Supplementary Material for:
Semismooth Newton Algorithm for Efficient Projections onto $\ell_{1,\infty}$-norm Ball

Dejun Chu¹  Changshui Zhang¹  Shiliang Sun²  Qing Tao³

1. Proof of Proposition 1

**Proposition 1** Let $s(\theta)$ be defined by (12). Then $s(\theta)$ is convex, strictly monotonically decreasing with $dm + 1$ breakpoints at most, and the equation (11) has unique root on the interval $[0, \max_i \sum_{j=1}^m A_{i,j}]$.

**Proof:** All the breakpoints are given by (9), in which $\mu_i$ is 0 or equal to each element of the $i$-th row of data matrix $A$. Thus, $s(\theta)$ has $dm + 1$ breakpoints at most.

Meanwhile, it is clear that for all $i$, $\bar{\mu}_i(\theta)$ is convex, continuous and monotonically decreasing in $[0, \max_i \sum_{j=1}^m A_{i,j}]$ with respect to $\theta$, and strictly monotonically decreasing in $[0, \sum_{j=1}^m A_{i,j}]$. Therefore, $s(\theta)$ is convex and strictly monotonically decreasing in $[0, \max_i \sum_{j=1}^m A_{i,j}]$.

It is easily verified that $\bar{\mu}_i = \max_j A_{i,j}$ given $\theta = 0$. Thus, we have $s(0) > 0$ from the assumption of the problem (4), and $s(\theta_{\max}) = -C < 0$ where $\theta_{\max} = \max_i \sum_{j=1}^m A_{i,j}$. According to the Intermediate Value Theorem, $s(\theta)$ has unique root on the interval $[0, \max_i \sum_{j=1}^m A_{i,j}]$.

2. Proof of Proposition 2

**Proposition 2** Assume $|\mathcal{I}(\mu_i(t))| \geq 1$ for $i = 1, 2, \ldots, d$. Then $v_{d+1}^{(t)}$ is the Newton step for $s(\theta)$ at $\theta^{(t)}$.

**Proof:** Substituting (19) into (20), we can rewrite the last element of $v$ as

$$v_{d+1}^{(t)} = \frac{-F_{d+1} + \sum_{i=1}^d \frac{F_i}{|\mathcal{I}(\mu_i^{(t)})|}}{\sum_{i=1}^d 1/|\mathcal{I}(\mu_i^{(t)})|} = \sum_{i=1}^d \mu_i^{(t)} - C + \sum_{i=1}^d \frac{\max_j(A_{i,j} - \mu_i^{(t)}), 0) - \theta^{(t)}}{|\mathcal{I}(\mu_i^{(t)})|} \sum_{i=1}^d 1/|\mathcal{I}(\mu_i^{(t)})| = \sum_{i=1}^d \frac{\sum_{j \in \mathcal{I}(\mu_i^{(t)})} A_{i,j} - \theta^{(t)}}{|\mathcal{I}(\mu_i^{(t)})|} - C$$

Thus, $v_{d+1}^{(t)}$ is the Newton step at $\theta^{(t)}$ for the search direction of (11).
3. Proof of Proposition 3

**Proposition 3** Suppose \( \theta^{(t)} \) lies between two breakpoints, i.e., \( \theta^{(t)} \in (\Theta_{[j-1]}, \Theta_{[j]}) \). Assume \( s(\Theta_{[j]}) > 0 \). There holds
\[
\theta^{(t)} \leq \Theta_{[j]} < \theta^{(t+1)}.
\]

**Proof:** We focus on the right inequality while the left one is obvious. From the update of \( \theta \), we obtain
\[
\theta^{(t+1)} = \theta^{(t)} + \frac{\sum_{i=1}^{d} \tilde{\mu}_i(\theta^{(t)}) - C}{\sum_{i=1}^{d} 1/|I(\mu_i^{(t)})|}.
\]

Recalling the definition of \( \tilde{\mu}_i(\theta^{(t)}) \), we have
\[
\theta^{(t+1)} = \frac{\sum_{i=1}^{d} \sum_{j \in I(\mu_i^{(t)})} A_{i,j}}{|I(\mu_i^{(t)})|} - C \frac{\sum_{i=1}^{d} 1}{|I(\mu_i^{(t)})|}.
\]

If \( s(\Theta_{[j]}) > 0 \), we have \( s(\theta^{(t)}) > 0 \) since \( s(\theta) \) is a strictly monotonically decreasing function. Meanwhile,
\[
s(\Theta_{[j]}) = \sum_{i=1}^{d} \frac{\sum_{j \in I(\mu_i^{(t)})} A_{i,j}}{|I(\mu_i^{(t)})|} - \Theta_{[j]} \frac{\sum_{i=1}^{d} 1}{|I(\mu_i^{(t)})|} - C > 0.
\]

