Sparse Additive Subspace Clustering (Supplement)

Xiao-Tong Yuan[1,2], Ping Li[2]

1. S-mart Lab, Nanjing Uinversity of Information Science and Technology
Nanjing, 210044, China
2. Department of Statistics and Biostatistics, Department of Computer Science, Rutgers University
Piscataway, New Jersey, 08854, USA
Email: xtyuan@nuist.edu.cn, pingli@stat.rutgers.edu

1 Proof of Technical Lemmas

This section is devoted to proving several technical lemmas used in the paper.

1.1 Proof of Lemma 1

We need the following proposition which is almost standard in compressive sensing.

Proposition 1. *Let $T \subseteq G$ be a subset of the groups. If the optimal solution*

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} \frac{1}{2p}\|\boldsymbol{z} - \boldsymbol{\Psi}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_{G,2} \quad \text{subject to } \boldsymbol{\beta}_{T^c} = \mathbf{0}$$

obeying $\|\frac{1}{p}\boldsymbol{\Psi}_{T^c}^\top(\boldsymbol{z} - \boldsymbol{\Psi}\boldsymbol{\beta}^*)\|_{T^c,\infty} < \lambda$, then any optimal solution

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \frac{1}{2p}\|\boldsymbol{z} - \boldsymbol{\Psi}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_{G,2}$$

must also satisfy $\hat{\boldsymbol{\beta}}_{T^c} = \mathbf{0}$.

Proof. Let us express $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^* + \boldsymbol{h}$. Some algebraic manipulations show that

$$\begin{aligned}
& \frac{1}{2p}\|\boldsymbol{z} - \boldsymbol{\Psi}(\boldsymbol{\beta}^* + \boldsymbol{h})\|^2 + \lambda\|\boldsymbol{\beta}^* + \boldsymbol{h}\|_{G,2} \\
\geq & \frac{1}{2p}\|\boldsymbol{z} - \boldsymbol{\Psi}\boldsymbol{\beta}^*\|^2 - \langle \frac{1}{p}\boldsymbol{\Psi}^\top(\boldsymbol{z} - \boldsymbol{\Psi}\boldsymbol{\beta}^*), \boldsymbol{h}\rangle + \frac{1}{2p}\|\boldsymbol{\Psi}\boldsymbol{h}\|^2 \\
& + \lambda\sum_{g\in G}\left(\|\boldsymbol{\beta}^*_g\| + \langle \mathrm{sgn}(\boldsymbol{\beta}^*_g), \boldsymbol{h}_g\rangle\right) + \lambda\sum_{g\in T^c}\|\boldsymbol{h}_g\| \\
= & \frac{1}{2p}\|\boldsymbol{z} - \boldsymbol{\Psi}\boldsymbol{\beta}^*\|^2 + \lambda\|\boldsymbol{\beta}^*\|_{G,2} + \frac{1}{2p}\|\boldsymbol{\Psi}\boldsymbol{h}\|^2 + \lambda\sum_{g\in T^c}(\|\boldsymbol{h}_g\| - \langle \frac{1}{p\lambda}\boldsymbol{\Psi}_g^\top(\boldsymbol{z} - \boldsymbol{\Psi}\boldsymbol{\beta}^*), \boldsymbol{h}_g\rangle),
\end{aligned}$$

where the "$\geq$" follows from the simple inequality $\|\boldsymbol{a}+\boldsymbol{b}\| \geq \|\boldsymbol{a}\| + \langle \boldsymbol{a}, \boldsymbol{b}\rangle / \|\boldsymbol{a}\|$, the "=" follows from the optimality conditions showing that $\forall g \in T$, $\frac{1}{p}\boldsymbol{\Psi}_g^\top(\boldsymbol{z} - \boldsymbol{\Psi}\boldsymbol{\beta}^*) = \lambda \mathrm{sgn}(\boldsymbol{\beta}_g^*)$. Assume that $\exists g \in T^c, \boldsymbol{h}_g \neq \mathbf{0}$. Since by assumption $\forall g \in T$, $\|\frac{1}{p}\boldsymbol{\Psi}_g^\top(\boldsymbol{z} - \boldsymbol{\Psi}\boldsymbol{\beta}^*)\| < \lambda$, from the preceding inequality we have

$$\frac{1}{2p}\|\boldsymbol{z} - \boldsymbol{\Psi}\hat{\boldsymbol{\beta}}\|^2 + \lambda\|\hat{\boldsymbol{\beta}}\|_{G,2} > \frac{1}{2p}\|\boldsymbol{z} - \boldsymbol{\Psi}\boldsymbol{\beta}^*\|^2 + \lambda\|\boldsymbol{\beta}^*\|_{G,2},$$

which contradicts the optimality of $\hat{\boldsymbol{\beta}}$. Therefore, we have $\boldsymbol{h}_{T^c} = \mathbf{0}$, which proves the claim. □

