An ADMM Algorithm for a Generic ℓ_0 Sparse Overlapping Group Lasso Problem

Youming Zhao
Email: youming0.zhao@gmail.com

First draft: March 31, 2022 Last update: November 12, 2023

Abstract

We present an alternating direction method of multipliers (ADMM) for a generic overlapping group lasso problem, where the groups can be overlapping in an arbitrary way. Meanwhile, we prove the lower bounds and upper bounds for both the ℓ_1 sparse group lasso problem and the ℓ_0 sparse group lasso problem. Also, we propose the algorithms for computing these bounds.

1 Generic ℓ_0 sparse overlapping group lasso

The generic ℓ_0 sparse overlapping group lasso problem is defined as

$$\min_{x\in\mathbf{R}^n}\left\{F(x) := \frac{1}{2s}\|x-v\|^2 + \lambda_0\|x\|_0 + \lambda_1\sum_{i=1}^{m}\|x_{G_i}\|_2\right\} \tag{1}$$

where m denotes the number of groups, and $G_i \subseteq \{1,2,\ldots,n\}$ contains the feature indices of the i-th group. Here s is the step size employed to get v based on x. Note that $\bigcap_{i=\{1,\ldots,m\}} G_i \neq \emptyset$. Now consider the problem,

$$\begin{array}{ll}\text{minimize}_{x_{G_i}\in\mathbf{R}^{n_i}, z\in\mathbf{R}^n} & \frac{1}{2s}\|z-v\|^2 + \lambda_0\|z\|_0 + \lambda_1\sum_{i=1}^m\|x_{G_i}\|_2 \\ \text{subject to} & x_{G_i} - z_i = 0, \quad i=1,2,\ldots,m\end{array} \tag{2}$$

where $z_i \in \mathbf{R}^{n_i}$ is defined by $(z_i)_j = z_{G(i,j)} = z_g$. Here, $G(i,j) = g$ denotes the global index (location) of the j-th element in the set (group) G_i. Hence, z is a linear function of $z_i, i \in \{1,\ldots,m\}$.

The augmented Lagrangian for (2) is

$$L_\rho(x_{G_i}, z, y_{G_i}) = \sum_{i=1}^{m}\left[\lambda_1\|x_{G_i}\|_2 + y_{G_i}^T(x_{G_i}-z_i) + \frac{\rho}{2}\|x_{G_i}-z_i\|_2^2\right] + \frac{1}{2s}\|z-v\|^2 + \lambda_0\|z\|_0. \tag{3}$$

$$x_{G_i}^{k+1} := \operatorname*{argmin}_{x_{G_i}}\left(\lambda_1\|x_{G_i}\|_2 + x_{G_i}^T y_{G_i}^k + \frac{\rho}{2}\|x_{G_i} - z_i^{k+1}\|_2^2\right), \quad i=1,2,\ldots,m \tag{4}$$

$$z^{k+1} := \operatorname*{argmin}_{z}\left(\frac{1}{2s}\|z-v\|_2^2 + \lambda_0\|z\|_0 + \sum_{i=1}^m\left(\frac{\rho}{2}\|x_{G_i}^{k+1} - z_i\|_2^2 - z_i^T y_{G_i}^k\right)\right) \tag{5}$$

$$y_{G_i}^{k+1} := y_{G_i}^k + \rho(x_{G_i}^{k+1} - z_i^{k+1}) \tag{6}$$

where x_{G_i} and z are primal variables, and y_{G_i} are dual variables. The x-update is actually a group lasso problem and can be solved with the proximal operator of group lasso as follows.

$$x_{G_i}^{k+1} = \operatorname*{argmin}_{x_{G_i}}\left(\lambda_1\|x_{G_i}\|_2 + x_{G_i}^T y_{G_i}^k + \frac{\rho}{2}\|x_{G_i} - z_i^{k+1}\|_2^2\right)$$

$$
\begin{aligned}
&= \underset{x_{G_i}}{\operatorname{argmin}} \left(\frac{1}{2}\|x_{G_i} - z_i^{k+1}\|_2^2 + x_{G_i}^T \frac{y_{G_i}^k}{\rho} + \frac{\lambda_1}{\rho}\|x_{G_i}\|_2 \right) \\
&= \underset{x_{G_i}}{\operatorname{argmin}} \left(\frac{1}{2}\|x_{G_i} - (z_i^{k+1} - \frac{y_{G_i}^k}{\rho})\|_2^2 + \frac{\lambda_1}{\rho}\|x_{G_i}\|_2 \right) \\
&= S_{\lambda_1/\rho}(z_i^{k+1} - \frac{y_{G_i}^k}{\rho}), \quad i = 1, 2, \ldots, m
\end{aligned}
$$

where $S_\lambda(\cdot)$ is a soft-thresholding operator defined as

$$
S_\lambda(a) = (\|a\|_2 - \lambda)_+ \frac{a}{\|a\|_2}.
$$

Now we derive the solution to the z-update.

$$
\begin{aligned}
z^{k+1} &= \underset{z}{\operatorname{argmin}} \left(\frac{1}{2s}\|z - v\|_2^2 + \lambda_0\|z\|_0 + \sum_{i=1}^{m} \left(\frac{\rho}{2}\|x_{G_i}^{k+1} - z_i\|_2^2 - z_i^T y_{G_i}^k \right) \right) \\
&= \underset{z}{\operatorname{argmin}} \sum_{g=1}^{n} \left(\frac{1}{2s}(z_g - v_g)^2 + \lambda_0\|z_g\|_0 + \sum_{G(i,j)=g} \left(\frac{\rho}{2}(z_i)_j^2 - (y_i^k)_j (z_i)_j - \rho(x_i^{k+1})_j (z_i)_j \right) \right) \\
&= \underset{z}{\operatorname{argmin}} \sum_{g=1}^{n} \left(\frac{1}{2s}(z_g - v_g)^2 + \lambda_0\|z_g\|_0 + \frac{k_g \rho}{2} z_g^2 - z_g \sum_{G(i,j)=g} \left((y_i^k)_j + \rho(x_i^{k+1})_j \right) \right) \\
&= \underset{z}{\operatorname{argmin}} \sum_{g=1}^{n} \left(\frac{1}{2s} z_g^2 - \frac{1}{s} v_g z_g + \frac{k_g \rho}{2} z_g^2 - z_g \sum_{G(i,j)=g} \left((y_i^k)_j + \rho(x_i^{k+1})_j \right) + \lambda_0\|z_g\|_0 \right) \\
&= \underset{z}{\operatorname{argmin}} \sum_{g=1}^{n} \left((\frac{1}{2s} + \frac{k_g \rho}{2}) z_g^2 - \frac{v_g}{s} z_g - z_g \sum_{G(i,j)=g} \left((y_i^k)_j + \rho(x_i^{k+1})_j \right) + \lambda_0\|z_g\|_0 \right) \\
&= \underset{z}{\operatorname{argmin}} \sum_{g=1}^{n} \left((\frac{1}{2s} + \frac{k_g \rho}{2}) z_g^2 - z_g \left(\frac{v_g}{s} + \sum_{G(i,j)=g} \left((y_i^k)_j + \rho(x_i^{k+1})_j \right) \right) + \lambda_0\|z_g\|_0 \right) \\
&= \underset{z}{\operatorname{argmin}} \sum_{g=1}^{n} \left(\frac{1}{2} z_g^2 - z_g \frac{v_g/s + \sum_{G(i,j)=g} \left((y_i^k)_j + \rho(x_i^{k+1})_j \right)}{1/s + k_g \rho} + \frac{\lambda_0}{1/s + k_g \rho}\|z_g\|_0 \right)
\end{aligned}
$$

