Supplementary material to "Fused sparsity and robust estimation for linear models with unknown variance" submitted to NIPS 2012

This supplement contains the proofs of the theoretical results stated in the main paper.

A Proof of Theorem 2.1

Let us begin with some simple relations one can deduce from the definitions $\overline{\mathbf{M}} = [\mathbf{M}^\top\ \mathbf{N}^\top]^\top$, $\overline{\mathbf{M}}^{-1} = [\mathbf{M}_\dagger\ \mathbf{N}_\dagger]$:

$$\mathbf{M}_\dagger\mathbf{M} + \mathbf{N}_\dagger\mathbf{N} = \mathbf{I}_p,$$

$$\mathbf{M}\mathbf{M}_\dagger = \mathbf{I}_q, \quad \mathbf{N}\mathbf{N}_\dagger = \mathbf{I}_{n-q}, \quad \mathbf{M}\mathbf{N}_\dagger = 0, \quad \mathbf{N}\mathbf{M}_\dagger = 0.$$

We introduce the following vector:

$$\bar{\boldsymbol{\beta}} = \mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^* + \mathbf{N}_\dagger(\mathbf{N}_\dagger^\top\mathbf{X}^\top\mathbf{X}\mathbf{N}_\dagger)^{-1}\mathbf{N}_\dagger^\top\mathbf{X}^\top(\boldsymbol{Y} - \mathbf{X}\mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^*),$$

which satisfies

$$\mathbf{M}\bar{\boldsymbol{\beta}} = \mathbf{M}\boldsymbol{\beta}^*, \quad \mathbf{N}\bar{\boldsymbol{\beta}} = (\mathbf{N}_\dagger^\top\mathbf{X}^\top\mathbf{X}\mathbf{N}_\dagger)^{-1}\mathbf{N}_\dagger^\top\mathbf{X}^\top(\boldsymbol{Y} - \mathbf{X}\mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^*),$$

and

$$\begin{aligned}
\mathbf{X}\bar{\boldsymbol{\beta}} &= \mathbf{X}\mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^* + \boldsymbol{\Pi}(\boldsymbol{Y} - \mathbf{X}\mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^*)\\
&= \boldsymbol{\Pi}\boldsymbol{Y} + (\mathbf{I}_n - \boldsymbol{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^*\\
&= \boldsymbol{\Pi}\boldsymbol{Y} + (\mathbf{I}_n - \boldsymbol{\Pi})\mathbf{X}(\mathbf{I} - \mathbf{N}_\dagger\mathbf{N})\boldsymbol{\beta}^*\\
&= \boldsymbol{\Pi}\boldsymbol{Y} + (\mathbf{I}_n - \boldsymbol{\Pi})\mathbf{X}\boldsymbol{\beta}^*\\
&= \mathbf{X}\boldsymbol{\beta}^* + \sigma^*\boldsymbol{\Pi}\boldsymbol{\xi}.
\end{aligned} \tag{14}$$

The main point in the present proof is the following: if we set

$$\widetilde{\boldsymbol{\beta}} = \bigg(1 + \sigma^*\frac{\boldsymbol{\xi}^\top(\mathbf{I}_n - \boldsymbol{\Pi})\mathbf{X}\bar{\boldsymbol{\beta}}}{\|\mathbf{X}\bar{\boldsymbol{\beta}}\|_2^2}\bigg)\bar{\boldsymbol{\beta}},$$

then, with high probability, for some $\widetilde{\sigma} > 0$, the pair $(\widetilde{\boldsymbol{\beta}}, \widetilde{\sigma})$ is feasible (*i.e.*, satisfies the constraint of the optimization problem we are dealing with). In what follows, we will repeatedly use the following property: for $\boldsymbol{m} = \frac{\mathbf{X}\bar{\boldsymbol{\beta}}}{\|\mathbf{X}\bar{\boldsymbol{\beta}}\|_2}$ it holds that

$$\begin{aligned}
\boldsymbol{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}} &= \boldsymbol{Y} - \mathbf{X}\bar{\boldsymbol{\beta}} - \sigma^*\boldsymbol{m}\boldsymbol{m}^\top(\mathbf{I}_n - \boldsymbol{\Pi})\boldsymbol{\xi}\\
&= \sigma^*(\mathbf{I}_n - \boldsymbol{\Pi})\boldsymbol{\xi} - \sigma^*\boldsymbol{m}\boldsymbol{m}^\top(\mathbf{I}_n - \boldsymbol{\Pi})\boldsymbol{\xi}\\
&= \sigma^*(\mathbf{I}_n - \boldsymbol{m}\boldsymbol{m}^\top)(\mathbf{I}_n - \boldsymbol{\Pi})\boldsymbol{\xi}.
\end{aligned} \tag{15}$$

Most of subsequent arguments will be derived on an event $\mathcal{B}$, having probability close to one, which can be represented as $\mathcal{B} = A \cap B \cap C$, where:

$$A = \Big\{\|\mathbf{M}_\dagger^\top\mathbf{X}_1^\top(\mathbf{I}_n - \boldsymbol{m}\boldsymbol{m}^\top)(\mathbf{I}_n - \boldsymbol{\Pi})\boldsymbol{\xi}\|_\infty \le \sqrt{2n\log(q/\delta)}\Big\},$$

$$B = \Big\{\|(\mathbf{I}_n - \boldsymbol{\Pi})\boldsymbol{\xi}\|_2^2 \ge n - r - 2\sqrt{(n-r)\log(1/\delta)}\Big\},$$

$$C = \Big\{|\boldsymbol{m}^\top(\mathbf{I}_n - \boldsymbol{\Pi})\boldsymbol{\xi}| \le \sqrt{2\log(1/\delta)}\Big\},$$

for some $\delta \in (0,1)$ close to zero. For the convenience of the reader, we recall that $r = rank\{\boldsymbol{\Pi}\} = rank\{\mathbf{X}\mathbf{N}_\dagger(\mathbf{N}_\dagger^\top\mathbf{X}^\top\mathbf{X}\mathbf{N}_\dagger)^{-1}\mathbf{N}_\dagger^\top\mathbf{X}^\top\}$.

