Supplementary for Locally Regularized Sparse Graph by Fast Proximal Gradient Descent

Dongfang Sun, Yingzhen Yang

1School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ 85281, USA , {dsun30, yingzhen.yang}@asu.edu

1 PROOFS AND MORE TECHNICAL RESULTS

Proposition 1.1. Define \( C^+ = \{ t : 1 \leq t \leq n, c_{ti} > 0 \} \), and \( C^- = \{ t : 1 \leq t \leq n, c_{ti} < 0 \} \). Let \( z^* \) be a critical point of function \( \tilde{F} \) in eq.(7) of the main paper. Then for arbitrary small positive number \( \varepsilon > 0 \), \( \tilde{z}^{*, \varepsilon} \in \mathbb{R}^n \) defined by

\[
\tilde{z}_k^{*, \varepsilon} = \begin{cases} 
  z_k^* & \text{if } z_k^* \neq 0 \text{ or } k \in C^+ \\
  \varepsilon & \text{otherwise}
\end{cases}
\]  

Then there exists \( u \in \partial F(\tilde{z}^{*, \varepsilon}) \) for \( F \) in eq.(6) of the main paper such that \( \| u \|_2 \leq L_f |C^-| \varepsilon \) where \( L_f := 2 \sigma_{\max}(X^T X) \).

Proof. Since the only different elements between \( \tilde{z}^{*, \varepsilon} \) and \( z^* \) are those with indices in \( A := C^- \setminus \{ k : z_k^* = 0 \} \), we have

\[
\| \nabla f(\tilde{z}^{*, \varepsilon}) - \nabla f(z^*) \|_2 \leq L_f \| \tilde{z}^{*, \varepsilon} - z^* \|_2 \leq L_f |C^-| \varepsilon,
\]

where \( L_f = 2 \sigma_{\max}(X^T X) \). Because \( z^* \) be a critical point of function \( \tilde{F} \), there exists \( q \in \partial h_{\gamma, c} \) such that \( \tilde{p} := \nabla f(z^*) + q = 0 \). Define \( \tilde{h}_{\gamma, c} = \gamma \sum_{k=1}^n c_k \mathbb{I}_{z_k \neq 0} \). With the definition of \( \tilde{z}^{*, \varepsilon} \), we have \( \tilde{q} \in \partial h_{\gamma, c}(\tilde{z}^{*, \varepsilon}) \) such that \( \tilde{q}_k = 0 \) for \( k \in A \) and \( \tilde{q}_k = q_k \) otherwise. Moreover, \( \tilde{q}_k = 0 \) for all \( k \in A \).

Therefore, let \( \tilde{p} \triangleq \nabla f(\tilde{z}^{*, \varepsilon}) + \tilde{q} \in \partial F(\tilde{z}^{*, \varepsilon}) \), we have

\[
\| \tilde{p} \|_2 = \| \tilde{p} - p \|_2 = \| \nabla f(\tilde{z}^{*, \varepsilon}) - \nabla f(z^*) \|_2 \leq L_f |C^-| \varepsilon.
\]

The claim of this proposition follows with \( u = \tilde{p} \).

We repeat critical equations in the main paper and define more notations before stating the proof of Theorem 3.2.

\[
\text{prox}_{s \gamma, c}(u) := \arg \min_{v \in \mathbb{R}^n} \frac{1}{2s} \| v - u \|_2^2 + h_{\gamma, c}(z) = T_{s, \gamma, c}(u),
\]

where \( s > 0 \) is the step size, \( T_{s, \gamma, c} \) is an element-wise hard thresholding operator. For \( 1 \leq t \leq n \),

\[
[T_{s, \gamma, c}(u)]_t = \begin{cases} 
  0 & : |u_t| \leq \sqrt{2s \gamma c_{ti}} \text{ and } c_{ti} > 0, \text{ or } t = i \\
  u_t & : \text{otherwise}
\end{cases}
\]  

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).
1.1 PROOF OF THEOREM 3.2

Proof of Theorem 3.2. First of all, it can be verified that supp($z_c^{(k)}$) \(\subseteq\) supp($z_c^{(k-1)}$) for all $k \geq 1$ when $s < \frac{\alpha_k}{\lambda_k}$. Therefore, there exists a finite $k' \geq 1$ such that $\{z_c^{(k)}\}_{k \geq k_0}$ have the same support $S$. We note that $\lambda$ can be also be slightly adjusted so that supp($v_c^{(k)}$) = $S$ for all $k \geq k_0$. Now we consider any $k > k'$ in the sequel, and let $z \in \mathbb{R}^n$ be a vector such that supp($z_c$) = $S$.