Proof (of Lemma 1). **Part (a)**: Let us consider a solution $\boldsymbol{\beta}^*$ to

$$\min_{\boldsymbol{\beta}} \frac{1}{2p}\|\boldsymbol{z} - \boldsymbol{\Psi}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_{G,2} \quad \text{subject to } \boldsymbol{\beta}_{T^c} = \mathbf{0}.$$

The optimality conditions show that

$$\frac{1}{p}\boldsymbol{\Psi}_T^\top(\boldsymbol{z} - \boldsymbol{\Psi}_T\boldsymbol{\beta}_T^*) = \lambda \mathrm{sgn}(\boldsymbol{\beta}_T^*),$$

which implies

$$\boldsymbol{\beta}_T^* = (\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}(\boldsymbol{\Psi}_T^\top\boldsymbol{z} - \lambda p \mathrm{sgn}(\boldsymbol{\beta}_T^*)). \tag{1}$$

The remaining task is to check that $\|\frac{1}{p}\boldsymbol{\Psi}_{T^c}^\top(\boldsymbol{z} - \boldsymbol{\Psi}\boldsymbol{\beta}^*)\|_{T^c,\infty} < \lambda$ for the selection of λ in the lemma and then apply Proposition 1. Indeed, $\forall g \in T^c$,

$$\begin{aligned}
&\|\frac{1}{p}\boldsymbol{\Psi}_g^\top(\boldsymbol{z} - \boldsymbol{\Psi}\boldsymbol{\beta}^*)\| = \|\frac{1}{p}\boldsymbol{\Psi}_g^\top(\boldsymbol{z} - \boldsymbol{\Psi}_T\boldsymbol{\beta}_T^*)\| \\
&\leq \frac{1}{p}\|\boldsymbol{\Psi}_g^\top\boldsymbol{z} - \boldsymbol{\Psi}_g^\top\boldsymbol{\Psi}_T(\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top\boldsymbol{z}\| + \lambda\|\boldsymbol{\Psi}_g^\top\boldsymbol{\Psi}_T(\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}\mathrm{sgn}(\boldsymbol{\beta}_T^*)\| \\
&< \delta\lambda + (1-\delta)\lambda = \lambda,
\end{aligned}$$

where the "$\leq$" uses (1) and triangle inequality, and the "$<$" follows the condition on λ. This proves part (a).

Part (b): The optimality condition of $\hat{\boldsymbol{\beta}}$ is

$$\frac{1}{p}\boldsymbol{\Psi}_T^\top(\boldsymbol{z} - \boldsymbol{\Psi}_T\hat{\boldsymbol{\beta}}_T) = \lambda \mathrm{sgn}(\hat{\boldsymbol{\beta}}_T).$$

After some algebraic manipulations we obtain

$$\begin{aligned}
\|\hat{\boldsymbol{\beta}}_T - \bar{\boldsymbol{\beta}}_T\|_\infty &= \|(\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}(\boldsymbol{\Psi}_T^\top(\boldsymbol{z} - \boldsymbol{\Psi}\bar{\boldsymbol{\beta}}) - \lambda p \mathrm{sgn}(\hat{\boldsymbol{\beta}}_T))\|_\infty \\
&\leq \|(\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}\|_{T,\infty}\|\boldsymbol{\Psi}_T^\top(\boldsymbol{z} - \boldsymbol{\Psi}\bar{\boldsymbol{\beta}}) - \lambda p \mathrm{sgn}(\hat{\boldsymbol{\beta}}_T)\|_{T,\infty} \\
&\leq \left\|\left(\frac{1}{p}\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T\right)^{-1}\right\|_{T,\infty}\left(\left\|\frac{1}{p}\boldsymbol{\Psi}_T^\top(\boldsymbol{z} - \boldsymbol{\Psi}\bar{\boldsymbol{\beta}})\right\|_{T,\infty} + \lambda\right).
\end{aligned} \tag{2}$$

This completes the proof. □

1.2 Proof of Lemma 2

We need the following proposition which gives a concentration bound on the element-wise infinity norm of a random matrix.

Proposition 2. *Let $\boldsymbol{A} = (a_{ij}) \in \mathbb{R}^{m\times n}$ is a random matrices whose entries has variance no larger than σ^2. Then we have*

$$\|A - \mathbb{E}[A]\|_{\infty,\infty} \leq \sigma\sqrt{\frac{mn}{\eta}}$$

holds with probability at least $1-\eta$.

Proof. *From Chebyshev's inequality we get that*

$$\mathbb{P}\left(|a_{ij} - \mathbb{E}(a_{ij})| \geq \sqrt{mn}\sigma/\sqrt{\eta}\right) \leq \frac{\eta}{mn}.$$

By union of bound we get

$$\mathbb{P}\left(\max_{i,j}|a_{ij} - \mathbb{E}(a_{ij})| \geq \sqrt{mn}\sigma/\sqrt{\eta}\right) \leq \eta.$$

This proves the claim.

