Electronic Journal of Statistics
ISSN: 1935-7524

Supplementary material to 'Non-negative least squares for high-dimensional linear models: consistency and sparse recovery without regularization'

Martin Slawski* **and Matthias Hein**

Saarland University
Campus E 1.1
Saarbrücken, Germany
e-mail: ms@cs.uni-saarland.de
e-mail: ms@cs.uni-saarland.de ; hein@cs.uni-saarland.de

1. Proofs of Propositions 3 and 4

We here provide proofs of Propositions 3 and 4 concerning random equi-correlation-like matrices. These proofs rely on a series of lemmas that are stated first.

1.1. Additional lemmas

We recall from Appendix A of the paper that a zero-mean random variable is called sub-Gaussian if there exists $\sigma > 0$ (referred to as sub-Gaussian parameter) so that the moment-generating function obeys the bound $\mathbf{E}[\exp(tZ)] \leq \exp(\sigma^2 t^2/2)\ \forall t \in \mathbb{R}$. If $Z_1, \dots, Z_n$ are i.i.d. copies of Z and $v_j \in \mathbb{R}^n$, $j = 1, \dots, p$, are fixed vectors, then

$$\mathbf{P}\left(\max_{1 \leq j \leq p} |v_j^\top \mathbf{Z}| > \sigma \max_{1 \leq j \leq p} \|v_j\|_2 \left(\sqrt{2 \log p} + z \right) \right) \leq 2 \exp\left(-\frac{1}{2} z^2 \right), \quad z \geq 0. \tag{1.1}$$

Bernstein-type inequality for squared sub-Gaussian random variables

The following exponential inequality combines Lemma 14, Proposition 16 and Remark 18 in [4].

Lemma 1. *Let $Z_1, \dots, Z_m$ be i.i.d. zero-mean sub-Gaussian random variables with parameter σ and the property that $\mathbf{E}[Z_1^2] \leq 1$. Then for any $z \geq 0$, one has*

$$\mathbf{P}\left(\sum_{i=1}^m Z_i^2 > m + zm \right) \leq \exp\left(-c \min\left\{ \frac{z^2}{\sigma^4}, \frac{z}{\sigma^2} \right\} m \right), \tag{1.2}$$

where $c > 0$ is an absolute constant.

Concentration of extreme singular values of sub-Gaussian random matrices

Let $s_{\min}(A)$ and $s_{\max}(A)$ denote the minimum and maximum singular value of a matrix A. The following lemma is a special case of Theorem 39 in [4].

*Supported by the Cluster of Excellence 'Multimodal Computing and Interaction' (MMCI) of Deutsche Forschungsgemeinschaft (DFG)

Lemma 2. *Let A be an $n \times s$ matrix with i.i.d. zero-mean sub-Gaussian entries with sub-Gaussian parameter σ and unit variance. Then for every $z \geq 0$, with probability at least $1 - 2\exp(-cz^2)$, one has*

$$s_{\max}\left(\frac{1}{n}A^\top A - I\right) \leq \max(\delta, \delta^2), \quad \textit{where } \delta = C\sqrt{\frac{s}{n}} + \frac{z}{\sqrt{n}}, \tag{1.3}$$

with C, c depending only on σ.

Entry-wise concentration of the Gram matrix associated with a sub-Gaussian random matrix

The next lemma results from Lemma 1 in [2] and the union bound.

Lemma 3. *Let X be an $n \times p$ random matrix of i.i.d. zero-mean, unit variance sub-Gaussian entries with parameter σ. Then*

$$\mathbf{P}\left(\max_{1 \leq j,k \leq p}\left|\left(\frac{1}{n}X^\top X - I\right)_{jk}\right| > z\right) \leq 4p^2 \exp\left(-\frac{nz^2}{128(1+4\sigma^2)^2}\right) \tag{1.4}$$

for all $z \in \big(0, 8(1+4\sigma^2)\big)$.