equivalently,

$$
\begin{aligned}
z_g^{k+1} &= \underset{z_g}{\operatorname{argmin}} \left(\frac{1}{2} z_g^2 - z_g \frac{v_g/s + \sum_{G(i,j)=g} \left((y_i^k)_j + \rho(x_i^{k+1})_j \right)}{1/s + k_g \rho} + \frac{\lambda_0}{1/s + k_g \rho}\|z_g\|_0 \right) \\
&= \underset{z_g}{\operatorname{argmin}} \left(\frac{1}{2} \left(z_g - \frac{v_g/s + \sum_{G(i,j)=g} \left((y_i^k)_j + \rho(x_i^{k+1})_j \right)}{1/s + k_g \rho} \right)^2 + \frac{\lambda_0}{1/s + k_g \rho}\|z_g\|_0 \right) \\
&= H_{\sqrt{2\lambda_0/(1/s + k_g \rho)}} \left(\frac{v_g/s + \sum_{G(i,j)=g} \left((y_i^k)_j + \rho(x_i^{k+1})_j \right)}{1/s + k_g \rho} \right), \quad \forall g \in \{1, 2, \ldots, n\}
\end{aligned}
$$

where $H_\lambda(\cdot)$ is a hard-thresholding operator defined as follows.

$$
H_\lambda(u) = \begin{cases} u, & \text{if } |u| > \lambda \\ 0, & \text{otherwise.} \end{cases}
$$

Note that when u is a vector, $H_\lambda(\cdot)$ is an element-wise hard-thresholding operator.

Thus, we obtain the final update formulas for x, z and y as follows.

$$
x_{G_i}^{k+1} = S_{\lambda_1/\rho}(z_i^{k+1} - \frac{y_{G_i}^k}{\rho}), \quad \forall i \in \{1, 2, \ldots, m\}
$$

$$z_g^{k+1} = H_{\sqrt{2\lambda_0/(1/s+k_g\rho)}}\left(\frac{v_g/s + \sum_{G(i,j)=g}\left((y_i^k)_j + \rho(x_i^{k+1})_j\right)}{1/s + k_g\rho}\right), \quad \forall g \in \{1,2,\ldots,n\}$$
$$y_{G_i}^{k+1} = y_{G_i}^k + \rho(x_{G_i}^{k+1} - z_i^{k+1}), \quad \forall i \in \{1,2,\ldots,m\}$$

1.1 The matrix form of the ℓ_0 sparse overlapping group lasso

Define $\tilde{x} = [x_{G_1}^T, x_{G_2}^T, \ldots, x_{G_m}^T]^T \in \mathbf{R}^{\tilde{n}}$ and $\tilde{z} = [z_{G_1}^T, z_{G_2}^T, \ldots, z_{G_m}^T]^T \in \mathbf{R}^{\tilde{n}}$ where $\tilde{n} = \sum_{i=1}^m n_i$. Then $\tilde{z}$ can be represented as

$$\tilde{z} = Gz$$

where each row of G has only one entry being 1 and other entries being 0. The corresponding definition of z-update becomes

$$\begin{aligned}
z^{k+1} &= \underset{z}{\operatorname{argmin}}\left(\frac{1}{2s}\|z-v\|_2^2 + \lambda_0\|z\|_0 + \frac{\rho}{2}\|\tilde{x}^k - Gz\|_2^2 - (\tilde{y}^k)^T Gz\right) \\
&= \underset{z}{\operatorname{argmin}}\left(\frac{1}{2}\left(\frac{z^Tz}{s} + \rho z^T G^T G z\right) - \left(\frac{v^Tz}{s} + (\rho\tilde{x}^k)^T Gz + (\tilde{y}^k)^T Gz\right) + \lambda_0\|z\|_0\right) \\
&= \underset{z}{\operatorname{argmin}}\left(\frac{1}{2}z^T\left(\frac{I}{s} + \rho G^T G\right)z - \left(\frac{v^Tz}{s} + (\rho\tilde{x}^k)^T Gz + (\tilde{y}^k)^T Gz\right) + \lambda_0\|z\|_0\right) \\
&= \underset{z}{\operatorname{argmin}}\left(\frac{1}{2}z^T\left(\frac{I}{s} + \rho G^T G\right)z - \left(\frac{v}{s} + G^T(\rho\tilde{x}^k + \tilde{y}^k)\right)^T z + \lambda_0\|z\|_0\right)
\end{aligned}$$

where $\tilde{y} = [y_{G_1}^T, y_{G_2}^T, \ldots, y_{G_m}^T]^T \in \mathbf{R}^{\tilde{n}}$. We observe that G^TG is a diagonal matrix of which the g-th diagonal entry corresponds to the number of groups that the global variable z_g involves. Since G^TG is positive semidefinite and $s, \rho > 0$, $I/s + \rho G^TG$ is definitely a positive definite matrix. Let $I/s + \rho G^TG = \operatorname{diag}(c_1, \ldots, c_n)$ and $C = \operatorname{diag}(\sqrt{c_1}, \ldots, \sqrt{c_n})$ where $c_g = 1/s + k_g\rho$. Thus, $C^TC = I/s + \rho G^TG$. With this setting, we have

$$\begin{aligned}
z^{k+1} &= \underset{z}{\operatorname{argmin}}\left(\frac{1}{2}z^TC^TCz - \left(\frac{v}{s} + G^T(\rho\tilde{x}^k + \tilde{y}^k)\right)^T C^{-1}Cz + \lambda_0\|z\|_0\right) \\
&= \underset{z}{\operatorname{argmin}}\left(\frac{1}{2}\|Cz - C^{-1}\left(\frac{v}{s} + G^T(\rho\tilde{x}^k + \tilde{y}^k)\right)\|_2^2 + \lambda_0\|z\|_0\right).
\end{aligned}$$

