Supplementary Material of
Correlation Adaptive Subspace Segmentation by Trace Lasso

Canyi Lu[1], Jiashi Feng[1], Zhouchen Lin[2,*], Shuicheng Yan[1]

[1] Department of Electrical and Computer Engineering, National University of Singapore;

[2] Key Laboratory of Machine Perception (MOE), School of EECS, Peking University

canyilu@gmail.com, a0066331@nus.edu.sg, zlin@pku.edu.cn, eleyans@nus.edu.sg

In this supplementary material, we prove the Theorem 1 which shows the grouping effect of CASS.

Theorem 1 *Given a data vector* $y \in \mathbb{R}^d$*, data points* $X = [x_1, \cdots, x_n] \in \mathbb{R}^{d\times n}$ *and parameter* $\lambda > 0$*. Let* $w^* = [w_1^*, \cdots, w_n^*]^T \in \mathbb{R}^n$ *be the optimal solution to the following problem*

$$\min_w f(w) = \frac{1}{2}||y - Xw||_2^2 + \lambda||X\mathit{Diag}(w)||_*. \tag{1}$$

If $x_i \to x_j$*, then* $w_i^* \to w_j^*$*.*

Theorem 1 says that if there are two columns x_i and x_j of X which are sufficiently close to each other, then the corresponding coefficients w_i^* and w_j^* are also sufficiently close to each other.

Suppose $X = [\hat{X}\ \tilde{X}]$, where $\tilde{X} \in \mathbb{R}^{d\times q}$ consists of q columns that are close to each other:

$$\max\{||\tilde{X} - \bar{x}_0\mathbf{1}^T||_*, ||\tilde{X} - \bar{x}_0\mathbf{1}^T||_2\} \leq \varepsilon, \tag{2}$$

where $\varepsilon > 0$, $\mathbf{1} \in \mathbb{R}^q$ is the all 1's vector, $\bar{x}_0$ is the mean of $\tilde{X}$, i.e. $\bar{x}_0 = \tilde{X}\mathbf{1}/q$, and $\hat{X} \in \mathbb{R}^{d\times(n-q)}$ consists of the rest columns of X. Accordingly $w^* = [\hat{w};\ \tilde{w}]$.

To prove Theorem 1, we only need to prove that if $||\tilde{w} - \bar{w}\mathbf{1}||_2$ is not small enough, then $f([\hat{w};\ \tilde{w}]) > f([\hat{w};\ \bar{w}\mathbf{1}])$, where $\bar{w} = \mathbf{1}^T\tilde{w}/q$ is the average of $\tilde{w}$.

We first prove two lemmas:

Lemma 1 $||A\mathit{Diag}(v)||_* \leq ||A||_F||v||_2$*, where* $v \in \mathbb{R}^N$*, and* $A \in \mathbb{R}^{D\times N}$*.*

Proof.

$$\begin{aligned}
||A\text{Diag}(v)||_* &= ||[A_1v_1\ A_2v_2\ \cdots\ A_Nv_N]||_* \\
&\leq \sum_{i=1}^N ||A_iv_i||_* \\
&= \sum_{i=1}^N ||A_i||_*|v_i| \\
&= \sum_{i=1}^N ||A_i||_2|v_i| \\
&\leq \sqrt{\sum_{i=1}^N ||A_i||_2^2 \sum_{i=1}^N |v_i|^2} \\
&= \sqrt{||A||_F^2||v||_2^2} \\
&= ||A||_F||v||_2.
\end{aligned} \tag{3}$$

*Corresponding author.