Because $f$ have $L_f$-Lipschitz continuous gradient, we have

$$f(z^{(k)}) \leq f(m^{(k)}) + \langle \nabla f(m^{(k)}), z^{(k)} - m^{(k)} \rangle + \frac{L_f}{2} \left\| z^{(k)} - m^{(k)} \right\|^2_2. \quad (3)$$

Also,

$$f(m^{(k)}) - (1 - \alpha_k) f(z^{(k-1)}) - \alpha_k f(z) = (1 - \alpha_k) \left( f(m^{(k)}) - f(z^{(k-1)}) \right) + \alpha_k \left( f(m^{(k)}) - f(z) \right) \quad (1)$$

$$\leq (1 - \alpha_k) \langle \nabla f(m^{(k)}), m^{(k)} - z^{(k-1)} \rangle + \alpha_k \langle \nabla f(m^{(k)}), m^{(k)} - z \rangle \leq \langle \nabla f(m^{(k)}), (1 - \alpha_k)(m^{(k)} - z^{(k-1)}) + \alpha_k (m^{(k)} - z) \rangle = \langle \nabla f(m^{(k)}), m^{(k)} - (1 - \alpha_k) z^{(k-1)} - \alpha_k z \rangle, \quad (4)$$

where $1$ is due to the convexity of $f$.

We have $\tilde{v}^{(k)} = v^{(k-1)} - \lambda_k$ $\nabla f(m^{(k)})$, and it follows that

$$\frac{1}{2\lambda_k} \left( \left\| v^{(k-1)} - z \right\|^2_2 - \left\| v^{(k)} - z \right\|^2_2 - \left\| v^{(k)} - v^{(k-1)} \right\|^2_2 \right)$$

$$= \frac{1}{\lambda_k} \langle z - v^{(k)}, v^{(k)} - v^{(k-1)} \rangle \quad (1)$$

$$\frac{1}{\lambda_k} \langle z - v^{(k)}, \tilde{v}^{(k)} - v^{(k-1)} \rangle = \langle \nabla f(m^{(k)}), v^{(k)} - z \rangle, \quad (5)$$

and $1$ is due to the fact that supp($z_c - v_c^{(k)}$) $\subseteq$ $S$ because supp($z_c$) = $S$, supp($v_c^{(k)}$) $\subseteq$ $S$.

Because supp($v_c^{(k)}$) $\subseteq$ supp($z_c$), we have

$$h_{\gamma,c}(v^{(k)}) \leq h_{\gamma,c}(z). \quad (6)$$

It follows by $(5)$ and $(6)$ that

$$\langle \nabla f(m^{(k)}), v^{(k)} - z \rangle + h_{\gamma,c}(v^{(k)}) \leq h_{\gamma,c}(z) + \frac{1}{2\lambda_k} \left( \left\| v^{(k-1)} - z \right\|^2_2 - \left\| v^{(k)} - z \right\|^2_2 - \left\| v^{(k)} - v^{(k-1)} \right\|^2_2 \right) \quad (7)$$

Similar to $(5)$, we have

$$\frac{1}{2s} \left( \left\| m^{(k)} - z \right\|^2_2 - \left\| z^{(k)} - z \right\|^2_2 - \left\| z^{(k)} - m^{(k)} \right\|^2_2 \right) = \frac{1}{s} \langle z - z^{(k)}, z^{(k)} - m^{(k)} \rangle. \quad (8)$$

For any $q \in \partial h_{\gamma,c}(z^{(k)})$, due to the fact that supp($z_c$) = supp($z_c^{(k)}$),

$$\langle z - z^{(k)}, q \rangle + h_{\gamma,c}(z^{(k)}) = h_{\gamma,c}(z). \quad (9)$$
By (8) and (9),
\[
\langle z - z^{(k)}, \frac{1}{s}(z^{(k)} - m^{(k)}) + q \rangle + h_{\gamma,c}(z^{(k)}) = h_{\gamma,c}(z) + \frac{1}{2s} \left( \|m^{(k)} - z\|_2^2 - \|z^{(k)} - z\|_2^2 - \|z^{(k)} - m^{(k)}\|_2^2 \right) \tag{10}
\]

