A Proofs of Theorems 1 and 2

A.1 ML-IALM

In the subsection we give the missing proof of theorem 1. First let us remind a few definitions. At each iteration of IALM algorithm we solve the following subproblem

$$\min_{\mathbf{L}\in\mathbb{R}^{m\times n}}\{\|\mathbf{L}\|_\star + \frac{1}{2\tau}\|\mathbf{M}-\mathbf{L}\|_F^2\}. \tag{1}$$

The minimiser of (1) is given in closed form via the singular value thresholding operator as

$$\hat{\mathbf{L}} = \mathcal{D}_\tau[\mathbf{M}], \tag{2}$$

where $\mathcal{D}_\tau[\mathbf{M}]$ is defined for the SVD of $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ as

$$\mathcal{D}_\tau[\mathbf{M}] = \mathbf{U}\mathcal{S}_\tau[\Sigma]\mathbf{V}^\top, \tag{3}$$

where $\mathcal{S}_\tau[\cdot]$ is the soft thresholding operator defined element-wise:

$$\mathcal{S}_\tau[x] = (|x|-\tau)_+\mathrm{sgn}(x). \tag{4}$$

Then for a full rank restriction operator $\mathbf{R}\in\mathbb{R}^{n\times n_H}$ we define the multilevel SVT (ML-SVT) operator

$$\mathcal{D}_\tau^H[\mathbf{M}] = \mathbf{U}_H\mathcal{S}_\tau[\mathbf{\Sigma}_H]\mathbf{V}_H^\top\mathbf{R}^\top, \tag{5}$$

where $\mathbf{M}\mathbf{R} = \mathbf{U}_H\mathbf{\Sigma}_H\mathbf{V}_H^\top$ is the SVD of the coarse $\mathbf{M}_H = \mathbf{M}\mathbf{R}$. Then Theorem 1 shows that $\mathcal{D}_\tau^H$ gives an approximate solution to the problem (1).

Theorem 1. *For any $\mathbf{R}$, such that $\|\mathbf{R}\|_2 \le 1$ and $0 < \tau \le \sigma_{H,1}$, the ML-SVT operator $\mathcal{D}_\tau^H[\mathbf{M}]$ gives a $\left(\frac{\sigma_{H,1}}{\tau^2}(\sigma_1+\sigma_{H,1}-\tau)\right)$-approximate solution to the problem (1), where $\sigma_{H,1}$ is the largest singular value of $\mathbf{M}\mathbf{R}$.*

Proof. The proof follows the steps of the proof of Theorem 2.1 of [1]. First note that (1) is a convex problem with optimality criteria

$$\mathbf{M}-\hat{\mathbf{L}} \in \tau\partial\|\hat{\mathbf{L}}\|_\star, \tag{6}$$

where $\partial\|\cdot\|_\star$ is the set of subgradients of the nuclear norm. Let $\mathbf{L}\in\mathbb{R}^{m\times n}$ be an arbitrary matrix and $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ be its SVD. It is known [2] that

$$\partial\|\mathbf{L}\|_\star = \{\mathbf{U}\mathbf{V}^\top + \mathbf{W} : \mathbf{W}\in\mathbb{R}^{m\times n}, \mathbf{U}^\top\mathbf{W} = \mathbf{0}, \mathbf{W}\mathbf{V} = \mathbf{0}, \|\mathbf{W}\|_2 \le 1\}. \tag{7}$$

Next we set $\mathbf{L}^H = \mathcal{D}_\tau^H[\mathbf{M}]$ and find an upper bound for the distance from $\mathbf{M}-\mathbf{L}^H$ to the set $\tau\partial\|\mathbf{L}^H\|_\star$. Then we decompose the SVDs of $\mathbf{M}$ and $\mathbf{M}\mathbf{R}$ as $\mathbf{M} = \mathbf{U}_0\mathbf{\Sigma}_0\mathbf{V}_0^\top + \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^\top$ and $\mathbf{M}\mathbf{R} = \mathbf{U}_{H,0}\mathbf{\Sigma}_{H,0}\mathbf{V}_{H,0}^\top + \mathbf{U}_{H,1}\mathbf{\Sigma}_{H,1}\mathbf{V}_{H,1}^\top$, where $\mathbf{U}_0$, $\mathbf{U}_{H,0}$, $\mathbf{V}_0$ and $\mathbf{V}_{H,0}$ (respectively, $\mathbf{U}_1$, $\mathbf{U}_{H,1}$, $\mathbf{V}_1$ and $\mathbf{V}_{H,1}$) are the corresponding singular vectors associated with the singular values greater than