Since C is diagonal, the z-update reduces to n subproblems as follows.

$$\begin{aligned}
z_g^{k+1} &= \underset{z_g}{\operatorname{argmin}}\left(\frac{1}{2}\left(\sqrt{c_g}z_g - \frac{1}{\sqrt{c_g}}\cdot\left[\frac{v}{s} + G^T(\rho\tilde{x}^k + \tilde{y}^k)\right]_g\right)^2 + \lambda_0\|z_g\|_0\right) \\
&= \underset{z_g}{\operatorname{argmin}}\left(\frac{1}{2}\left(z_g - \frac{1}{c_g}\cdot\left[\frac{v}{s} + G^T(\rho\tilde{x}^k + \tilde{y}^k)\right]_g\right)^2 + \frac{\lambda_0}{c_g}\|z_g\|_0\right) \\
&= H_{\sqrt{2\lambda_0/c_g}}\left(\frac{1}{c_g}\cdot\left[\frac{v}{s} + G^T(\rho\tilde{x}^k + \tilde{y}^k)\right]_g\right)
\end{aligned}$$

which is exactly the same as the previous counterpart result.

1.2 Solving the dual problem via ADMM

The dual of (2) is,

$$\begin{aligned}
&\min_{x_{G_i},z}\frac{1}{2s}\|z-v\|^2 + \lambda_0\|z\|_0 + \lambda_1\sum_{i=1}^m\|x_{G_i}\|_2 + \sum_{i=1}^m y_{G_i}^T(x_{G_i} - z_i) \\
&= \min_z\left(\frac{1}{2s}\|z-v\|^2 + \lambda_0\|z\|_0 - \sum_{i=1}^m y_{G_i}^T z_i\right) + \min_{x_{G_i}}\left(\sum_{i=1}^m(\lambda_1\|x_{G_i}\|_2 + y_{G_i}^T x_{G_i})\right) \\
&= \min_z\left(\frac{1}{2s}(z^Tz - 2v^Tz + v^Tv) - \tilde{y}^T\tilde{z} + \lambda_0\|z\|_0\right) + \sum_{i=1}^m\min_{x_{G_i}}\left(\lambda_1\|x_{G_i}\|_2 + y_{G_i}^T x_{G_i}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \min_z \left(\frac{1}{2s}(z^T z - 2v^T z - 2s\tilde{y}^T G z + v^T v) + \lambda_0\|z\|_0\right) + \sum_{i=1}^m \min_{x_{G_i}} \left(\lambda_1\|x_{G_i}\|_2 + y_{G_i}^T x_{G_i}\right) \\
&= \min_z \left(\frac{1}{2s}(z^T z - 2(v + sG^T\tilde{y})^T z + v^T v) + \lambda_0\|z\|_0\right) + \sum_{i=1}^m \min_{x_{G_i}} \left(\lambda_1\|x_{G_i}\|_2 + y_{G_i}^T x_{G_i}\right) \\
&= \min_z \left(\frac{1}{2s}\left(\|z - (v + sG^T\tilde{y})\|^2 + v^T v - (v + sG^T\tilde{y})^T(v + sG^T\tilde{y})\right) + \lambda_0\|z\|_0\right) + \sum_{i=1}^m \min_{x_{G_i}} \left(\lambda_1\|x_{G_i}\|_2 + y_{G_i}^T x_{G_i}\right) \\
&= \min_z \left(\frac{1}{2s}\|z - (v + sG^T\tilde{y})\|^2 + \lambda_0\|z\|_0\right) + \sum_{i=1}^m \min_{x_{G_i}} \left(\lambda_1\|x_{G_i}\|_2 + y_{G_i}^T x_{G_i}\right) + \frac{1}{2s}(\|v\|^2 - \|v + sG^T\tilde{y}\|^2) \\
&= \min_z \left(\frac{1}{2s}\|z - (v + sG^T\tilde{y})\|^2 + \lambda_0\|z\|_0\right) - \frac{1}{2s}\|v + sG^T\tilde{y}\|^2 + \frac{1}{2s}\|v\|^2, \quad \|y_{G_i}\|_2 \leq \lambda_1, i = 1, \ldots, m
\end{aligned}
$$

After dropping the constant term, the dual problem of (2) becomes

$$
\max_{\tilde{y}\in\Omega} \min_{z\in\mathbf{R}^n} \left\{\psi(z, \tilde{y}) = \frac{1}{2s}\|z - (v + sG^T\tilde{y})\|^2 + \lambda_0\|z\|_0 - \frac{1}{2s}\|v + sG^T\tilde{y}\|^2\right\}. \tag{7}
$$

where Ω is defined as follows:

$$
\Omega = \{\tilde{y} \in \mathbf{R}^{\tilde{n}} \mid \|y_{G_i}\|_2 \leq \lambda_1, i = 1, 2, \ldots, m\}.
$$

For a given $\tilde{y}^{k-1}$, the optimal z minimizing $\psi(z, \tilde{y}^{k-1})$ in (7) is given by

$$
z^k = H_{\sqrt{2s\lambda_0}}(v + sG^T\tilde{y}^{k-1}). \tag{8}
$$

Plugging (8) into (7), we get the following maximization problem with respect to $\tilde{y}$:

$$
\max_{\tilde{y}\in\Omega} \{\omega(\tilde{y}) = -\psi(z^k, \tilde{y})\} \tag{9}
$$

which is equivalent to the following problem

$$
\max_{\tilde{y}\in\Omega} \ \tilde{y}^T G(z^k - 2v) \tag{10}
$$

which can be solved analytically as follows.

$$
\tilde{y}_{G_i} = \frac{[G(z^k - 2v)]_i}{\|[G(z^k - 2v)]_i\|_2} \tag{11}
$$

where $[G(z^k - 2v)]_i \in \mathbf{R}^{n_i}$ denotes the counterpart corresponding to the group G_i. Finally, our methodology for minimizing the problem defined in (2) is to alternate update z and $\tilde{y}$.

1.3 The bounds on the optimal value of the overlapping group lasso

Before presenting the results regarding the bounds of the optimal value of the ℓ_0 sparse group lasso, we introduce three lemmas which lead to the upcoming theorem. For completeness, we describe a well known result as the following lemma, namely the quadratic mean (QM) is no less than the arithmetic mean (AM).