By the optimality condition of the proximal mapping in eq.(10) in Algorithm 1, we can choose \( q \in \partial h_{\gamma,c}(z^{(k)}) \) such that \( z^{(k)} = m^{(k)} - s(\nabla f(m^{(k)}) + q) \). Plugging such \( q \) in \( (10) \), we have
\[
\langle \nabla f(m^{(k)}), z^{(k)} - z \rangle + h_{\gamma,c}(z^{(k)}) = h_{\gamma,c}(z) + \frac{1}{2s} \left( \|m^{(k)} - z\|_2^2 - \|z^{(k)} - z\|_2^2 - \|z^{(k)} - m^{(k)}\|_2^2 \right) \tag{11}
\]

Setting \( z = (1 - \alpha_k)z^{(k-1)} + \alpha_k v^{(k)} \) in \( (11) \), we have
\[
\langle \nabla f(m^{(k)}), z^{(k)} - (1 - \alpha_k)z^{(k-1)} - \alpha_k v^{(k)} \rangle + h_{\gamma,c}(z^{(k)}) \leq h_{\gamma,c}((1 - \alpha_k)z^{(k-1)} + \alpha_k v^{(k)}) + \frac{1}{2s} \left( \|m^{(k)} - (1 - \alpha_k)z^{(k-1)} - \alpha_k v^{(k)}\|_2^2 - \|z^{(k)} - m^{(k)}\|_2^2 \right) \tag{12}
\]

where \( (1) \) is due to the fact that \( \text{supp}(v^{(c)}_c) = \text{supp}(z^{(k-1)}_c) \) and \( h_{\gamma,c} \) satisfies \( h_{\gamma,c}((1 - \tau)u + \tau v) \leq (1 - \tau)h_{\gamma,c}(u) + \tau h_{\gamma,c}(v) \) for any two vectors \( u, v \) with \( \text{supp}(u_c) = \text{supp}(v_c) \) and any \( \tau \in (0, 1) \). \( (2) \) is due to the fact that \( m^{(k)} - (1 - \alpha_k)z^{(k-1)} - \alpha_k v^{(k)} = \alpha_k(v^{(k-1)} - v^{(k)}) \) according to eq.(9) in Algorithm 1.

Computing \( \alpha_k \times (7) + (12) \), we have
\[
\langle \nabla f(m^{(k)}), z^{(k)} - (1 - \alpha_k)z^{(k-1)} - \alpha_k z \rangle + h_{\gamma,c}(z^{(k)}) \leq (1 - \alpha_k)h_{\gamma,c}(z^{(k-1)} + \alpha_k h_{\gamma,c}(z) + \frac{\alpha_k}{2\lambda_k} \left( \|v^{(k-1)} - z\|_2^2 - \|v^{(k-1)} - v^{(k)}\|_2^2 \right) \tag{13}
\]

and \( (1) \) is due to \( \lambda_k \alpha_k \leq s \).

Combining (3), (4) and (13), and noting that \( \tilde{F}(z) = f(z) + h_{\gamma,c}(z) \), we have
\[
\tilde{F}(z^{(k)}) \leq (1 - \alpha_k)\tilde{F}(z^{(k-1)}) + \alpha_k \tilde{F}(z) - \left( \frac{1}{2s} - \frac{L_f}{2} \right) \|z^{(k)} - m^{(k)}\|_2^2 + \frac{\alpha_k}{2\lambda_k} \left( \|v^{(k-1)} - z\|_2^2 - \|v^{(k-1)} - v^{(k)}\|_2^2 \right) \tag{14}
\]

