}\| \cdot \|\boldsymbol{x_1} - \boldsymbol{x_2}\| \\
&\leq \|\boldsymbol{A^*}\| \cdot \|\boldsymbol{A}\| \cdot \|\boldsymbol{x_1} - \boldsymbol{x_2}\| \\
&= \|\boldsymbol{A}\|^2 \cdot \|\boldsymbol{x_1} - \boldsymbol{x_2}\|
\end{aligned}$$

Therefore, ∇f is $\|\boldsymbol{A}\|^2$-Lipschitz.

(2)

When $\boldsymbol{x}$ lies in $\mathbb{R}^n$

$$\mathrm{prox}_{\alpha\|\cdot\|_1}(\boldsymbol{w}) = \underset{\boldsymbol{x}\in\mathbb{C}^n}{\mathrm{argmin}} \left\{ \alpha\|\boldsymbol{x}\|_1 + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{w}\|_2^2 \right\}$$

$$\begin{aligned}
\left(\frac{\partial(\alpha\|\boldsymbol{x}\|_1 + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{w}\|_2^2)}{\partial \boldsymbol{x}} \right)_i &= \left(\frac{\partial \alpha\|\boldsymbol{x}\|_1}{\partial \boldsymbol{x}} + \boldsymbol{x} - \boldsymbol{w} \right)_i \\
&= \begin{cases} \alpha + x_i - w_i & , x_i > 0 \\ x_i - w_i & , x_i = 0 \\ -\alpha + x_i - w_i & , x_i < 0 \end{cases}
\end{aligned}$$

Let the above equation be 0, we can get the closed-form expression for the proximal mapping in real space.

$$0 = \begin{cases} \alpha + x_i - w_i & , x_i > 0 \\ x_i - w_i & , x_i = 0 \\ -\alpha + x_i - w_i & , x_i < 0 \end{cases}$$

$\Rightarrow$

$$x_i = \begin{cases} w_i - \alpha & , x_i > 0 \\ 0 & , x_i = 0 \\ w_i + \alpha & , x_i < 0 \end{cases}$$
$$= \begin{cases} w_i - \alpha & , w_i > \alpha \\ 0 & , -\alpha \leq w_i \leq \alpha \\ w_i + \alpha & , w_i < -\alpha \end{cases}$$

When $\boldsymbol{x}$ lies in $\mathbb{C}^n$

Let $\boldsymbol{x} = |\boldsymbol{x}|e^{i\phi(\boldsymbol{x})}$ and $\boldsymbol{w} = |\boldsymbol{w}|e^{i\phi(\boldsymbol{w})}$, then the proximal mapping can be written as

$$\begin{aligned} \operatorname{prox}_{\alpha\|\cdot\|_1}(\boldsymbol{w}) &= \arg\min_{\boldsymbol{x}\in\mathbb{C}^n} \left\{ \alpha\|\boldsymbol{x}\|_1 + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{w}\|_2^2 \right\} \\ &= \arg\min \left\{ \alpha \left\| |\boldsymbol{x}|e^{i\phi(\boldsymbol{x})} \right\|_1 + \frac{1}{2} \left\| |\boldsymbol{x}|e^{i\phi(\boldsymbol{x})} - |\boldsymbol{w}|e^{i\phi(\boldsymbol{x})} \right\|_2^2 \right\} \\ &= \arg\min \left\{ \alpha \||\boldsymbol{x}|\|_1 + \frac{1}{2} \left\| |\boldsymbol{x}| - |\boldsymbol{w}|e^{i(\phi(\boldsymbol{x}) - \phi(\boldsymbol{x}))} \right\|_2^2 \right\} \\ &= \arg\min \left\{ \alpha \||\boldsymbol{x}|\|_1 + \frac{1}{2} \left[|\boldsymbol{x}|^2 + |\boldsymbol{w}|^2 - |\boldsymbol{x}||\boldsymbol{w}| \left(e^{-i(\phi(\boldsymbol{w}) - \phi(\boldsymbol{x}))} + e^{i(\phi(\boldsymbol{w}) - \phi(\boldsymbol{x}))} \right) \right] \right\} \\ &= \arg\min \left\{ \alpha \||\boldsymbol{x}|\|_1 + \frac{1}{2} \left[|\boldsymbol{x}|^2 + |\boldsymbol{w}|^2 - 2|\boldsymbol{x}||\boldsymbol{w}| \cos(\phi(\boldsymbol{w}) - \phi(\boldsymbol{x})) \right] \right\} \end{aligned}$$

First consider the phase and fix the magnitude, to minimize the expression, we need to maximize $\cos(\phi(\boldsymbol{w}) - \phi(\boldsymbol{x}))$. Thus $\phi(\boldsymbol{w}) = \phi(\boldsymbol{x})$.