This means that
\[
\Theta_{[j]} < \frac{\sum_{i=1}^{d} \sum_{j \in I(\mu_i^{(t)})} A_{i,j}}{|I(\mu_i^{(t)})|} - C \frac{1}{\sum_{i=1}^{d} 1/|I(\mu_i^{(t)})|} = \theta^{(t+1)}.
\]

4. Proof of Proposition 4

**Proposition 4** Assume \( \mu_i^{(t)} \) is updated via (23) for \( t \geq 0 \). Then we have
\[
\sum_{i=1}^{d} \mu_i^{(t+1)} - C = 0.
\]

**Proof:** From the update of \( \theta^{(t)} \), we have
\[
\theta^{(t+1)} = \left( \sum_i \sum_{j \in I(\mu_i^{(t)})} A_{i,j} \right) / \sum_i 1/|I(\mu_i^{(t)})|,
\]
which means
\[
\sum_i \frac{\sum_{j \in I(\mu_i^{(t)})} A_{i,j}}{|I(\mu_i^{(t)})|} - \theta^{(t+1)} \sum_i 1/|I(\mu_i^{(t)})| - C = 0
\]
\[
\iff \sum_i \frac{\sum_{j \in I(\mu_i^{(t)})} A_{i,j} - \theta^{(t+1)}}{|I(\mu_i^{(t)})|} - C = 0
\]
\[
\iff \sum_{i=1}^{d} \mu_i^{(t+1)} - C = 0.
\]
5. Proof of Lemma 2

Lemma 2 Assume that \( \mu_i^{(t)} \in [0, \max_j A_{i,j}] \), \( \theta^{(t)} \geq 0 \) and the following two inequalities hold:

(i) \( \sum_{j=1}^{m} \max \left( A_{i,j} - \mu_i^{(t)}, 0 \right) \geq \theta^{(t)} \), (ii) \( s(\theta^{(t)}) \geq 0 \), then it can be obtained that

\[
\sum_{j=1}^{m} \max \left( A_{i,j} - \mu_i^{(t+1)}, 0 \right) \geq \theta^{(t+1)}.
\]

Proof: According to the inequality (i), we have

\[
\sum_{j \in I(\mu_{i}^{(t)})} A_{i,j} - |I(\mu_{i}^{(t)})| \mu_i^{(t)} - \theta^{(t)} \geq 0
\]

\( \Leftrightarrow \sum_{j \in I(\mu_{i}^{(t)})} A_{i,j} \geq |I(\mu_{i}^{(t)})| \mu_i^{(t)} + \theta^{(t)}. \)

Meanwhile, from the definition of \( I(\mu_{i}^{(t)}) \) and using

\[
\mu_i^{(t+1)} = \frac{\sum_{j \in I(\mu_{i}^{(t)})} A_{i,j} - \theta^{(t+1)}}{|I(\mu_{i}^{(t)})|},
\]

it can be obtained

\[
\sum_{j=1}^{m} \max \{ A_{i,j} - \mu_i^{(t+1)}, 0 \} - \theta^{(t+1)}
\]

\[
= \sum_{j=1}^{m} \max \{ A_{i,j} - \sum_{k \in I(\mu_{i}^{(t)})} A_{i,k} - \theta^{(t+1)} \frac{|I(\mu_{i}^{(t)})|}{|I(\mu_{i}^{(t)})|}, 0 \} - \theta^{(t+1)}
\]

\[
\geq \sum_{j \in I(\mu_{i}^{(t)})} A_{i,j} - |I(\mu_{i}^{(t)})| \sum_{j \in I(\mu_{i}^{(t)})} A_{i,j} - \theta^{(t+1)} - \theta^{(t+1)}
\]

\[
= 0.
\]

6. Proof of Corollary 1

Corollary 1 Assume that \( \sum_{j=1}^{m} \max \left( A_{i,j} - \mu_i^{(t)}, 0 \right) \geq \theta^{(t)} \). Then we can obtain

\( \tilde{\mu}_i(\theta^{(t)}) \geq \mu_i^{(t)}. \)

Proof: From the definition of \( \tilde{\mu}_i(\theta^{(t)}) \), we have

\[
\tilde{\mu}_i(\theta^{(t)}) = \frac{\sum_{j \in I(\mu_{i}^{(t)})} A_{i,j} - \theta^{(t)}}{|I(\mu_{i}^{(t)})|}
\]

\[
= \mu_i^{(t)} + \frac{\sum_{j \in I(\mu_{i}^{(t)})} A_{i,j} - |I(\mu_{i}^{(t)})| \mu_i^{(t)} - \theta^{(t)}}{|I(\mu_{i}^{(t)})|}
\]

\[
\geq \mu_i^{(t)}.
\]

The last inequality comes from the assumption which concludes the proof.