Step I: Evaluation of the probability of $\mathcal{B}$ Let us check that the conditions involved in the definition of $\mathcal{B}$ are satisfied with probability at least $1-5\delta$. Since all the diagonal entries of $\frac{1}{n}\mathbf{M}_\dagger^\top\mathbf{X}^\top\mathbf{X}\mathbf{M}_\dagger$ are equal to 1, we have $\|(\mathbf{X}\mathbf{M}_\dagger)_j\|_2^2 = n$ for all $j=1,...,q$. Then we have:

$$\begin{aligned}
\mathbb{P}(A^c) &= \mathbb{P}\Big(\|\mathbf{M}_\dagger^\top\mathbf{X}^\top(\mathbf{I}_n - \boldsymbol{m}\boldsymbol{m}^\top)(\mathbf{I}_n-\boldsymbol{\Pi})\boldsymbol{\xi}\|_\infty \geq \sqrt{2n\log(q/\delta)}\Big)\\
&\leq \sum_{j=1}^q \mathbb{P}\Big(|(\mathbf{X}\mathbf{M}_\dagger)_j^\top(\mathbf{I}_n - \boldsymbol{m}\boldsymbol{m}^\top)(\mathbf{I}_n-\boldsymbol{\Pi})\boldsymbol{\xi}| \geq \sqrt{2n\log(q/\delta)}\Big)\\
&= \sum_{j=1}^q \mathbb{P}\Big(|\eta|\|(\mathbf{X}\mathbf{M}_\dagger)_j^\top(\mathbf{I}_n - \boldsymbol{m}\boldsymbol{m}^\top)(\mathbf{I}_n-\boldsymbol{\Pi})\|_2 \geq \sqrt{2n\log(q/\delta)}\Big)
\end{aligned}$$

where $\eta\sim\mathcal{N}(0,1)$. Using the inequality $\|(\mathbf{X}\mathbf{M}_\dagger)_j^\top(\mathbf{I}_n - \boldsymbol{m}\boldsymbol{m}^\top)(\mathbf{I}_n-\boldsymbol{\Pi})\|_2 \leq \|(\mathbf{X}\mathbf{M}_\dagger)_j\|_2$ and the well known bound on the tails of the Gaussian distribution, we get

$$\begin{aligned}
\mathbb{P}(A^c) &\leq \sum_{j=1}^q \mathbb{P}\Big(|\eta|\|(\mathbf{X}\mathbf{M}_\dagger)_j^\top\|_2 \geq \sqrt{2n\log(q/\delta)}\Big)\\
&= q\,\mathbb{P}\Big(|\eta|\sqrt{n} \geq \sqrt{2n\log(q/\delta)}\Big)\\
&= 2q\,\mathbb{P}\Big(\eta \geq \sqrt{2\log(q/\delta)}\Big)\\
&\leq 2q\exp\{-\frac{1}{2}(\sqrt{2\log(q/\delta)})^2\} = 2\delta.
\end{aligned}$$

For the set B, we recall that $\boldsymbol{\xi}^\top(\mathbf{I}_n-\boldsymbol{\Pi})\boldsymbol{\xi}$ is a chi-squared random variable with $n-r$ degrees of freedom: $\boldsymbol{\xi}^\top(\mathbf{I}_n-\boldsymbol{\Pi})\boldsymbol{\xi}\sim\chi^2(n-r)$. Therefore:

$$\mathbb{P}(B^c) = \mathbb{P}(\chi^2(n-r) \leq n-r-2\sqrt{(n-r)\log(1/\delta)}) \leq e^{-\log(1/\delta)} = \delta$$

Finally, to bound the probability of C^c, we use that $\boldsymbol{m}^\top(\mathbf{I}_n-\boldsymbol{\Pi})\boldsymbol{\xi}\sim\|(\mathbf{I}_n-\boldsymbol{\Pi})\boldsymbol{m}\|_2\mathcal{N}(0,1)$. This yields:

$$\begin{aligned}
\mathbb{P}(C^c) &= \mathbb{P}(|\eta|\|(\mathbf{I}_n-\boldsymbol{\Pi})\boldsymbol{m}\|_2 \geq \sqrt{2\log(1/\delta)})\\
&\leq \mathbb{P}(|\eta|\|\boldsymbol{m}\|_2 \geq \sqrt{2\log(1/\delta)})\\
&\leq 2\mathbb{P}(\eta \geq \sqrt{2\log(1/\delta)}) = 2\delta.
\end{aligned}$$

Because of $\mathcal{B} = A\cap B\cap C$, we can conclude that:

$$\mathbb{P}(\mathcal{B}^c) \leq \mathbb{P}(A^c)+\mathbb{P}(B^c)+\mathbb{P}(C^c) \leq 5\delta$$

or, equivalently, $\mathbb{P}(\mathcal{B}) \geq 1-5\delta$.

Step II: feasibility of $\widetilde{\boldsymbol{\beta}}$ The goal here is to check that if λ and μ satisfy the condition:

$$\frac{\lambda^2}{\mu} \geq \frac{2n^2\log(q/\delta)}{n-r-2\sqrt{(n-r)\log(1/\delta)}-2\log(1/\delta)} \tag{16}$$

then, on the event $\mathcal{B}$, there exists $\widetilde{\sigma}\leq\sigma^*/\sqrt{\mu}$ such that the pair $(\widetilde{\boldsymbol{\beta}},\widetilde{\sigma})$ is feasible.