■

Lemma 2 *If $\lambda_i \geq \mu_i \geq 0$, $i = 1, \cdots, N$, and $C = \sum_{i=1}^{N}(\lambda_i - \mu_i)$, then $\sum_{i=1}^{N}\sqrt{\lambda_i} \geq \sum_{i=1}^{N}\sqrt{\mu_i} + \frac{C}{2\sqrt{\max\{\lambda_i\}}}$.*

Proof.

$$\begin{aligned}\sum_{i=1}^{N}\sqrt{\lambda_i} - \sum_{i=1}^{N}\sqrt{\mu_i} &= \sum_{i=1}^{N}(\sqrt{\lambda_i} - \sqrt{\mu_i}) \\ &= \sum_{i=1}^{N}\frac{\lambda_i - \mu_i}{\sqrt{\lambda_i} + \sqrt{\mu_i}} \\ &\geq \sum_{i=1}^{N}\frac{\lambda_i - \mu_i}{2\sqrt{\max\{\lambda_i\}}} \\ &= \frac{1}{2\sqrt{\max\{\lambda_i\}}}\sum_{i=1}^{N}(\lambda_i - \mu_i) \\ &= \frac{C}{2\sqrt{\max\{\lambda_i\}}}.\end{aligned} \tag{4}$$

■

Next we prove the following theorem which is equivalent to the Theorem 1:

Theorem 2 *For any $\varepsilon > 0$, if $||\tilde{w} - \bar{w}\mathbf{1}||_2 > \delta$, where*

$$\delta = \left(\frac{2((\lambda + ||y - \hat{X}\hat{w} - (\mathbf{1}^T\tilde{w})\bar{x}_0||_2)||\tilde{w}||_2 + \lambda|\bar{w}|)||[\hat{X}_{\hat{w}}\ \bar{w}\bar{x}_0\mathbf{1}^T]||_2}{\lambda||\bar{x}_0||_2^2} + 1\right)\varepsilon, \tag{5}$$

then $f([\hat{w};\ \tilde{w}]) > f([\hat{w};\ \bar{w}\mathbf{1}])$.

Proof.

$$\begin{aligned}f([\hat{w};\ \tilde{w}]) =& \frac{1}{2}||y - \hat{X}\hat{w} - \tilde{X}\tilde{w}||_2^2 + \lambda||[\hat{X}_{\hat{w}}\ \tilde{X}\mathrm{Diag}(\tilde{w})]||_* \\ =& \frac{1}{2}||(y - \hat{X}\hat{w} - \bar{x}_0\mathbf{1}^T\tilde{w}) + (\bar{x}_0\mathbf{1}^T - \tilde{X})\tilde{w}||_2^2 + \lambda||[\hat{X}_{\hat{w}}\ \ \bar{x}_0\mathbf{1}^T\mathrm{Diag}(\tilde{w})] + [0\ \ (\tilde{X} - \bar{x}_0\mathbf{1}^T)\mathrm{Diag}(\tilde{w})]||_* \\ \geq& \frac{1}{2}||y - \hat{X}\hat{w} - (\mathbf{1}^T\tilde{w})\bar{x}_0||_2^2 + \frac{1}{2}||(\bar{x}_0\mathbf{1}^T - \tilde{X})\tilde{w}||_2^2 - ||y - \hat{X}\hat{w} - (\mathbf{1}^T\tilde{w})\bar{x}_0||_2||(\bar{x}_0\mathbf{1}^T - \tilde{X})\tilde{w}||_2 \\ &+ \lambda||[\hat{X}_{\hat{w}}\ \ \bar{x}_0\mathbf{1}^T\mathrm{Diag}(\tilde{w})]||_* - \lambda||(\tilde{X} - \bar{x}_0\mathbf{1}^T)\mathrm{Diag}(\tilde{w})||_* \\ \geq& \frac{1}{2}||y - \hat{X}\hat{w} - (\mathbf{1}^T\tilde{w})\bar{x}_0||_2^2 - ||(y - \hat{X}\hat{w} - (\mathbf{1}^T\tilde{w})\bar{x}_0||_2||\tilde{w}||_2||\bar{x}_0\mathbf{1}^T - \tilde{X}||_2 \\ &+ \lambda||[\hat{X}_{\hat{w}}\ \bar{x}_0\tilde{w}^T]||_* - \lambda||\tilde{w}||_2||\tilde{X} - \bar{x}_0\mathbf{1}^T||_F \\ \geq& \frac{1}{2}||y - \hat{X}\hat{w} - q\bar{w}\bar{x}_0||_2^2 + \lambda||[\hat{X}_{\hat{w}}\ \bar{x}_0\tilde{w}^T]||_* - (\lambda + ||y - \hat{X}\hat{w} - (\mathbf{1}^T\tilde{w})\bar{x}_0||_2)||\tilde{w}||_2\varepsilon \\ =& \frac{1}{2}||y - \hat{X}\hat{w} - \tilde{X}(\bar{w}\mathbf{1})||_2^2 + \lambda||[\hat{X}_{\hat{w}}\ \bar{x}_0\tilde{w}^T]||_* - (\lambda + ||y - \hat{X}\hat{w} - (\mathbf{1}^T\tilde{w})\bar{x}_0||_2)||\tilde{w}||_2\varepsilon,\end{aligned} \tag{6}$$