τ (respectively, smaller than or equal to τ). Then we have $\mathbf{L}^H = \mathbf{U}_{H,0}(\mathbf{\Sigma}_{H,0} - \tau\mathbf{I})\mathbf{V}_{H,0}^\top\mathbf{R}^\top$, and therefore using $\mathbf{U}_0\mathbf{U}_0^\top + \mathbf{U}_1\mathbf{U}_1^\top = \mathbf{I}$ and $\mathbf{U}_0^\top\mathbf{U}_0 = \mathbf{I}$ we get

$$\begin{aligned}
&\mathbf{M} - \mathbf{L}^H \\
&= \mathbf{U}_0\mathbf{\Sigma}_0\mathbf{V}_0^\top + \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^\top - \mathbf{U}_{H,0}(\mathbf{\Sigma}_{H,0} - \tau\mathbf{I})\mathbf{V}_{H,0}^\top\mathbf{R}^\top \\
&= \mathbf{U}_0\mathbf{\Sigma}_0\mathbf{V}_0^\top + \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^\top - (\mathbf{U}_0\mathbf{U}_0^\top + \mathbf{U}_1\mathbf{U}_1^\top)\mathbf{U}_{H,0}(\mathbf{\Sigma}_{H,0} - \tau\mathbf{I})\mathbf{V}_{H,0}^\top\mathbf{R}^\top \\
&= \tau[\tau^{-1}\mathbf{U}_1(\mathbf{\Sigma}_1\mathbf{V}_1^\top - \mathbf{U}_1^\top\mathbf{U}_{H,0}(\mathbf{\Sigma}_{H,0} - \tau\mathbf{I})\mathbf{V}_{H,0}^\top\mathbf{R}^\top) \\
&\quad + \tau^{-1}\mathbf{U}_0(\mathbf{\Sigma}_0\mathbf{V}_0^\top - \mathbf{U}_0^\top\mathbf{U}_{H,0}(\mathbf{\Sigma}_{H,0} - \tau\mathbf{I})\mathbf{V}_{H,0}^\top\mathbf{R}^\top)] \\
&:= \tau[\mathbf{W} + \mathbf{U}\mathbf{V}^\top],
\end{aligned} \tag{8}$$

where

$$\mathbf{W} := \tau^{-1}\mathbf{U}_1(\mathbf{\Sigma}_1\mathbf{V}_1^\top - \mathbf{U}_1^\top\mathbf{U}_{H,0}(\mathbf{\Sigma}_{H,0} - \tau\mathbf{I})\mathbf{V}_{H,0}^\top\mathbf{R}^\top), \tag{9}$$

$$\mathbf{V}^\top := \tau^{-1}(\mathbf{\Sigma}_0\mathbf{V}_0^\top - \mathbf{U}_0^\top\mathbf{U}_{H,0}(\mathbf{\Sigma}_{H,0} - \tau\mathbf{I})\mathbf{V}_{H,0}^\top\mathbf{R}^\top) \tag{10}$$

and $\mathbf{U} := \mathbf{U}_0$. Then $\mathbf{U}^\top\mathbf{W} = \mathbf{0}$ and since $\|\mathbf{\Sigma}_1\|_2 \leq \tau$ and $\sigma_{H,1} \geq \tau$ we also have

$$\begin{aligned}
&\|\mathbf{W}\|_2 \\
&= \tau^{-1}\|\mathbf{\Sigma}_1\mathbf{V}_1^\top - \mathbf{U}_1^\top\mathbf{U}_{H,0}(\mathbf{\Sigma}_{H,0} - \tau\mathbf{I})\mathbf{V}_{H,0}^\top\mathbf{R}^\top\|_2 \\
&\leq \tau^{-1}(\|\mathbf{\Sigma}_1\|_2 + \|\mathbf{\Sigma}_{H,0} - \tau\mathbf{I}\|_2\|\mathbf{R}^\top\|_2) \\
&\leq \tau^{-1}(\tau + \sigma_{H,1} - \tau) \\
&= \frac{\sigma_{H,1}}{\tau}
\end{aligned} \tag{11}$$

Furthermore,

$$\begin{aligned}
&\|\mathbf{W}\mathbf{V}\|_2 \\
&= \tau^{-2}\|\mathbf{U}_1(\mathbf{\Sigma}_1\mathbf{V}_1^\top - \mathbf{U}_1^\top\mathbf{U}_{H,0}(\mathbf{\Sigma}_{H,0} - \tau\mathbf{I})\mathbf{V}_{H,0}^\top\mathbf{R}^\top) \\
&\qquad\qquad (\mathbf{\Sigma}_0\mathbf{V}_0^\top - \mathbf{U}_0^\top\mathbf{U}_{H,0}(\mathbf{\Sigma}_{H,0} - \tau\mathbf{I})\mathbf{V}_{H,0}^\top\mathbf{R}^\top)^\top\|_2 \\
&\leq \tau^{-2}(\|\mathbf{\Sigma}_1\|_2 + \|\mathbf{\Sigma}_{H,0} - \tau\mathbf{I}\|_2)(\|\mathbf{\Sigma}_0\|_2 + \|\mathbf{\Sigma}_{H,0} - \tau\mathbf{I}\|_2) \\
&\leq \tau^{-2}(\tau + \sigma_{H,1} - \tau)(\sigma_1 + \sigma_{H,1} - \tau) \\
&= \frac{\sigma_{H,1}}{\tau^2}(\sigma_1 + \sigma_{H,1} - \tau),
\end{aligned} \tag{12}$$