The matrix $\mathbf{I}_n-\boldsymbol{m}\boldsymbol{m}^\top$ is the orthogonal projector onto the $(n-1)$-dimensional subspace of $\mathbb{R}^n$ containing all the vectors orthogonal to $\mathbf{X}\widetilde{\boldsymbol{\beta}}$. Therefore, using (14), we arrive at

$$\begin{aligned}
\boldsymbol{Y}^\top(\boldsymbol{Y}-\mathbf{X}\widetilde{\boldsymbol{\beta}}) &= (\mathbf{X}\boldsymbol{\beta}^*)^\top(\boldsymbol{Y}-\mathbf{X}\widetilde{\boldsymbol{\beta}}) + \sigma^*\boldsymbol{\xi}^\top(\boldsymbol{Y}-\mathbf{X}\widetilde{\boldsymbol{\beta}})\\
&= \sigma^*\boldsymbol{\xi}^\top(\mathbf{I}_n-\boldsymbol{\Pi})(\mathbf{I}_n-\boldsymbol{m}\boldsymbol{m}^\top)\mathbf{X}\boldsymbol{\beta}^* + (\sigma^*)^2\boldsymbol{\xi}^\top(\mathbf{I}_n-\boldsymbol{m}\boldsymbol{m}^\top)(\mathbf{I}_n-\boldsymbol{\Pi})\boldsymbol{\xi}\\
&= (\sigma^*)^2\boldsymbol{\xi}^\top(\mathbf{I}_n-\boldsymbol{\Pi})(\mathbf{I}_n-\boldsymbol{m}\boldsymbol{m}^\top)\boldsymbol{\Pi}\boldsymbol{\xi} + (\sigma^*)^2\boldsymbol{\xi}^\top(\mathbf{I}_n-\boldsymbol{m}\boldsymbol{m}^\top)(\mathbf{I}_n-\boldsymbol{\Pi})\boldsymbol{\xi}\\
&= (\sigma^*)^2\boldsymbol{\xi}^\top(\mathbf{I}_n-\boldsymbol{\Pi})(\mathbf{I}_n-\boldsymbol{m}\boldsymbol{m}^\top)(\mathbf{I}_n-\boldsymbol{\Pi})\boldsymbol{\xi}\\
&= (\sigma^*)^2\|(\mathbf{I}_n-\boldsymbol{\Pi})\boldsymbol{\xi}\|_2^2 - (\sigma^*)^2[\boldsymbol{m}^\top(\mathbf{I}_n-\boldsymbol{\Pi})\boldsymbol{\xi}]^2.
\end{aligned}$$

On the event $\mathcal{B}$, we have:

$$\|(\mathbf{I}_n - \mathbf{\Pi})\boldsymbol{\xi}\|_2^2 \geq n - r - 2\sqrt{(n-r)\log(1/\delta)}, \quad [\boldsymbol{m}^\top(\mathbf{I}_n - \mathbf{\Pi})\boldsymbol{\xi}]^2 \leq 2\log(1/\delta).$$

So we know:

$$\boldsymbol{Y}^\top(\boldsymbol{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) \geq (\sigma^*)^2\big(n - r - 2\sqrt{(n-r)\log(1/\delta)} - 2\log(1/\delta)\big) \geq (\sigma^*)^2 n\mu$$

Setting $\widetilde{\sigma} = \sigma^*\big(n - r - 2\sqrt{(n-r)\log(1/\delta)} - 2\log(1/\delta)\big)^{1/2}(n\mu)^{-1/2}$ we get that the pair $(\widetilde{\boldsymbol{\beta}}, \widetilde{\sigma})$ satisfies the third constraint and that $\widetilde{\sigma} \leq \sigma^*/\sqrt{\mu}$. It is obvious that the second constraint is satisfied as well. To check the first constraint, we note that

$$\mathbf{M}_\dagger^\top\mathbf{X}^\top(\boldsymbol{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) = \sigma^*\mathbf{M}_\dagger^\top\mathbf{X}_1^\top(\mathbf{I}_n - \boldsymbol{m}\boldsymbol{m}^\top)(\mathbf{I}_n - \mathbf{\Pi})\boldsymbol{\xi},$$

and therefore

$$\|\mathbf{M}_\dagger^\top\mathbf{X}^\top(\boldsymbol{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}\|_\infty = \sigma^*\|\mathbf{M}_\dagger^\top\mathbf{X}_1^\top(\mathbf{I}_n - \boldsymbol{m}\boldsymbol{m}^\top)(\mathbf{I}_n - \mathbf{\Pi})\boldsymbol{\xi}\|_\infty \leq \sigma^*\sqrt{2n\log(q/\delta)}.$$

Under the condition stated in (16) above, the right-hand side of the last inequality is upper bounded by $\lambda\widetilde{\sigma}$. This completes the proof of the fact that the pair $(\widetilde{\boldsymbol{\beta}}, \widetilde{\sigma})$ is a feasible solution on the event $\mathcal{B}$.

Step III: proof of (7) and (8) On the event $\mathcal{B}$, the pair $(\widetilde{\boldsymbol{\beta}}, \widetilde{\sigma})$ is feasible and therefore $\|\mathbf{M}\widehat{\boldsymbol{\beta}}\|_1 \leq \|\mathbf{M}\widetilde{\boldsymbol{\beta}}\|_1$. Let $\boldsymbol{\Delta} = \mathbf{M}\widehat{\boldsymbol{\beta}} - \mathbf{M}\widetilde{\boldsymbol{\beta}}$ and J be the set of indexes corresponding to the nonzero elements of $\mathbf{M}\boldsymbol{\beta}^*$. We have $|J| \leq s$. Note that J is also the set of indexes corresponding to nonzero elements of $\mathbf{M}\widetilde{\boldsymbol{\beta}} \propto \mathbf{M}\bar{\boldsymbol{\beta}} = \mathbf{M}\boldsymbol{\beta}^*$. This entails that:

$$\begin{aligned}
\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\boldsymbol{\Delta}\|_2^2 &= \boldsymbol{\Delta}^\top\mathbf{M}_\dagger^\top\mathbf{X}^\top(\mathbf{I}_n - \mathbf{\Pi})^2\mathbf{X}\mathbf{M}_\dagger\boldsymbol{\Delta} \\
&= \boldsymbol{\Delta}^\top\mathbf{M}_\dagger^\top\mathbf{X}^\top(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\boldsymbol{\Delta} \\
&\leq \|\boldsymbol{\Delta}\|_1\|\mathbf{M}_\dagger^\top\mathbf{X}^\top(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\boldsymbol{\Delta}\|_\infty. \qquad (17)
\end{aligned}$$