Then consider the magnitude and fix phase as $\phi(\boldsymbol{x}) = \phi(\boldsymbol{w})$, i.e. $\boldsymbol{x}$ and $\boldsymbol{w}$ lie on a line. Then the proximal mapping is written as

$$\begin{aligned} \operatorname{prox}_{\alpha\|\cdot\|_1}(\boldsymbol{w}) &= \arg\min \left\{ \alpha \||\boldsymbol{x}|\|_1 + \frac{1}{2} \left(|\boldsymbol{x}|^2 + |\boldsymbol{w}|^2 - 2|\boldsymbol{x}||\boldsymbol{w}| \right) \right\} \\ &= \arg\min \left\{ \alpha \||\boldsymbol{x}|\|_1 + \frac{1}{2} \left(|\boldsymbol{x}| - |\boldsymbol{w}| \right)^2 \right\} \end{aligned}$$

Similar to the case when $\boldsymbol{x}$ is real,

$$\left(\frac{\partial \left[\alpha \||\boldsymbol{x}|\|_1 + \frac{1}{2} \left(|\boldsymbol{x}| - |\boldsymbol{w}| \right)^2 \right]}{\partial \boldsymbol{x}} \right)_i = \begin{cases} \alpha + |x_i| - |w_i| & , |x_i| > 0 \\ |x_i| - |w_i| & , |x_i| = 0 \\ -\alpha - |x_i| + |w_i| & , |x_i| < 0 \end{cases}$$

Thus

$$|x_i| = \begin{cases} |w_i| - \alpha & , |w_i| > \alpha \\ 0 & , |w_i| \leq \alpha \end{cases}$$

Therefore, we can get the closed-form expression for the proximal mapping in complex space:

$$
\begin{aligned}
x_i = |x_i| e^{i\phi(\boldsymbol{x})} &= \begin{cases} \left(|w_i| - \alpha\right) e^{i\phi(\boldsymbol{w})} & , |w_i| > \alpha \\ 0 & , |w_i| \le \alpha \end{cases} \\
&= \begin{cases} w_i \times \dfrac{|w_i| - \alpha}{|w_i|} & , |w_i| > \alpha \\ 0 & , |w_i| \le \alpha \end{cases}
\end{aligned}
$$

2.3.3

1.

$$
\mathcal{I}_{[\epsilon, 1-\epsilon]}(x) = \begin{cases} 0 & , \epsilon \le x \le 1-\epsilon \\ +\infty & , \text{else} \end{cases}
$$

Thus

$$
\partial \mathcal{I}_{[\epsilon, 1-\epsilon]}(x) = \begin{cases} [-\infty, 0] & , x = \epsilon \\ [0, +\infty] & , x = 1-\epsilon \\ 0 & , \text{else} \end{cases}
$$

Similar with former

$$
\operatorname{prox}_{\alpha_\gamma \mathcal{I}_{[\epsilon, 1-\epsilon]}}(x) = \begin{cases} \epsilon & , x < \epsilon \\ x & , \epsilon \le x \le 1-\epsilon \\ 1-\epsilon & , x > 1-\epsilon \end{cases}
$$

2.

$$
H(\boldsymbol{x}) = \frac{1}{2} \left\| \boldsymbol{a}_\gamma \otimes \boldsymbol{x} - \boldsymbol{y} \right\|_2^2
$$

$$
\boldsymbol{a} \otimes \boldsymbol{x} = \boldsymbol{C}_a \boldsymbol{x}
$$

Thus

$$
\nabla_{\boldsymbol{x}} H(\boldsymbol{x}, \gamma) = \boldsymbol{C}_a^T \times \left(a_\gamma \otimes \boldsymbol{x} - \boldsymbol{y}\right)
$$

$$
\begin{aligned}
\nabla_\gamma H(\boldsymbol{x}, \gamma) &= \left(\frac{\partial \left(\boldsymbol{a}_\gamma \otimes \boldsymbol{x}\right)}{\partial \gamma} \right)^T \times \frac{\partial H}{\partial \left(\boldsymbol{a}_\gamma \otimes \boldsymbol{x}\right)} \\
&= \left(\frac{\partial \boldsymbol{a}_\gamma}{\partial \gamma} \otimes \boldsymbol{x} \right)^T \times \left(\boldsymbol{a}_\gamma \otimes \boldsymbol{x} - \boldsymbol{y}\right)
\end{aligned}
$$