It follows by (14) that
\[
\tilde{F}(z^{(k)}) - \tilde{F}(z) \leq (1 - \alpha_k) \left( \tilde{F}(z^{(k-1)}) - \tilde{F}(z) \right) - \left( \frac{1}{2s} - \frac{L_f}{2} \right) \|z^{(k)} - m^{(k)}\|_2^2 + \frac{\alpha_k}{2\lambda_k} \left( \|v^{(k-1)} - z\|_2^2 - \|v^{(k-1)} - v^{(k)}\|_2^2 \right) \tag{15}
\]
Define a sequence \( \{T_k\}_{k=1}^{\infty} \) as \( T_1 = 1 \), and \( T_k = (1 - \alpha_k)T_{k-1} \) for \( k \geq 2 \). Dividing both sides of (15) by \( T_k \), we have

\[
\frac{\hat{F}(z^{(k)}) - \hat{F}(z)}{T_k} \leq \frac{\hat{F}(z^{(k-1)}) - \hat{F}(z)}{T_{k-1}} - \frac{1 - Lfs}{2sT_k} \|z^{(k)} - m^{(k)}\|_2^2 \\
+ \frac{\alpha_k}{2\lambda_k T_k} \left( \|v^{(k-1)} - z\|_2^2 - \|v^{(k)} - z\|_2^2 \right).
\]

Since we choose \( \alpha_k = \frac{2}{k+1} \), it follows that \( T_k = \frac{2}{k+1} \) for all \( k \geq 1 \). Plugging the values of \( \alpha_k \) and \( T_k \) in \( \frac{\alpha_k}{2\lambda_k T_k} \) in (16), we have

\[
\frac{\hat{F}(z^{(k)}) - \hat{F}(z)}{T_k} \leq \frac{\hat{F}(z^{(k-1)}) - \hat{F}(z)}{T_{k-1}} - \frac{1 - Lfs}{2sT_k} \|z^{(k)} - m^{(k)}\|_2^2 \\
+ \frac{k}{2\lambda_k} \left( \|v^{(k-1)} - z\|_2^2 - \|v^{(k)} - z\|_2^2 \right) \\
\stackrel{\circ}{\leq} \frac{\hat{F}(z^{(k-1)}) - \hat{F}(z)}{T_{k-1}} - \frac{1 - Lfs}{2sT_k} \|z^{(k)} - m^{(k)}\|_2^2 + \frac{k}{2\lambda_k} \|v^{(k-1)} - z\|_2^2 \\
- \frac{k+1}{2\lambda_k} \|v^{(k)} - z\|_2^2,
\]

where \( \circ \) is due to the condition that \( \lambda_{k+1} \geq \frac{k+1}{k} \lambda_k \) for \( k \geq 1 \).

Set \( k_0 = k' + 1 \). Summing the above inequality for \( k = k_0, \ldots, m \) with \( m \geq k_0 \), we have

\[
\frac{\hat{F}(z^{(m)}) - \hat{F}(z)}{T_m} \leq \frac{\hat{F}(z^{(k_0-1)}) - \hat{F}(z)}{T_{k_0-1}} + \frac{k_0}{2\lambda_{k_0}} \|v^{(k_0-1)} - z\|_2^2 - \sum_{k=k_0}^{m-1} \frac{1 - Lfs}{2sT_k} \|z^{(k)} - m^{(k)}\|_2^2 \\
\leq \frac{k_0(k_0 - 1)}{2} \left( \frac{\hat{F}(z^{(k_0-1)}) - \hat{F}(z)}{T_{k_0-1}} \right) + \frac{\|v^{(k_0-1)} - z\|_2^2}{2\eta}.
\]

Since \( T_m = \frac{2}{m(m+1)} \), it follows by (18) with \( z = z^* \) that

\[
\hat{F}(z^{(m)}) - \hat{F}(z^*) \leq \frac{1}{m(m+1)} \left( k_0(k_0 - 1) \left( \frac{\hat{F}(z^{(k_0-1)}) - \hat{F}(z^*)}{T_{k_0-1}} \right) + \frac{\|v^{(k_0-1)} - z^*\|_2^2}{\eta} \right).
\]

Changing \( m \) to \( k \) in (19) completes the proof.

\[\square\]

2 ADDITIONAL ILLUSTRATION

Figure 1 illustrates the comparison between the weighed adjacency matrix of \( \ell^1 \)-graph and SRSG.

References


Figure 1: The comparison between the weighed adjacency matrix $W$ of the sparse graph produced by $\ell^1$-graph (right) and SRSG (left) on the Extended Yale Face Database B, where each white dot indicates an edge in the sparse graph.