Using the relations $\mathbf{M}_\dagger\mathbf{M} = \mathbf{I}_p - \mathbf{N}_\dagger\mathbf{N}$ and $(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{N}_\dagger = 0$ yields

$$\begin{aligned}
\|\mathbf{M}_\dagger^\top\mathbf{X}^\top(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\boldsymbol{\Delta}\|_\infty &= \|\mathbf{M}_\dagger^\top\mathbf{X}^\top(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}(\widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}})\|_\infty \\
&= \|\mathbf{M}_\dagger^\top\mathbf{X}^\top(\mathbf{I}_n - \mathbf{\Pi})(\mathbf{X}\widetilde{\boldsymbol{\beta}} - \mathbf{X}\widehat{\boldsymbol{\beta}})\|_\infty.
\end{aligned}$$

Taking into account the fact that both $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ satisfy the second constraint, we get $\mathbf{\Pi}(\mathbf{X}\widetilde{\boldsymbol{\beta}} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{\Pi}(\mathbf{X}\widetilde{\boldsymbol{\beta}} - Y) - \mathbf{\Pi}(\mathbf{X}\widehat{\boldsymbol{\beta}} - Y) = 0$. From the first constraint, we deduce:

$$\|\mathbf{M}_\dagger^\top\mathbf{X}^\top(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\boldsymbol{\Delta}\|_\infty = \|\mathbf{M}_\dagger^\top\mathbf{X}^\top(\mathbf{X}\widetilde{\boldsymbol{\beta}} - \mathbf{X}\widehat{\boldsymbol{\beta}})\|_\infty \leq \lambda(\widehat{\sigma} + \widetilde{\sigma}). \qquad (18)$$

To bound $\|\boldsymbol{\Delta}\|_1$, we use a standard argument from [4]:

$$\|\boldsymbol{\Delta}_{J^c}\|_1 = \|\mathbf{M}\widehat{\boldsymbol{\beta}}_{J^c}\|_1 = \|\mathbf{M}\widehat{\boldsymbol{\beta}}\|_1 - \|\mathbf{M}\widehat{\boldsymbol{\beta}}_J\|_1.$$

Since $\widetilde{\boldsymbol{\beta}}$ is a feasible solution while $\widehat{\boldsymbol{\beta}}$ is an optimal one, $\|\mathbf{M}\widehat{\boldsymbol{\beta}}\|_1 \leq \|\mathbf{M}\widetilde{\boldsymbol{\beta}}\|_1$, and we have:

$$\|\boldsymbol{\Delta}_{J^c}\|_1 \leq \|\mathbf{M}\widetilde{\boldsymbol{\beta}}\|_1 - \|\mathbf{M}\widehat{\boldsymbol{\beta}}_J\|_1 = \|\mathbf{M}\widetilde{\boldsymbol{\beta}}_J\|_1 - \|\mathbf{M}\widehat{\boldsymbol{\beta}}_J\|_1 \leq \|(\mathbf{M}\widetilde{\boldsymbol{\beta}} - \mathbf{M}\widehat{\boldsymbol{\beta}})_J\|_1 = \|\boldsymbol{\Delta}_J\|_1.$$

This yields the bound

$$\|\boldsymbol{\Delta}\|_1 \leq 2\|\boldsymbol{\Delta}_J\|_1 \leq 2s^{1/2}\|\boldsymbol{\Delta}_J\|_2$$

and also allows us to use the condition of RE$(s, 1)$, which implies that:

$$\|\boldsymbol{\Delta}_J\|_2 \leq \frac{\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\boldsymbol{\Delta}\|_2}{\kappa\sqrt{n}}. \qquad (19)$$

Combining these estimates, we get

$$\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\boldsymbol{\Delta}\|_2^2 \leq \frac{2\lambda(\widehat{\sigma} + \sigma^*)\sqrt{s}\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\boldsymbol{\Delta}\|_2}{\kappa\sqrt{n}},$$

and, after simplification

$$\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger \mathbf{\Delta}\|_2 \le 2\lambda(\widehat{\sigma} + \sigma^*)\frac{\sqrt{s}}{\kappa\sqrt{n}}, \qquad \|\mathbf{\Delta}_J\|_2 \le 2\lambda(\widehat{\sigma} + \sigma^*)\frac{\sqrt{s}}{n\kappa^2}. \tag{20}$$

Furthermore:

$$\|\mathbf{\Delta}\|_1 = \|\mathbf{\Delta}_J\|_1 + \|\mathbf{\Delta}_{J^c}\|_1 \le 2\|\mathbf{\Delta}\|_1 \le 2\sqrt{s}\|\mathbf{\Delta}\|_2 \le 4\lambda(\widehat{\sigma} + \sigma^*)\frac{s}{n\kappa^2}$$

So we have:

$$\|\mathbf{M}\widehat{\boldsymbol{\beta}} - \mathbf{M}\widetilde{\boldsymbol{\beta}}\|_1 \le 4\lambda(\widehat{\sigma} + \sigma^*)\frac{s}{n\kappa^2}, \quad \|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})\|_2 \le 2\lambda(\widehat{\sigma} + \sigma^*)\frac{\sqrt{s}}{\kappa\sqrt{n}}$$

To complete this step, we decompose $\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}$ into the sum of the terms $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$ and $\boldsymbol{\beta}^* - \widetilde{\boldsymbol{\beta}}$ and estimate the latter in prediction norm and in ℓ_1-norm. For the ℓ_1-norm, this gives

$$\begin{aligned}
\|\mathbf{M}\widetilde{\boldsymbol{\beta}} - \mathbf{M}\boldsymbol{\beta}^*\|_1 &= \sigma^*\Big\|\frac{\boldsymbol{\xi}^\top(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\bar{\boldsymbol{\beta}}}{\|\mathbf{X}\bar{\boldsymbol{\beta}}\|_2^2}\mathbf{M}\boldsymbol{\beta}^*\Big\|_1 = \sigma^*|\boldsymbol{m}^\top(\mathbf{I}_n - \mathbf{\Pi})\boldsymbol{\xi}|\frac{\|\mathbf{M}\boldsymbol{\beta}^*\|_1}{\|\mathbf{X}\bar{\boldsymbol{\beta}}\|_2} \\
&\le \sigma^*\sqrt{2\log(1/\delta)}\frac{\|\mathbf{M}\boldsymbol{\beta}^*\|_1}{\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\bar{\boldsymbol{\beta}}\|_2} = \sigma^*\sqrt{2\log(1/\delta)}\frac{\|\mathbf{M}\boldsymbol{\beta}^*\|_1}{\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^*\|_2} \\
&\le \sigma^*\sqrt{2\log(1/\delta)}\frac{\sqrt{s}\|\mathbf{M}\boldsymbol{\beta}^*\|_2}{\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^*\|_2} \le \sigma^*\sqrt{2\log(1/\delta)}\frac{\sqrt{s}}{\kappa\sqrt{n}} \\
&= \sqrt{2\log(1/\delta)}\frac{\sigma^*\sqrt{s}}{\kappa\sqrt{n}}.
\end{aligned}$$