Lemma 1.1 (QM≥AM). *Given* $\mathbf{x} \in \mathbf{R}^n$*, the following*

$$
\sqrt{\frac{\sum_{i=1}^n x_i^2}{n}} \geq \frac{\sum_{i=1}^n |x_i|}{n}
$$

holds. The equality holds if and only if $x_1 = x_2 = \ldots = x_n$.

Proof. According to Cauchy-Schwartz inequality which says that given two vectors $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$, $\|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \geq |\mathbf{x}^T\mathbf{y}|$, we have

$$\begin{aligned}\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} &= \sqrt{\sum_{i=1}^{n} (\frac{|x_i|}{\sqrt{n}})^2} \cdot \overbrace{\sqrt{\sum_{i=1}^{n} (\frac{1}{\sqrt{n}})^2}}^{=1} \\ &\geq \frac{\sum_{i=1}^{n} |x_i|}{n} \geq \frac{\sum_{i=1}^{n} x_i}{n},\end{aligned}$$

where the equalities in the first and second inequalities hold if and only if $|x_1| = |x_2| = \cdots = |x_n|$ and $x_1 = x_2 = \cdots = x_n$, respectively. This completes the proof. □

1.3.1 Lower bound on the overlapping group lasso

Lemma 1.2 (lower bound on the overlapping group lasso). *Given $\mathbf{x} \in \mathbf{R}^n$, $\mathbf{w} \in \mathbf{R}^m_{++}$ and some groups $G_i \subseteq \{1, 2, \ldots, n\}, i = 1, 2, \ldots, m$, let $I_{G_i}(j)$ denote an indicator function whose value is 1 if $j \in G_i$ and 0 otherwise. Then, the following*

$$\|\mathbf{L}\mathbf{x}\|_1 \leq \sum_{i=1}^{m} w_i \|\mathbf{x}_{G_i}\|_2$$

holds, where $\mathbf{L} = \mathbf{diag}(\mathbf{l})$ with elements $l_j = \sum_{i=1}^{m} \frac{w_i}{\sqrt{|G_i|}} \odot \mathbb{I}(j \in G_i), j = 1, \ldots, n$ and $\mathbb{I}(e) = 1$ if e is true, 0 otherwise. The equality holds if and only if for every G_i, the entries of $\mathbf{x}_{G_i}$ are identical.

Proof.

$$\begin{aligned}\sum_{i=1}^{m} w_i \|\mathbf{x}_{G_i}\|_2 &= w_1 \|\mathbf{x}_{G_1}\|_2 + \cdots + w_m \|\mathbf{x}_{G_m}\|_2 \\ &= \sum_{i=1}^{m} w_i \sqrt{\sum_{j=1}^{|G_i|} |x_j|^2} = \sum_{i=1}^{m} w_i \sqrt{|G_i|} \cdot \sqrt{\frac{\sum_{j=1}^{|G_i|} |x_j|^2}{|G_i|}} \\ &\geq \sum_{i=1}^{m} w_i \sqrt{|G_i|} \cdot \frac{\sum_{j=1}^{|G_i|} |x_j|}{|G_i|} = \sum_{i=1}^{m} \frac{w_i}{\sqrt{|G_i|}} \cdot \sum_{j=1}^{|G_i|} |x_j| \\ &= \sum_{j=1}^{n} \sum_{i=1}^{m} \left(\frac{w_i}{\sqrt{|G_i|}} \odot \mathbb{I}(j \in G_i) \right) |x_j|\end{aligned}$$

where $\mathbb{I}(e)$ is an indicator function defined as follows:

$$\begin{cases} 1, & \text{if } e \text{ is true} \\ 0, & \text{otherwise.} \end{cases}$$

The second line follows from the definition of p-norm ($p \geq 1$) and the third line from Lemma 1.1. The equality holds if and only if the entries belonging to the same group are identical for all the groups. Let $\mathbf{L} = \mathbf{diag}(\mathbf{l})$, the diagonal matrix with elements $l_j = \sum_{i=1}^{m} \frac{w_i}{\sqrt{|G_i|}} \odot \mathbb{I}(j \in G_i), j = 1, \ldots, n$, so we have

$$\sum_{i=1}^{m} w_i \|\mathbf{x}_{G_i}\|_2 \geq \|\mathbf{L}\mathbf{x}\|_1 \tag{12}$$

□

1.3.2 Computing the lower bound on the overlapping group lasso

Since we have found the lower bound on the overlapping group lasso operator, the overlapping group lasso problem reduces to solving a weighted lasso problem as follows.

$$\min_{\mathbf{x}\in\mathbf{R}^n}\left\{f_{\text{GL\_lb}}(\mathbf{x}) := \frac{1}{2}\|\mathbf{x}-\mathbf{v}\|_2^2+\lambda\|\mathbf{L}\mathbf{x}\|_1\right\} \tag{13}$$

Since $f_{\text{GL\_lb}}(\mathbf{x})$ is separable w.r.t $\mathbf{x}$, this is equivalent to solving the following subproblem for each i.

$$\min_{x_i\in\mathbf{R}}\left\{f_{\text{GL\_lb}}(x_i) := \frac{1}{2}(x_i-v_i)^2+\lambda l_i|x_i|\right\}$$

If $x_i > 0$, then $f_{\text{GL\_lb}}(x_i) = \frac{1}{2}(x_i-v_i)^2+\lambda l_i x_i$. By the first-order optimality condition,

$$\nabla f_{\text{GL\_lb}}(x_i) = x_i - v_i + \lambda l_i = 0 \Longleftrightarrow x_i = v_i - \lambda l_i > 0 \Longleftrightarrow v_i > \lambda l_i.$$

For $x_i < 0$, we have the following similar argument.

$$\nabla f_{\text{GL\_lb}}(x_i) = x_i - v_i - \lambda l_i = 0 \Longleftrightarrow x_i = v_i + \lambda l_i < 0 \Longleftrightarrow v_i < -\lambda l_i.$$

In the case of $x_i = 0$, let ν be the subdifferential of $|x_i|$ at $x_i = 0$, then $\nu \in [-1, 1]$. Thus,

$$0 \in \partial f_{\text{GL\_lb}}(0) = 0 - v_i + \lambda l_i \nu \Longleftrightarrow \frac{v_i}{\lambda l_i} \in \nu \Longleftrightarrow |v_i| \le \lambda l_i$$

where the RHS follows from the fact that $|\nu| \le 1$. To sum up, the solution to the subproblem is

$$x_i = \begin{cases} v_i - \lambda l_i, & \text{if } v_i > \lambda l_i \\ 0, & \text{if } |v_i| \le \lambda l_i \\ v_i + \lambda l_i, & \text{if } v_i < -\lambda l_i. \end{cases} \quad i = 1, 2, \ldots, n. \tag{14}$$