where $\hat{X}_{\hat{w}} = \hat{X}\mathrm{Diag}(\hat{w})$. The last equation uses the fact that $q\bar{x}_0 = \tilde{X}\mathbf{1}$.

Denote $Y = \hat{X}_{\hat{w}}\hat{X}_{\hat{w}}^T$, and $\lambda_i(M)$, $i = 1, \cdots, d$, are the ordered eigenvalues of a matrix $M \in \mathbb{R}^{d\times d}$, i.e. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. We show that if $||\tilde{w} - \bar{w}\mathbf{1}||_2 > \delta$, then

$$||[\hat{X}_{\hat{w}}\ \ \bar{x}_0\tilde{w}^T]||_* > ||[\hat{X}_{\hat{w}}\ \bar{w}\bar{x}_0\mathbf{1}^T]||_* + \eta,\ \text{with}\ \eta > 0. \tag{7}$$

Indeed, since

$$\begin{aligned}
\sum_{i=1}^{d} \lambda_i(Y + ||\tilde{w}||_2^2 \bar{x}_0 \bar{x}_0^T) =& \text{tr}(Y + ||\tilde{w}||_2^2 \bar{x}_0 \bar{x}_0^T) \\
=& \text{tr}[(Y + ||\bar{w}\mathbf{1}||_2^2 \bar{x}_0 \bar{x}_0^T) + (||\tilde{w}||_2^2 - ||\bar{w}\mathbf{1}||_2^2)\bar{x}_0 \bar{x}_0^T] \\
=& \text{tr}(Y + ||\bar{w}\mathbf{1}||_2^2 \bar{x}_0 \bar{x}_0^T) + \text{tr}((||\tilde{w}||_2^2 - ||\bar{w}\mathbf{1}||_2^2)\bar{x}_0 \bar{x}_0^T) \\
=& \text{tr}(Y + ||\bar{w}\mathbf{1}||_2^2 \bar{x}_0 \bar{x}_0^T) + (||\tilde{w}||_2^2 - ||\bar{w}\mathbf{1}||_2^2)||\bar{x}_0||_2^2 \\
=& \sum_{i=1}^{d} \lambda_i(Y + ||\bar{w}\mathbf{1}||_2^2 \bar{x}_0 \bar{x}_0^T) + (||\tilde{w}||_2^2 - ||\bar{w}\mathbf{1}||_2^2)||\bar{x}_0||_2^2.
\end{aligned} \tag{8}$$

Note that $||\bar{w}\mathbf{1}||_2^2$ is the minimum value of $||\tilde{w}||_2^2$ under the constraint $\mathbf{1}^T\tilde{w} = q\bar{w}$. So $\lambda_i(Y + ||\tilde{w}||_2^2 \bar{x}_0 \bar{x}_0^T) \geq \lambda_i(Y + ||\bar{w}\mathbf{1}||_2^2 \bar{x}_0 \bar{x}_0^T) \geq 0$. Moreover, since $\mathbf{1}^T\tilde{w} = q\bar{w}$, we have $||\tilde{w}||_2^2 - ||\bar{w}\mathbf{1}||_2^2 = ||\tilde{w} - \bar{w}\mathbf{1}||_2^2$.