While for the prediction norm:

$$\begin{aligned}
\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|_2 &= \|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|_2 \\
&= \sigma^*\Big\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\frac{\boldsymbol{\xi}^\top(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\bar{\boldsymbol{\beta}}}{\|\mathbf{X}\bar{\boldsymbol{\beta}}\|_2^2}\mathbf{M}\boldsymbol{\beta}^*\Big\|_2 \\
&= \sigma^*|\boldsymbol{m}^\top(\mathbf{I}_n - \mathbf{\Pi})\boldsymbol{\xi}|\frac{\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^*\|_2}{\|\mathbf{X}\bar{\boldsymbol{\beta}}\|_2} \\
&\le \sigma^*|\boldsymbol{m}^\top(\mathbf{I}_n - \mathbf{\Pi})\boldsymbol{\xi}|\frac{\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^*\|_2}{\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\bar{\boldsymbol{\beta}}\|_2} \\
&= \sigma^*|\boldsymbol{m}^\top(\mathbf{I}_n - \mathbf{\Pi})\boldsymbol{\xi}|\frac{\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^*\|_2}{\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^*\|_2} \\
&\le \sigma^*\sqrt{2\log(1/\delta)}.
\end{aligned}$$

We conclude that:

$$\|\mathbf{M}\widehat{\boldsymbol{\beta}} - \mathbf{M}\boldsymbol{\beta}^*\|_1 \le 4\lambda(\widehat{\sigma} + \sigma^*)\frac{s}{n\kappa^2} + \sqrt{2\log(1/\delta)}\sigma^*\frac{\sqrt{s}}{\kappa\sqrt{n}}$$

$$\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|_2 \le 2\lambda(\widehat{\sigma} + \sigma^*)\frac{\sqrt{s}}{\kappa\sqrt{n}} + \sigma^*\sqrt{2\log(1/\delta)}.$$

To finish, we remark that

$$\begin{aligned}
\|\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|_2 &\le \|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|_2 + \|\mathbf{\Pi}(\mathbf{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{Y})\|_2 + \sigma^*\|\mathbf{\Pi}\boldsymbol{\xi}\|_2 \\
&= \|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|_2 + \sigma^*\|\mathbf{\Pi}\boldsymbol{\xi}\|_2 \qquad \text{(in view of the second constraint)} \\
&\le 2\lambda(\widehat{\sigma} + \sigma^*)\frac{\sqrt{s}}{\kappa\sqrt{n}} + \sigma^*(\sqrt{2\log(1/\delta)} + r + \sqrt{2\log(1/\delta)}),
\end{aligned}$$

the last inequality being true with a probability at least $1 - 6\delta$.

Step IV: Proof of (9) Here we define J_0 a subset of $\{1, ..., q\}$ corresponding to the s largest value coordinates of $\mathbf{\Delta}$ outside of J, so $J_1 = J \cup J_0$. It is easy to see that the kth largest in absolute value element of $\mathbf{\Delta}_{J^c}$ satisfies $|\mathbf{\Delta}_{J^c}|_{(k)} \leq \|\mathbf{\Delta}_{J^c}\|_1/k$. Thus,

$$\|\mathbf{\Delta}_{J_1^c}\|_2^2 \leq \|\mathbf{\Delta}_{J^c}\|_1^2 \sum_{k\geq s+1} \frac{1}{k^2} \leq \frac{1}{s}\|\mathbf{\Delta}_{J^c}\|_1^2$$

On the event $\mathcal{B}$, with $c_0 = 1$ we get:

$$\|\mathbf{\Delta}_{J_1^c}\|_2 \leq \frac{\|\mathbf{\Delta}_J\|_1}{\sqrt{s}} \leq \sqrt{\frac{s}{s}}\|\mathbf{\Delta}_J\|_2 \leq \|\mathbf{\Delta}_{J_1}\|_2$$

Then, on $\mathcal{B}$,

$$\|\mathbf{\Delta}\|_2 \leq \|\mathbf{\Delta}_{J_1^c}\|_2 + \|\mathbf{\Delta}_{J_1}\|_2 \leq 2\|\mathbf{\Delta}_{J_1}\|_2$$

On the other hand, from (20),

$$\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{\Delta}\|_2^2 \leq 2\lambda(\widehat{\sigma} + \sigma^*)\sqrt{s}\|\mathbf{\Delta}_J\|_2 \leq 2\lambda(\widehat{\sigma} + \sigma^*)\sqrt{s}\|\mathbf{\Delta}_{J_1}\|_2$$

Combining this inequality with the Assumption RE$(s, s, 1)$,

$$\frac{\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{\Delta}\|_2}{\sqrt{n}\|\mathbf{\Delta}_{J_1}\|_2} \geq \kappa, \qquad \kappa\sqrt{n}\|\mathbf{\Delta}_{J_1}\|_2 \leq \|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{\Delta}\|_2$$

we obtain on $\mathcal{B}$,

$$\|\mathbf{\Delta}_{J_1}\|_2 \leq 2\frac{\widehat{\sigma} + \widetilde{\sigma}}{\kappa^2}\frac{\sqrt{s}\lambda}{n}$$

with the condition $\|\mathbf{\Delta}\|_2 \leq 2\|\mathbf{\Delta}_{J_1}\|_2$, we get:

$$\|\mathbf{M}\widehat{\boldsymbol{\beta}} - \mathbf{M}\widetilde{\boldsymbol{\beta}}\|_2 \leq 4\frac{\widehat{\sigma} + \widetilde{\sigma}}{\kappa^2}\frac{\sqrt{s}\lambda}{n}.$$