1.3.3 Upper bound on the overlapping group lasso

Lemma 1.3 (upper bound on the overlapping group lasso). *Given $\mathbf{x} \in \mathbf{R}^n$, $\mathbf{w} \in \mathbf{R}^m_{++}$ and some groups $G_i \subseteq \{1, 2, \ldots, n\}, i = 1, \ldots, m$, denote the total number of appearances in all groups by $k_j, j = 1, \ldots, n$, and let $\mathbf{U} = \mathbf{diag}(\sqrt{k_1}\|\mathbf{w}\|_2, \ldots, \sqrt{k_n}\|\mathbf{w}\|_2)$. Then, the following*

$$\sum_{i=1}^m w_i\|\mathbf{x}_{G_i}\|_2 \le \|\mathbf{U}\mathbf{x}\|_2$$

holds. The equality holds if and only if $\frac{\|\mathbf{x}_{G_1}\|_2}{w_1} = \ldots = \frac{\|\mathbf{x}_{G_m}\|_2}{w_m}$.

Proof.

$$\begin{aligned}\sum_{i=1}^m w_i\|\mathbf{x}_{G_i}\|_2 &= w_1\|\mathbf{x}_{G_1}\|_2 + \cdots + w_m\|\mathbf{x}_{G_m}\|_2 \\ &\le \sqrt{w_1^2+\cdots+w_m^2}\cdot\sqrt{\|\mathbf{x}_{G_1}\|_2^2+\cdots+\|\mathbf{x}_{G_m}\|_2^2} \\ &= \sqrt{\sum_{i=1}^m w_i^2}\cdot\sqrt{\sum_{g=1}^n k_g x_g^2} = \sqrt{\sum_{g=1}^n(\sum_{i=1}^m w_i^2)k_g x_g^2}\end{aligned}$$

where the second line follows from Cauchy-Schwarz inequality and the equality holds if and only if $\frac{\|\mathbf{x}_{G_1}\|_2}{w_1} = \ldots = \frac{\|\mathbf{x}_{G_m}\|_2}{w_m}$. Let $\mathbf{U}_0 = \mathbf{diag}(\sqrt{k_1}, \ldots, \sqrt{k_n})$ and $\mathbf{w} = (w_1, \ldots, w_m)$, then we have

$$\sum_{i=1}^m w_i\|\mathbf{x}_{G_i}\|_2 \le \|\mathbf{w}\|_2\cdot\sqrt{\mathbf{x}^T\,\mathbf{diag}(k_1, \ldots, k_n)\mathbf{x}} = \|\mathbf{w}\|_2\cdot\|\mathbf{U}_0\mathbf{x}\|_2 \tag{15}$$

By the positive homogeneity of $\|\cdot\|_2$, $\|\mathbf{w}\|_2$ can be absorbed into $\mathbf{U}_0$ as follows.

$$\sum_{i=1}^m w_i\|\mathbf{x}_{G_i}\|_2 \le \|\mathbf{U}\mathbf{x}\|_2 \tag{16}$$

where $\mathbf{U} = \mathbf{diag}(\mathbf{u})$ and $\mathbf{u} = (\sqrt{k_1}\|\mathbf{w}\|_2, \ldots, \sqrt{k_n}\|\mathbf{w}\|_2)$. □

1.3.4 Computing the upper bound on the overlapping group lasso

After replacing the overlapping group lasso operator with $\|\mathbf{Ux}\|_2$, the upper bound on the overlapping group lasso is equal to the optimal value of the following problem.

$$\min_{\mathbf{x}\in\mathbf{R}^n}\left\{f_{\text{GL\_ub}}(\mathbf{x}) := \frac{1}{2}\|\mathbf{x}-\mathbf{v}\|_2^2+\lambda\|\mathbf{Ux}\|_2\right\} \tag{17}$$

Let $\mathbf{s}$ be an element of the subdifferential of $\|\cdot\|_2$ at $\mathbf{0}$. Then if $\mathbf{x}=\mathbf{0}$, solving the following zero subgradient equation gives

$$\mathbf{0}-\mathbf{v}+\lambda\mathbf{U}^T\mathbf{s}=\mathbf{0} \Longleftrightarrow \mathbf{v}=\lambda\mathbf{U}^T\mathbf{s} \Longleftrightarrow \mathbf{U}^{-1}\mathbf{v}=\lambda\mathbf{s} \Longleftrightarrow \|\mathbf{U}^{-1}\mathbf{v}\|_2\leq\lambda$$

Thus, we obtain that if $\|\mathbf{U}^{-1}\mathbf{v}\|_2\leq\lambda$, the optimal minimizer is $\mathbf{0}$.

When $\mathbf{x}\neq\mathbf{0}$, we have

$$\nabla f_{\text{GL\_ub}}(\mathbf{x})=\mathbf{x}-\mathbf{v}+\frac{\lambda\mathbf{U}^T\mathbf{Ux}}{\|\mathbf{Ux}\|_2}=0 \Longleftrightarrow \left(\mathbf{I}+\frac{\lambda\mathbf{U}^T\mathbf{U}}{\|\mathbf{Ux}\|_2}\right)\mathbf{x}=\mathbf{v} \Longleftrightarrow \mathbf{x}=\left(\mathbf{I}+\frac{\lambda\mathbf{U}^T\mathbf{U}}{\|\mathbf{Ux}\|_2}\right)^{-1}\mathbf{v}.$$

Thus, we reformulate the optimality condition as $\mathbf{x}=T(\mathbf{x})$ for the case of $\mathbf{x}\neq\mathbf{0}$, where T is the operator

$$T(\mathbf{x}) := \left(\mathbf{I}+\frac{\lambda\mathbf{U}^T\mathbf{U}}{\|\mathbf{Ux}\|_2}\right)^{-1}\mathbf{v}. \tag{18}$$

Then we have the following result concerning $T(\mathbf{x})$.

Theorem 1.4. *$T(\mathbf{x})$ has a unique fixed point $\mathbf{x}^*$. In other words, the corresponding fixed point iteration*

$$\mathbf{x}^{(k+1)} := T(\mathbf{x}^{(k)})$$

converges to a unique $\mathbf{x}^*$.