So by Lemma 2, we have

$$\begin{aligned}
||[\hat{X}_{\hat{w}} \ \bar{x}_0 \tilde{w}^T]||_* =& \sum_{i=1}^{d} \sqrt{\lambda_i(Y + ||\tilde{w}||_2^2 \bar{x}_0 \bar{x}_0^T)} \\
\geq& \sum_{i=1}^{d} \sqrt{\lambda_i(Y + ||\bar{w}\mathbf{1}||_2^2 \bar{x}_0 \bar{x}_0^T)} + \frac{||\tilde{w} - \bar{w}\mathbf{1}||_2^2 ||\bar{x}_0||_2^2}{2\sqrt{\lambda_1(Y + ||\bar{w}\mathbf{1}||_2^2 \bar{x}_0 \bar{x}_0^T)}} \\
=& ||[\hat{X}_{\hat{w}} \ \bar{w}\bar{x}_0 \mathbf{1}^T]||_* + \frac{||\tilde{w} - \bar{w}\mathbf{1}||_2^2 ||\bar{x}_0||_2^2}{2||[\hat{X}_{\hat{w}} \ \bar{w}\bar{x}_0 \mathbf{1}^T]||_2} \\
>& ||[\hat{X}_{\hat{w}} \ \bar{w}\bar{x}_0 \mathbf{1}^T]||_* + \frac{||\bar{x}_0||_2^2}{2||[\hat{X}_{\hat{w}} \ \bar{w}\bar{x}_0 \mathbf{1}^T]||_2}\delta.
\end{aligned} \tag{9}$$

Furthermore,

$$\begin{aligned}
||[\hat{X}_{\hat{w}} \ \bar{w}\bar{x}_0 \mathbf{1}^T]||_* =& ||[\hat{X}_{\hat{w}} \ \tilde{X}\text{Diag}(\bar{w}\mathbf{1})] + [0 \ \ \bar{w}\bar{x}_0 \mathbf{1}^T - \tilde{X}\text{Diag}(\bar{w}\mathbf{1})]||_* \\
\geq& ||[\hat{X}_{\hat{w}} \ \tilde{X}\text{Diag}(\bar{w}\mathbf{1})]||_* - ||\bar{w}\bar{x}_0 \mathbf{1}^T - \tilde{X}\text{Diag}(\bar{w}\mathbf{1})||_* \\
=& ||[\hat{X}_{\hat{w}} \ \tilde{X}\text{Diag}(\bar{w}\mathbf{1})]||_* - |\bar{w}|||\bar{x}_0 \mathbf{1}^T - \tilde{X}||_* \\
\geq& ||[\hat{X}_{\hat{w}} \ \tilde{X}\text{Diag}(\bar{w}\mathbf{1})]||_* - |\bar{w}|\varepsilon.
\end{aligned} \tag{10}$$

Combining Eqn (6)(9) and (10) together, we have

$$\begin{aligned}
f([\hat{w};\ \tilde{w}]) \geq& \frac{1}{2}||y - \hat{X}\hat{w} - \tilde{X}(\bar{w}\mathbf{1})||_2^2 + \lambda||[\hat{X}_{\hat{w}} \ \tilde{X}\text{Diag}(\bar{w}\mathbf{1})]||_* - \lambda|\bar{w}|\varepsilon + \frac{\lambda||\bar{x}_0||_2^2}{2||[\hat{X}_{\hat{w}} \ \bar{w}\bar{x}_0 \mathbf{1}^T]||_2}\delta \\
&- (\lambda + ||y - \hat{X}\hat{w} - \mathbf{1}^T\tilde{w}\bar{x}_0||_2)||\tilde{w}||_2\varepsilon \\
=& f([\hat{w};\ \bar{w}\mathbf{1}]) + \frac{\lambda||\bar{x}_0||_2^2}{2||[\hat{X}_{\hat{w}} \ \bar{w}\bar{x}_0 \mathbf{1}^T]||_2}\delta - ((\lambda + ||y - \hat{X}\hat{w} - (\mathbf{1}^T\tilde{w})\bar{x}_0||_2)||\tilde{w}||_2 + \lambda|\bar{w}|)\varepsilon.
\end{aligned} \tag{11}$$

Then by the choice of δ in Eqn (5), it is easy to see that

$$f([\hat{w};\ \tilde{w}]) > f([\hat{w};\ \bar{w}\mathbf{1}]). \tag{12}$$

Thus the Theorem 2 is proved. ■