In addition, we have:

$$\begin{aligned}
\|\mathbf{M}\widetilde{\boldsymbol{\beta}} - \mathbf{M}\boldsymbol{\beta}^*\|_2 &= \sigma^*\left\|\frac{\boldsymbol{\xi}^\top(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\bar{\boldsymbol{\beta}}}{\|\mathbf{X}\bar{\boldsymbol{\beta}}\|_2^2}\mathbf{M}\boldsymbol{\beta}^*\right\|_2 = \sigma^*|\boldsymbol{m}^\top(\mathbf{I}_n - \mathbf{\Pi})\boldsymbol{\xi}|\frac{\|\mathbf{M}\boldsymbol{\beta}^*\|_2}{\|\mathbf{X}\bar{\boldsymbol{\beta}}\|_2} \\
&\leq \sigma^*\sqrt{2\log(1/\delta)}\frac{\|\mathbf{M}\boldsymbol{\beta}^*\|_2}{\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\bar{\boldsymbol{\beta}}\|_2} = \sigma^*\sqrt{2\log(1/\delta)}\frac{\|\mathbf{M}\boldsymbol{\beta}^*\|_2}{\|(\mathbf{I}_n - \mathbf{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^*\|_2} \\
&\leq \sigma^*\sqrt{2\log(1/\delta)}\frac{1}{\kappa\sqrt{n}} = \frac{\sigma^*}{\kappa\sqrt{n}}\sqrt{2\log(1/\delta)}.
\end{aligned}$$

Putting these estimates together and using the obvious inequality

$$\|\mathbf{M}\widehat{\boldsymbol{\beta}} - \mathbf{M}\boldsymbol{\beta}^*\|_2 \leq \|\mathbf{M}\widehat{\boldsymbol{\beta}} - \mathbf{M}\widetilde{\boldsymbol{\beta}}\|_2 + \|\mathbf{M}\widetilde{\boldsymbol{\beta}} - \mathbf{M}\boldsymbol{\beta}^*\|_2$$

we arrive at

$$\|\mathbf{M}\widehat{\boldsymbol{\beta}} - \mathbf{M}\boldsymbol{\beta}^*\|_2 \leq 4\frac{\widehat{\sigma} + \sigma^*}{\kappa^2}\frac{\sqrt{s}\lambda}{n} + \frac{\sigma^*}{\kappa}\sqrt{\frac{2\log(1/\delta)}{n}}.$$

Replacing $\lambda = \sqrt{2n\gamma\log(p/\delta)}$, we get the inequality in (9).

Step V: proof of an upper bound on $\widehat{\sigma}$ To complete the proof, one needs to check that $\widehat{\sigma}$ is of the order of σ^*. This is done by using the following chain of relations:

$$\begin{aligned}
n\mu\widehat{\sigma}^2 &\leq \|\boldsymbol{Y}\|_2^2 - \boldsymbol{Y}^\top\mathbf{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{Y}^\top(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) \\
&= (\boldsymbol{\beta}^*)^\top\mathbf{X}^\top(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) + \sigma^*\boldsymbol{\xi}^\top(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) \\
&= (\boldsymbol{\beta}^*)^\top\mathbf{M}^\top\mathbf{M}_\dagger^\top\mathbf{X}^\top(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) + (\boldsymbol{\beta}^*)^\top\mathbf{N}^\top\mathbf{N}_\dagger^\top\mathbf{X}^\top(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) + \sigma^*\boldsymbol{\xi}^\top(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}).
\end{aligned}$$

The second term of the last expression vanishes since $\widehat{\boldsymbol{\beta}}$ satisfies the second constraint. To bound the first term, we will use the first constraint while for bounding the third term, we will use the relation

$\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}} = (\mathbf{I}_n - \boldsymbol{\Pi})(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \sigma^*(\mathbf{I}_n - \boldsymbol{\Pi})\boldsymbol{\xi} + (\mathbf{I}_n - \boldsymbol{\Pi})\mathbf{X}(\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}) = \sigma^*(\mathbf{I}_n - \boldsymbol{\Pi})\boldsymbol{\xi} + (\mathbf{I}_n - \boldsymbol{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}(\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}})$. This leads to

$$n\mu\widehat{\sigma}^2 \le \widehat{\sigma}\lambda\|\mathbf{M}\boldsymbol{\beta}^*\|_1 + \sigma^*\boldsymbol{\xi}^\top(\mathbf{I}_n - \boldsymbol{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}(\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}) + (\sigma^*)^2\boldsymbol{\xi}^\top(\mathbf{I}_n - \boldsymbol{\Pi})\boldsymbol{\xi}$$

On the event $\mathcal{B}$, we have:

$$|\boldsymbol{\xi}^\top(\mathbf{I}_n - \boldsymbol{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}\widehat{\boldsymbol{\beta}}| \le \|\mathbf{M}\widehat{\boldsymbol{\beta}}\|_1\|\mathbf{M}_\dagger^\top\mathbf{X}^\top(\mathbf{I}_n - \boldsymbol{\Pi})\boldsymbol{\xi}\|_\infty \le \|\mathbf{M}\widehat{\boldsymbol{\beta}}\|_1\sqrt{2n\log(q/\delta)},$$
$$|\boldsymbol{\xi}^\top(\mathbf{I}_n - \boldsymbol{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^*| \le \|\mathbf{M}\boldsymbol{\beta}^*\|_1\|\mathbf{M}_\dagger^\top\mathbf{X}^\top(\mathbf{I}_n - \boldsymbol{\Pi})\boldsymbol{\xi}\|_\infty \le \|\mathbf{M}\boldsymbol{\beta}^*\|_1\sqrt{2n\log(1/\delta)}.$$

Also with a probability at least $1 - \delta$:

$$\boldsymbol{\xi}^\top(\mathbf{I}_n - \boldsymbol{\Pi})\boldsymbol{\xi} \le n - r + 2\sqrt{(n-r)\log(1/\delta)} + 2\log(1/\delta) \le (\sqrt{n-r} + \sqrt{2\log(1/\delta)})^2.$$