Proof. Since $\mathbf{U}$ is a diagonal matrix, $\mathbf{U}^T\mathbf{U}=\mathbf{U}^2$. Let $\mathbf{G}=\mathbf{I}+\frac{\lambda\mathbf{U}^T\mathbf{U}}{\|\mathbf{Ux}\|_2}$, then $\mathbf{G}$ is a diagonal matrix with elements $G_{ii}=1+\frac{\lambda u_i^2}{\|\mathbf{Ux}\|_2}, i=1,2,\ldots,n$. Thus, $\mathbf{G}^{-1}$ is also a diagonal matrix with elements

$$0<\rho_i=\frac{\|\mathbf{Ux}\|_2}{\|\mathbf{Ux}\|_2+\lambda u_i^2}=1-\frac{\lambda u_i^2}{\|\mathbf{Ux}\|_2+\lambda u_i^2}<1. \tag{19}$$

Then we have $x_i=\rho_i v_i$ for each i. Let $\boldsymbol{\rho}$ be a vector whose i-th entry is ρ_i with $\rho_i\in(0,1)$. So, $\mathbf{x}=\boldsymbol{\rho}\odot\mathbf{v}$ where $\odot$ is the element-wise Hadamard product, which indicates $\mathbf{x}$ is a contracted version of $\mathbf{v}$. Let $\rho_i(y)=1-\frac{\lambda u_i^2}{y+\lambda u_i^2}$ with $y=\|\mathbf{Ux}\|_2>0$. Then

$$\rho_i'(y)=\frac{\lambda u_i^2}{(y+\lambda u_i^2)^2}>0.$$

which shows ρ_i is a strictly increasing function of y, i.e., $\|\mathbf{Ux}\|_2$. By (19), a smaller (bigger) $\|\mathbf{Ux}\|_2$ gives a smaller (bigger) ρ_i for each i which in turn generates smaller (bigger) x_i via $x_i=\rho_i v_i$ for each i, and then small (greater) $\|\mathbf{Ux}\|_2$. Thanks to this interplay between $\|\mathbf{Ux}\|_2$ and $\boldsymbol{\rho}$, the sequences regarding $\|\mathbf{Ux}\|_2$ and $\boldsymbol{\rho}$ generated by performing $T(\mathbf{x})$ are monotone. For example, let us start the iteration with $\mathbf{x}^{(0)}\neq\mathbf{0}$. Then $\mathbf{x}^{(1)}=\boldsymbol{\rho}^{(0)}\odot\mathbf{v}$. Suppose $\|\mathbf{Ux}^{(0)}\|_2>\|\mathbf{Ux}^{(1)}\|_2$, then

$$\rho_i^{(0)}=1-\frac{\lambda u_i^2}{\|\mathbf{Ux}^{(0)}\|_2+\lambda u_i^2}>1-\frac{\lambda u_i^2}{\|\mathbf{Ux}^{(1)}\|_2+\lambda u_i^2}=\rho_i^{(1)},\quad i=1,2,\ldots,n. \tag{20}$$

Thus,

$$x_i^{(2)}=\rho_i^{(1)}v_i<\rho_i^{(0)}v_i=x_i^{(1)},\quad i=1,2,\ldots,n.$$

Since $\mathbf{U}$ is a diagonal matrix with nonnegative diagonal entries u_i, we have

$$\|\mathbf{Ux}^{(2)}\|_2=\sqrt{\sum_{i=1}^{n}(u_ix_i^{(2)})^2}<\sqrt{\sum_{i=1}^{n}(u_ix_i^{(1)})^2}=\|\mathbf{Ux}^{(1)}\|_2$$

So, $\|\mathbf{U}\mathbf{x}^{(0)}\|_2 > \|\mathbf{U}\mathbf{x}^{(1)}\|_2 > \|\mathbf{U}\mathbf{x}^{(2)}\|_2$. Substituting $\|\mathbf{U}\mathbf{x}^{(1)}\|_2$ and $\|\mathbf{U}\mathbf{x}^{(2)}\|_2$ into (20), we get $\rho_i^{(1)} > \rho_i^{(2)} > \rho_i^{(3)}$. By repeating this, the contraction interplay between $\|\mathbf{U}\mathbf{x}^{(k)}\|_2$ and $\rho_i(k)$ lead to that both $\{\|\mathbf{U}\mathbf{x}^{(k)}\|_2\}$ and $\{\rho_i^{(k)}\}_{i=1,2,\ldots,n}$ are decreasing sequences. Also, $\{\|\mathbf{U}\mathbf{x}^{(k)}\|_2\}$ and $\{\rho_i^{(k)}\}$ are both bounded below by 0. Since monotone bounded sequences converge, $\{\|\mathbf{U}\mathbf{x}^{(k)}\|_2\}$ and $\{\rho_i^{(k)}\}$ are convergent. Assuming $\lim_{k\to\infty}\|\mathbf{U}\mathbf{x}^{(k)}\|_2 = c$ and $\lim_{k\to\infty}\rho_i^{(k)} = \rho_i$ yield

$$x_i = \frac{cv_i}{c+\lambda u_i^2}, \quad i = 1, 2, \ldots, n.$$

Multiplying both sides by u_i, squaring both sides and summing over i gives

$$c^2 = \sum_{i=1}^{n}(u_i x_i)^2 = \sum_{i=1}^{n}\frac{(cu_i v_i)^2}{(c+\lambda u_i^2)^2} \Longrightarrow 1 = \sum_{i=1}^{n}\frac{(u_i v_i)^2}{(c+\lambda u_i^2)^2} \tag{21}$$

The solution to the equation on the RHS of (21) is c which is unique since $c = \|\mathbf{U}\mathbf{x}^*\|_2 > 0$. We can show this by contradiction. Specifically, suppose $c' > c$ is the solution to (21). Since $\lambda, u_i > 0, i = 1, 2, \ldots, n$, then we get

$$\frac{(u_i v_i)^2}{(c'+\lambda u_i^2)^2} < \frac{(u_i v_i)^2}{(c+\lambda u_i^2)^2} \Longrightarrow \sum_{i=1}^{n}\frac{(u_i v_i)^2}{(c'+\lambda u_i^2)^2} < \sum_{i=1}^{n}\frac{(u_i v_i)^2}{(c+\lambda u_i^2)^2} = 1$$

which contradicts the supposition $c' > c$ is the solution to (21), i.e., $\sum_{i=1}^{n}\frac{(u_i v_i)^2}{(c'+\lambda u_i)^2} = 1$. Similar arguments hold for the case when $c' < c$. Thus, c, i.e., $\|\mathbf{U}\mathbf{x}^*\|_2$ is unique. Furthermore, $\boldsymbol{\rho}$ is unique because of $\rho_i = 1 - \frac{\lambda u_i^2}{\|\mathbf{U}\mathbf{x}\|_2 + \lambda u_i^2}, i = 1, 2, \ldots, n$.