where $\sigma_{H,1} = \|\mathbf{\Sigma}_{H,0}\|_2$ is the largest singular value of $\mathbf{MR}$ and for the last inequality we used the assumptions that $\sigma_{H,1} \geq \tau$ and $\|\mathbf{R}\|_2 \leq 1$. Therefore, since $\frac{\sigma_{H,1}}{\tau^2}(\sigma_1 + \sigma_{H,1} - \tau) \geq \frac{\sigma_{H,1}}{\tau}$ then $\mathcal{D}_\tau^H[\mathbf{M}]$ is at most $\frac{\sigma_{H,1}}{\tau^2}(\sigma_1 + \sigma_{H,1} - \tau)$ away from a zero subdifferential of (1). $\square$

A.2 ML-AltProj

In this section we give the proof of Theorem 2. Here the optimisation problem in question is given as

$$\min_{\mathbf{L}\in\mathbb{R}^{m\times n}} \|\mathbf{D}-\mathbf{L}-\mathbf{S}\|_2 \quad s.t. \quad \text{rank}(\mathbf{L}) \leq l. \tag{13}$$

Then we solve (13) using the multilevel hard thresholding operator defined as

$$\mathbf{L}^H = \mathbf{U}_H \mathcal{H}_k[\mathbf{\Sigma}_H]\mathbf{V}_H^\top \mathbf{R}^\top, \tag{14}$$

where $\mathbf{MR} = \mathbf{M}_H = \mathbf{U}_H\mathbf{\Sigma}_H\mathbf{V}_H^\top$ is a SVD of the coarse model, $\mathcal{H}$ is the hard thresholding operator and $\mathbf{R}$ is the restirction operator.

Theorem 2. *The multilevel low rank approximation procedure given in (14) gives a $(\sigma_1+\sigma_{H,1})$-approximate solution to the problem (13), where $\sigma_{H,1}$ is the largest singular value of $\mathbf{M}_H = \mathbf{MR}$.*

Proof. First note that for any $\mathbf{B} = \mathbf{X}\mathbf{Y}^\top$ $(\mathbf{X},\mathbf{Y}\in\mathbb{R}^{m\times k})$ we can find a vector $\omega = \sum_{i=1}^{k+1}\gamma_i \mathbf{v}_i$ so that $\omega^\top\mathbf{X} = \mathbf{0}$ and $\|\omega\|_2 = \sum_{i=1}^{k+1}\gamma_i^2 = 1$, where $\mathbf{v}_i$ are the columns of $\mathbf{V}$. Then

$$\|\mathbf{M}-\mathbf{B}\|_2^2 \geq \|\omega^\top(\mathbf{M}-\mathbf{B})\|_2^2 = \|\omega^\top\mathbf{M}\|_2^2 = \sum_{i=1}^{k+1}\gamma_i^2\sigma_i^2 \geq \sigma_{k+1}^2. \tag{15}$$

On the other hand

$$\begin{aligned} &\|\mathbf{M}-\mathbf{L}^H\|_2 \\ &= \|\mathbf{U}_0\mathbf{\Sigma}_0\mathbf{V}_0^\top + \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^\top - \mathbf{U}_{H,0}\mathbf{\Sigma}_{H,0}\mathbf{V}_{H,0}^\top\mathbf{R}^\top\|_2 \\ &\leq \sigma_1 + \sigma_{k+1} + \sigma_{H,1}, \end{aligned} \tag{16}$$

where $\mathbf{U}_0$, $\mathbf{V}_0$, $\mathbf{U}_{H,0}$ and $\mathbf{V}_{H,0}$ (respectively, $\mathbf{U}_1$, $\mathbf{V}_1$, $\mathbf{U}_{H,1}$ and $\mathbf{V}_{H,1}$) are correspondingly the singular vectors associated with the largest k (respectively, smallest $n-k$ and n_H-k) singular values of $\mathbf{M}$ and $\mathbf{MR}$. From these two inequalities we have that for any $\mathbf{B}$

$$\|\mathbf{M}-\mathbf{L}^H\|_2 - \|\mathbf{M}-\mathbf{B}\|_2 \leq \sigma_1 + \sigma_{H,1}. \tag{17}$$

□

References

[1] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[2] G. A. Watson. Characterization of the subdifferential of some matrix norms. *Linear algebra and its applications*, 170:33–45, 1992.