So combining all these relations, we get with probability at least $1 - 7\delta$:

$$\widehat{\sigma}^2 \le \widehat{\sigma}\frac{\lambda\|\mathbf{M}\boldsymbol{\beta}^*\|_1}{n\mu} + \frac{(\sigma^*)^2(\sqrt{n-r} + \sqrt{2\log(1/\delta)})^2}{n\mu} + (\|\mathbf{M}\widehat{\boldsymbol{\beta}}\|_1 + \|\mathbf{M}\boldsymbol{\beta}^*\|_1)\frac{\sigma^*}{\mu}\sqrt{\frac{2\log(q/\delta)}{n}}.$$

All the subsequent relations, even if it is not explicitly mentioned, are true on an event of probability at least $1 - 7\delta$. Combining simple algebra and the condition $RE(s)$, we get that:

$$\begin{aligned}
\|\mathbf{M}\widehat{\boldsymbol{\beta}}\|_1 \le \|\mathbf{M}\widetilde{\boldsymbol{\beta}}\|_1 &\le \|\mathbf{M}\boldsymbol{\beta}^*\|_1 + \sigma^*\frac{|\boldsymbol{\xi}^\top(\mathbf{I}_n - \boldsymbol{\Pi})\mathbf{X}\bar{\boldsymbol{\beta}}|}{\|\mathbf{X}\bar{\boldsymbol{\beta}}\|_2^2}\|\mathbf{M}\boldsymbol{\beta}^*\|_1 \\
&\le \|\mathbf{M}\boldsymbol{\beta}^*\|_1 + \sigma^*|\boldsymbol{m}^\top(\mathbf{I}_n - \boldsymbol{\Pi})\boldsymbol{\xi}|\frac{\|\mathbf{M}\boldsymbol{\beta}^*\|_1}{\|(\mathbf{I}_n - \boldsymbol{\Pi})\mathbf{X}\mathbf{M}_\dagger\mathbf{M}\boldsymbol{\beta}^*\|_2} \\
&\le \|\mathbf{M}\boldsymbol{\beta}^*\|_1 + \frac{\sigma^*}{\kappa}\sqrt{\frac{2s\log(1/\delta)}{n}}.
\end{aligned}$$

Then,

$$\begin{aligned}
\Big(\widehat{\sigma} - \frac{\lambda\|\mathbf{M}\boldsymbol{\beta}^*\|_1}{2n\mu}\Big)^2 \le \Big(\frac{\lambda\|\mathbf{M}\boldsymbol{\beta}^*\|_1}{2n\mu}\Big)^2 &+ \frac{(\sigma^*)^2(\sqrt{n} + \sqrt{2\log(1/\delta)})^2}{n\mu} \\
&+ \frac{2s^{1/2}(\sigma^*)^2\log(q/\delta)}{n\kappa\mu} + 2\|\mathbf{M}\boldsymbol{\beta}^*\|_1\frac{\sigma^*}{\mu}\sqrt{\frac{2\log(1/\delta)}{n}}
\end{aligned}$$

From the fact that $\sqrt{a^2 + b^2 + c} \le a + b + \frac{c}{2b}$, we have:

$$\widehat{\sigma} \le \frac{\lambda\|\mathbf{M}\boldsymbol{\beta}^*\|_1}{n\mu} + \frac{\sigma^*}{\sqrt{\mu}}\Big(1 + \sqrt{\frac{2\log(1/\delta)}{n}}\Big) + \frac{s^{1/2}\sigma^*\log(q/\delta)}{n\kappa\mu^{1/2}} + \|\mathbf{M}\boldsymbol{\beta}^*\|_1\sqrt{\frac{2\log(1/\delta)}{n\mu}}.$$

This yields the desired result.

B Proof of Theorem 3.1

All the claims of this theorem, except the bound on $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2$ are direct consequences of the corresponding claims in Theorem 2.1. Therefore, we focus here only on the proof of an upper bound on $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2$ taking all the other claims of Theorem 3.1 as granted.

Since $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\omega}}, \widehat{\sigma})$ is a feasible solution to (SRDS), it satisfies the second constraint:

$$\mathbf{A}^\top(\mathbf{A}\widehat{\boldsymbol{\theta}} + \sqrt{n}\,\widehat{\boldsymbol{\omega}} - \mathbf{A}\boldsymbol{\theta}^* - \sqrt{n}\,\boldsymbol{\omega}^* - \sigma^*\boldsymbol{\xi}) = 0,$$

which implies that

$$\|\mathbf{A}^\top\mathbf{A}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\|_2 \le \sqrt{n}\|\mathbf{A}^\top(\boldsymbol{\omega}^* - \widehat{\boldsymbol{\omega}})\|_2 + \sigma^*\|\mathbf{A}^\top\boldsymbol{\xi}\|_2.$$

Recall that ν_* stands for the smallest eigenvalue of $(\frac{1}{n}\mathbf{A}^\top\mathbf{A})^{1/2}$. This yields

$$\nu_*^2\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2 \le \frac{1}{n}\|\mathbf{A}^\top\mathbf{A}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\|_2 \le \frac{1}{\sqrt{n}}\|\mathbf{A}^\top(\boldsymbol{\omega}^* - \widehat{\boldsymbol{\omega}})\|_2 + \frac{\sigma^*}{n}\|\mathbf{A}^\top\boldsymbol{\xi}\|_2.$$

Since ν^* is the largest eigenvalue of $(\frac{1}{n}\mathbf{A}^\top\mathbf{A})^{1/2}$, we have

$$\frac{1}{\sqrt{n}}\|\mathbf{A}^\top(\boldsymbol{\omega}^* - \widehat{\boldsymbol{\omega}})\|_2 \le \nu^*\|\boldsymbol{\omega}^* - \widehat{\boldsymbol{\omega}}\|_2 \le \frac{4\nu^*(\widehat{\sigma}+\sigma^*)}{\kappa^2}\sqrt{\frac{2s\log(n/\delta)}{n}} + \frac{\nu^*\sigma^*}{\kappa}\sqrt{\frac{2\log(1/\delta)}{n}}.$$