For the case of $\|\mathbf{U}\mathbf{x}^{(0)}\|_2 < \|\mathbf{U}\mathbf{x}^{(1)}\|_2$, similar arguments give increasing and bounded sequences $\{\|\mathbf{U}\mathbf{x}^{(k)}\|_2\}$ and $\{\rho_i^{(k)}\}$. Thus, they are convergent as well. If $\|\mathbf{U}\mathbf{x}^{(0)}\|_2 = \|\mathbf{U}\mathbf{x}^{(1)}\|_2$, we luckily hit the fixed point in one step. Finally, $\mathbf{x}^*$ is unique due to $\mathbf{x}^* = \boldsymbol{\rho} \odot \mathbf{v}$. Therefore, $T(\mathbf{x})$ is a fixed point operator. This completes our proof. □

1.4 The bounds on the optimal value of the ℓ_1 sparse overlapping group lasso

We have found the bounds for the overlapping group lasso operator in the previous section. Now it is natural to transform the bounds on the optimal value of the ℓ_1 sparse overlapping group lasso into solving two problems.

1.4.1 Lower bound on the optimal value of the ℓ_1 sparse overlapping group lasso

The lower bound can be obtained by solving the following problem.

$$\min_{\mathbf{x}\in\mathbf{R}^n}\left\{f_{\ell_1\text{\_GL\_lb}}(\mathbf{x}) := \frac{1}{2}\|\mathbf{x}-\mathbf{v}\|_2^2 + \lambda\|\mathbf{L}\mathbf{x}\|_1 + \lambda_1\|\mathbf{x}\|_1\right\}. \tag{22}$$

Since $\mathbf{L}$ is a diagonal matrix, it can be rewritten as

$$\min_{\mathbf{x}\in\mathbf{R}^n}\left\{f_{\ell_1\text{\_GL\_lb}}(\mathbf{x}) := \frac{1}{2}\|\mathbf{x}-\mathbf{v}\|_2^2 + \|(\lambda\mathbf{L}+\lambda_1\mathbf{I})\mathbf{x}\|_1\right\}. \tag{23}$$

which shares the same form as (13) and can be solved in a similar way. For brevity, we present its solution directly as follows.

$$x_i = \begin{cases} v_i - \lambda l_i, & \text{if } v_i > \lambda l_i + \lambda_1 \\ 0, & \text{if } |v_i| \le \lambda l_i + \lambda_1 \\ v_i + \lambda l_i, & \text{if } v_i < -\lambda l_i - \lambda_1. \end{cases} \quad i = 1, 2, \ldots, n.$$

1.4.2 Upper bound on the optimal value of the ℓ_1 sparse overlapping group lasso

The upper bound can be obtained by solving the following problem.

$$\min_{\mathbf{x}\in\mathbf{R}^n}\left\{f_{\ell_1\text{\_GL\_ub}}(\mathbf{x}) := \frac{1}{2}\|\mathbf{x}-\mathbf{v}\|_2^2+\lambda\|\mathbf{U}\mathbf{x}\|_2+\lambda_1\|\mathbf{x}\|_1\right\}. \tag{24}$$

Let $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ be the subdifferentials of $\|\mathbf{x}\|_2$ and $\|\mathbf{x}\|_1$ at $\mathbf{x}=\mathbf{0}$, where $\|\boldsymbol{\mu}\|_2\leq 1$ and $\|\boldsymbol{\nu}\|_\infty\leq 1$. Δ and Λ are defined as

$$\Delta = \{\boldsymbol{\mu}\in\mathbf{R}^n \mid \|\boldsymbol{\mu}\|_2\leq 1\},\quad \Lambda = \{\boldsymbol{\nu}\in\mathbf{R}^n \mid \|\boldsymbol{\nu}\|_\infty\leq 1\}.$$

If $\mathbf{x}=\mathbf{0}$, using the first-order optimality condition gives

$$\mathbf{0}\in\mathbf{0}-\mathbf{v}+\lambda\mathbf{U}^T\boldsymbol{\mu}+\lambda_1\boldsymbol{\nu} \iff \mathbf{U}^{-1}\mathbf{v}\in\lambda\boldsymbol{\mu}+\lambda_1\mathbf{U}^{-1}\boldsymbol{\nu} \iff \mathbf{U}^{-1}(\mathbf{v}-\lambda_1\boldsymbol{\nu}) = \lambda\boldsymbol{\mu} \iff \max_{\boldsymbol{\nu}\in\Lambda}\|\mathbf{U}^{-1}(\mathbf{v}-\lambda_1\boldsymbol{\nu})\|_2\leq\lambda$$

Since $\mathbf{U}^{-1}$ is also a diagonal matrix with positive diagonal entries, the maximum value of $\|\mathbf{U}^{-1}(\mathbf{v}-\lambda_1\boldsymbol{\nu})\|_2$ is attained at $\boldsymbol{\nu}=-\operatorname{sgn}(\mathbf{v})\odot\mathbf{1}$. Here, sgn is an elementwise sign function whose value is 1 for positive inputs, -1 for negative inputs and 0 otherwise, and $\mathbf{1}$ is a vector with all entries being 1. With these settings, we have

$$\max_{\boldsymbol{\nu}\in\Lambda}\|\mathbf{U}^{-1}(\mathbf{v}-\lambda_1\boldsymbol{\nu})\|_2 = \|\mathbf{U}^{-1}\left(\mathbf{v}+\lambda_1\operatorname{sgn}(\mathbf{v})\odot\mathbf{1}\right)\|_2 = \sqrt{\sum_{i=1}^{n}\left(\frac{v_i+\lambda_1\operatorname{sgn}(v_i)}{u_i}\right)^2}\leq\lambda$$

Hence, we get that $\mathbf{x}=\mathbf{0}$ if and only if $\|\mathbf{U}^{-1}\left(\mathbf{v}+\lambda_1\operatorname{sgn}(\mathbf{v})\odot\mathbf{1}\right)\|_2\leq\lambda$.