Proof (of Lemma 2). Since $\|\frac{1}{p}(\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}\|_{T,\infty} \leq l$ holds with high probability and the elements of $\boldsymbol{\Psi}$ are bounded, we have that the entries of $(\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_{T^c}$ have variance $o(1/p)$. From Proposition 2 we get that

$$\|(\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_{T^c} - \mathbb{E}[(\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_{T^c}]\|_{\infty,\infty} = o(\frac{n}{\sqrt{p\eta}}).$$

holds with probability at least $1-\eta$. Therefore, with high probability

$$\begin{aligned}
&\|(\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_{T^c}\|_{T^c,\infty} \\
&\leq \|\mathbb{E}[(\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_{T^c}]\|_{T^c,\infty} + \|\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_{T^c} - \mathbb{E}[\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_{T^c}]\|_{T^c,\infty} \\
&\leq 1 - 2\delta + q^{1/2}n\|\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_{T^c} - \mathbb{E}[\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_{T^c}]\|_{\infty,\infty} \\
&\leq 1 - 2\delta + o\left(\frac{n^2}{\sqrt{p\eta}}\right) \leq 1 - \delta, \qquad (3)
\end{aligned}$$

where the last inequality follows when p is sufficiently large. Next, we show that with high probability

$$\lambda > \frac{\|\boldsymbol{\Psi}_{T^c}^\top\boldsymbol{z} - \boldsymbol{\Psi}_{T^c}^\top\boldsymbol{\Psi}_T(\boldsymbol{\Psi}_T^\top\boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top\boldsymbol{z}\|_{T^c,\infty}}{p\delta}.$$

Indeed, since $\boldsymbol{z} = \boldsymbol{\Psi}_T \bar{\boldsymbol{\beta}}_T + \boldsymbol{\varepsilon}$, we have that

$$
\begin{aligned}
&\frac{1}{p}\|\boldsymbol{\Psi}_{T^c}^\top \boldsymbol{z} - \boldsymbol{\Psi}_{T^c}^\top \boldsymbol{\Psi}_T(\boldsymbol{\Psi}_T^\top \boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top \boldsymbol{z}\|_{T^c,\infty} \\
&= \frac{1}{p}\|\boldsymbol{\Psi}_{T^c}^\top \boldsymbol{\varepsilon} - \boldsymbol{\Psi}_{T^c}^\top \boldsymbol{\Psi}_T(\boldsymbol{\Psi}_T^\top \boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top \boldsymbol{\varepsilon}\|_{T^c,\infty} \\
&= \frac{1}{p}\|\boldsymbol{\Psi}_{T^c}^\top (\boldsymbol{I} - \boldsymbol{\Psi}_T(\boldsymbol{\Psi}_T^\top \boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top)\boldsymbol{\varepsilon}\|_{T^c,\infty} \\
&\leq \sqrt{q}\left\|\frac{1}{p}\boldsymbol{\Psi}_{T^c}^\top (\boldsymbol{I} - \boldsymbol{\Psi}_T(\boldsymbol{\Psi}_T^\top \boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top)\boldsymbol{\varepsilon}\right\|_\infty .
\end{aligned}
$$

Since $\boldsymbol{I} - \boldsymbol{\Psi}_T(\boldsymbol{\Psi}_T^\top \boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top$ is a projection matrix, we get that the rows of the matrix $\boldsymbol{\Psi}_{T^c}^\top(\boldsymbol{I} - \boldsymbol{\Psi}_T(\boldsymbol{\Psi}_T^\top \boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top)$ lie within a bounded ball. Also, it is easy to see that $\boldsymbol{\varepsilon}$ and $\boldsymbol{\Psi}_{T^c}^\top(\boldsymbol{I} - \boldsymbol{\Psi}_T(\boldsymbol{\Psi}_T^\top \boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top)$ are uncorrelated, and thus $\mathbb{E}[\frac{1}{p}\boldsymbol{\Psi}_{T^c}^\top(\boldsymbol{I} - \boldsymbol{\Psi}_T(\boldsymbol{\Psi}_T^\top \boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top)\boldsymbol{\varepsilon}] = \mathbf{0}$. Therefore, from Proposition 2 we know that with probability at least $1 - \eta$,

$$
\frac{1}{p}\|\boldsymbol{\Psi}_{T^c}^\top \boldsymbol{z} - \boldsymbol{\Psi}_{T^c}^\top \boldsymbol{\Psi}_T(\boldsymbol{\Psi}_T^\top \boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top \boldsymbol{z}\|_{T^c,\infty} \leq \frac{c\sigma\sqrt{n}}{\sqrt{p\eta}}.
$$

Therefore, with probability at least $1 - \eta$,

$$
\lambda > \frac{\|\boldsymbol{\Psi}_{T^c}^\top \boldsymbol{z} - \boldsymbol{\Psi}_{T^c}^\top \boldsymbol{\Psi}_T(\boldsymbol{\Psi}_T^\top \boldsymbol{\Psi}_T)^{-1}\boldsymbol{\Psi}_T^\top \boldsymbol{z}\|_{T^c,\infty}}{p\delta}.
$$

The remaining follows Lemma 1. This proves the desired result.