To bound $\frac{\sigma^*}{n}\|\mathbf{A}^\top\boldsymbol{\xi}\|_2 = \|\frac{1}{\sqrt{n}}(\mathbf{A}^\top\mathbf{A})^{1/2}\boldsymbol{\xi}\|_2$ we denote by $\{\nu_i\}$ the eigenvalues of $\frac{1}{\sqrt{n}}(\mathbf{A}^\top\mathbf{A})^{1/2}$ and use the singular value decomposition of $\mathbf{A}^\top$:

$$\mathbf{A}^\top = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^\top$$

where $\mathbf{U}$ is a $k\times k$ orthogonal matrix, $\mathbf{V}$ is a $n\times n$ orthogonal matrix and $\boldsymbol{\Delta}$ is a $k\times n$ matrix with (assume $n>k$):

$$\boldsymbol{\Delta} = [\text{diag}\{\nu_1,\ldots,\nu_k\},\ \mathbf{0}_{k\times(n-k)}].$$

Setting $\boldsymbol{\eta} = \mathbf{V}^\top\boldsymbol{\xi}$, we get

$$\|\mathbf{A}^\top\boldsymbol{\xi}\|_2^2 = \|\mathbf{U}\boldsymbol{\Delta}\mathbf{V}^\top\boldsymbol{\xi}\|_2^2 = \|\boldsymbol{\Delta}\mathbf{V}^\top\boldsymbol{\xi}\|_2^2 = \|\boldsymbol{\Delta}\eta\|_2^2 \le \nu^*(\eta_1^2+...+\eta_k^2) \triangleq \nu^*\|\boldsymbol{\eta}_{1:k}\|_2^2.$$

Using the well-known inequality on the tails of chi-squared distribution:

$$\mathbb{P}(\|\boldsymbol{\eta}_{1:k}\|_2^2 \ge k + 2\sqrt{kx} + 2x) \le e^{-x}$$

with $x = \log(1/\delta)$, we obtain that with a probability at least $1-\delta$:

$$\|\boldsymbol{\eta}_{1:k}\|_2^2 \le k + 2\sqrt{k\log(1/\delta)} + 2\log(1/\delta) \le (\sqrt{k} + \sqrt{2\log(1/\delta)})^2.$$

Combined with the previous estimates, this leads to the desired result.

C Proof of Lemma 3.2

Let J be a subset of $\{1,\ldots,n\}$ of cardinality s and let $\boldsymbol{\delta}$ be a vector of $\mathbb{R}^n$ satisfying $\|\boldsymbol{\delta}_{J^c}\|_1 \le \|\boldsymbol{\delta}_J\|_1$. Let us denote by $\boldsymbol{\delta}_1$ the projection of $\boldsymbol{\delta}$ onto the image of $\mathbf{A}$ and by $\boldsymbol{\delta}_2$ the projection onto the orthogonal complement. We are interested in lower bounding the quotient

$$\frac{\|\boldsymbol{\delta}_2\|_2}{\sqrt{\|\boldsymbol{\delta}_1\|_2^2 + \|\boldsymbol{\delta}_2\|_2^2}} = \frac{\|\boldsymbol{\delta}_2\|_2/\|\boldsymbol{\delta}_1\|_2}{\sqrt{1+(\|\boldsymbol{\delta}_2\|_2/\|\boldsymbol{\delta}_1\|_2)^2}}. \tag{21}$$

To this end, we use the following sequence of inequalities:

$$\begin{aligned}
\|\boldsymbol{\delta}_1\|_1 &= \|(\boldsymbol{\delta}_1)_{J^c}\|_1 + \|(\boldsymbol{\delta}_1)_J\|_1 \\
&\le \|\boldsymbol{\delta}_{J^c}\|_1 + \|(\boldsymbol{\delta}_2)_{J^c}\|_1 + s\|\boldsymbol{\delta}_1\|_\infty \\
&\le \|\boldsymbol{\delta}_J\|_1 + \|(\boldsymbol{\delta}_2)_{J^c}\|_1 + s\|\boldsymbol{\delta}_1\|_\infty \\
&\le \|(\boldsymbol{\delta}_1)_J\|_1 + \|\boldsymbol{\delta}_2\|_1 + s\|\boldsymbol{\delta}_1\|_\infty \\
&\le \sqrt{n}\|\boldsymbol{\delta}_2\|_2 + 2s\|\boldsymbol{\delta}_1\|_\infty
\end{aligned}$$

This entails that

$$\|\boldsymbol{\delta}_2\|_2 \ge \frac{\|\boldsymbol{\delta}_1\|_1 - 2s\|\boldsymbol{\delta}\|_\infty}{\sqrt{n}} \ge \|\boldsymbol{\delta}_1\|_2 \inf_{\boldsymbol{w}\in\text{Im}(\mathbf{A})} \frac{\|\boldsymbol{w}\|_1 - 2s\|\boldsymbol{w}\|_\infty}{\sqrt{n}\|\boldsymbol{w}\|_2}$$

Let $\boldsymbol{v}\in\mathbb{R}^k$ be a vector such that $\mathbf{A}\boldsymbol{v} = \boldsymbol{w}$. We have $\|\boldsymbol{w}\|_\infty = \|\mathbf{A}\boldsymbol{v}\|_\infty \le \|\mathbf{A}\|_{2,\infty}\|\boldsymbol{v}\|_2$. Furthermore, $\|\boldsymbol{w}\|_2 = \|\mathbf{A}\boldsymbol{v}\|_2 \le \sqrt{n}\nu^*\|\boldsymbol{v}\|_2$. Thus

$$\frac{\|\boldsymbol{\delta}_2\|_2}{\|\boldsymbol{\delta}_2\|_2} \ge \inf_{\boldsymbol{v}} \frac{1}{n\nu^*}\frac{\|\mathbf{A}\boldsymbol{v}\|_1}{\|\boldsymbol{v}\|_2} - \frac{2s\|\mathbf{A}\|_{2,\infty}}{\nu^*\sqrt{n}} \ge \frac{\zeta_s(\mathbf{A})}{\nu^*}.$$

Injecting this bound in (21), the assertion of the lemma follows.