Now we talk about the case of $\mathbf{x}\neq\mathbf{0}$. By the first-order optimality condition, we have

$$\mathbf{0}\in\mathbf{x}-\mathbf{v}+\lambda\frac{\mathbf{U}^T\mathbf{U}\mathbf{x}}{\|\mathbf{U}\mathbf{x}\|_2}+\lambda_1\mathbf{x}'$$

where $\mathbf{x}'$ denotes the subdifferential of $\|\mathbf{x}\|$ at $\mathbf{x}\neq\mathbf{0}$ defined as follows.

$$x_i' = \begin{cases} 1, & \text{if } x_i>0\\ -1, & \text{if } x_i<0\\ \nu\in[-1,1], & \text{if } x_i=0.\end{cases}$$

If $x_i=0$, we have

$$x_i=0\iff 0\in 0-v_i+0+\lambda_1\nu\iff v_i\in\lambda_1\nu\iff |v_i|\leq\lambda_1.$$

By contradiction, if $x_i\neq 0$, it is clear to see that the optimal x_i shares the common sign with v_i, otherwise it will lead to greater objective values. Thus, we get

$$\tilde{\mathbf{x}}+\lambda\frac{\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}\tilde{\mathbf{x}}}{\|\tilde{\mathbf{U}}\tilde{\mathbf{x}}\|_2} = \tilde{\mathbf{v}}-\lambda_1\operatorname{sgn}(\tilde{\mathbf{v}})\odot\mathbf{1}\iff\tilde{\mathbf{x}} = T_{\ell_1\text{\_ub}}\left(\mathbf{I}+\lambda\frac{\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}\tilde{\mathbf{x}}}{\|\tilde{\mathbf{U}}\tilde{\mathbf{x}}\|_2}\right)^{-1}(\tilde{\mathbf{v}}-\lambda_1\operatorname{sgn}(\tilde{\mathbf{v}})\odot\mathbf{1})$$

where $\tilde{\mathbf{x}}$ represents the reduced $\mathbf{x}$ after removing zero entries, and $\tilde{\mathbf{v}},\tilde{\mathbf{U}}$ are the corresponding notations. By Theorem 1.4, $T_{\ell_1\text{\_ub}}(\tilde{\mathbf{x}})$ is a fixed point operator.

1.5 The bounds on the optimal value of the ℓ_0 sparse overlapping group lasso

1.5.1 Lower bound on the optimal value of the ℓ_0 sparse overlapping group lasso

The lower bound can be obtained by solving the following problem.

$$\min_{\mathbf{x}\in\mathbf{R}^n}\left\{f_{\ell_0\text{\_GL\_lb}}(\mathbf{x}) := \frac{1}{2}\|\mathbf{x}-\mathbf{v}\|_2^2+\lambda\|\mathbf{L}\mathbf{x}\|_1+\lambda_0\|\mathbf{x}\|_0\right\}. \tag{25}$$

which is separable and can be divided into subproblems as follows.

$$\min_{x_i \in \mathbf{R}} \left\{ f_{\ell_0\_\mathrm{GL\_lb}}(x_i) := \frac{1}{2}(x_i - v_i)^2 + \lambda l_i \|x_i\|_1 + \lambda_0 \|x_i\|_0 \right\}. \tag{26}$$

If $x_i \neq 0$, then (26) can be reduced to solving the following simpler problem.

$$\min_{x_i \in \mathbf{R}} f_{\ell_0\_\mathrm{GL\_lb}}(x_i) = \frac{1}{2}(x_i - v_i)^2 + \lambda l_i |x_i| + \lambda_0$$

whose solution is given by (14), namely, $x_i = v_i - \lambda l_i \operatorname{sgn}(v_i)$. In this case, if $|v_i| > \lambda l_i$, the corresponding objective value is

$$f_{\ell_0\_\mathrm{GL\_lb}}(v_i - \lambda l_i \operatorname{sgn}(v_i)) = \frac{1}{2}(\lambda l_i)^2 + \lambda l_i |v_i - \lambda l_i \operatorname{sgn}(v_i)| + \lambda_0,$$

and if $|v_i| \leq \lambda l_i$, we are done and definitely $x_i = 0$. However, in the case of $|v_i| > \lambda l_i$, we still need to compare $\frac{1}{2}(\lambda l_i)^2 + \lambda l_i |v_i - \lambda l_i \operatorname{sgn}(v_i)| + \lambda_0$ with $f_{\ell_0\_\mathrm{GL\_lb}}(0) = \frac{1}{2}v_i^2$ due to the existence of the additional term λ_0. If $f_{\ell_0\_\mathrm{GL\_lb}}(0) \leq f_{\ell_0\_\mathrm{GL\_lb}}(v_i - \lambda l_i \operatorname{sgn}(v_i))$, the solution is 0 rather than $v_i - \lambda l_i \operatorname{sgn}(v_i)$.

1.5.2 Upper bound on the optimal value of the ℓ_0 sparse overlapping group lasso

The upper bound can be obtained by solving the following problem.

$$\min_{\mathbf{x} \in \mathbf{R}^n} \left\{ f_{\ell_0\_\mathrm{GL\_ub}}(\mathbf{x}) := \frac{1}{2}\|\mathbf{x} - \mathbf{v}\|_2^2 + \lambda \|\mathbf{U}\mathbf{x}\|_2 + \lambda_0 \|\mathbf{x}\|_0 \right\}. \tag{27}$$

By the definition of the induced norm of $\|\cdot\|_{a,b}$, we have $\|\mathbf{U}\mathbf{x}\|_2 \leq \|\mathbf{U}\|_2 \|\mathbf{x}\|_2$. When $a = b = 2$, $\|\mathbf{U}\|_2$ is called the spectral norm and it is equal to the maximum singular value of $\mathbf{U}$, denoted as $\sigma_{\max}(\mathbf{U})$. Thus, the upper bound can be relaxed as follows.

$$\min_{\mathbf{x} \in \mathbf{R}^n} \left\{ f_{\ell_0\_\mathrm{GL\_ub}}(\mathbf{x}) := \frac{1}{2}\|\mathbf{x} - \mathbf{v}\|_2^2 + \lambda \sigma_{\max}(\mathbf{U}) \|\mathbf{x}\|_2 + \lambda_0 \|\mathbf{x}\|_0 \right\}. \tag{28}$$

which has a closed-form solution proposed by Shao et al. (2022).

2 Acknowledgement

The inspiration of this work is from reading Boyd et al. (2011).

Bibliography

Boyd, S. P., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3:1–122.

Shao, Y., Zhao, K., Cao, Z., Peng, Z., Peng, X., Li, P., Wang, Y., and Ma, J. (2022). Mobileprune: Neural network compression via ℓ_0 sparse group lasso on the mobile system. *Sensors*, 22(11):4081.