1.2. Application to Ens$_+$

Recall that the class Ens$_+$ is given by

$$\text{Ens}_+ : \; X = (x_{ij})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}, \; \{x_{ij}\} \text{ i.i.d. from a sub-Gaussian distribution on } \mathbb{R}_+. \tag{1.5}$$

We shall make use of the following decomposition valid for any X from (1.5).

$$X = \widetilde{X} + \mu \mathbb{1}, \tag{1.6}$$

where the entries $\{\widetilde{x}_{ij}\}$ of $\widetilde{X}$ are zero mean sub-Gaussian random variables with parameter σ, say, $\mu = \mathbf{E}[x_{11}]$ and $\mathbb{1}$ is an $n \times p$-matrix of ones. In the sequel, we specialize to the case where the entries of X are scaled such that

$$\Sigma^* = \mathbf{E}\left[\frac{1}{n}X^\top X\right] = (1-\rho)I + \rho \mathbf{1}\mathbf{1}^\top \tag{1.7}$$

for $\rho \in (0,1)$, i.e. the population Gram matrix has equi-correlation structure. Then, decomposition (1.6) becomes

$$X = \widetilde{X} + \sqrt{\rho}\mathbb{1}, \quad \text{and } \mathbf{E}[\widetilde{x}_{11}^2] = (1-\rho). \tag{1.8}$$

Accordingly, we have the following expansion of $\Sigma = \frac{1}{n}X^\top X$.

$$\Sigma = \frac{1}{n}\widetilde{X}^\top \widetilde{X} + \sqrt{\rho}\left(\frac{1}{n}\widetilde{X}^\top \mathbb{1} + \frac{1}{n}\mathbb{1}^\top \widetilde{X}\right) + \rho \mathbf{1}\mathbf{1}^\top, \quad \text{where } \; \mathbf{E}\left[\frac{1}{n}\widetilde{X}^\top \widetilde{X}\right] = (1-\rho)I. \tag{1.9}$$

Observe that

$$n^{-1}\widetilde{X}^\top \mathbb{1} = D\mathbf{1}\mathbf{1}^\top, \quad \text{and } \; n^{-1}\mathbb{1}^\top \widetilde{X} = \mathbf{1}\mathbf{1}^\top D, \tag{1.10}$$

where $D \in \mathbb{R}^{p \times p}$ is a diagonal matrix with diagonal entries $d_{jj} = n^{-1}\sum_{i=1}^n \widetilde{x}_{ij}$, $j = 1, \ldots, p$. It hence follows from (1.1) that

$$\mathbf{P}\left(\max_{j,k}\left|n^{-1}\widetilde{X}^\top \mathbb{1}\right|_{jk} > 2\sigma\sqrt{\frac{2\log(p \vee n)}{n}}\right) \leq \frac{2}{p \vee n}, \tag{1.11}$$

Combining (1.7), (1.9), (1.11) and Lemma 3, it follows that there exists a constant $C > 0$ depending only on σ such that

$$\mathbf{P}\left(\max_{j,k}\left|\left(\frac{X^\top X}{n} - \Sigma^*\right)_{jk}\right| > C\sqrt{\frac{\log(p \vee n)}{n}}\right) \leq \frac{6}{p \vee n}. \tag{1.12}$$

Let now $S \subset \{1, \dots, p\}$, $|S| = s < n$ be given. Without loss of generality, let us assume that $S = \{1, \dots, s\}$. In the sequel, we control $s_{\max}(\Sigma^*_{SS} - \Sigma_{SS})$. From decomposition (1.9), we obtain that

$$s_{\max}(\Sigma^*_{SS} - \Sigma_{SS}) \leq (1-\rho) s_{\max}\left(\frac{1}{1-\rho}\frac{\widetilde{X}_S^\top \widetilde{X}_S}{n} - I\right) + 2\sqrt{\rho}\, s_{\max}\left(\frac{\widetilde{X}_S^\top \mathbb{1}_S}{n}\right) \tag{1.13}$$

Introduce $w = \left(\sum_{i=1}^n \widetilde{x}_{i1}/n, \dots, \sum_{i=1}^n \widetilde{x}_{is}/n\right)^\top$ as the vector of column means of $\widetilde{X}_S$. We have that

$$s_{\max}\left(\frac{\widetilde{X}_S^\top \mathbb{1}_S}{n}\right) = \sup_{\|u\|_2=1}\sup_{\|v\|_2=1} u^\top \frac{\widetilde{X}_S^\top \mathbb{1}_S}{n} v = \sup_{\|u\|_2=1}\sup_{\|v\|_2=1} u^\top w \mathbf{1}^\top v = \sqrt{s}\|w\|_2. \tag{1.14}$$

Moreover,

$$\|w\|_2^2 = \sum_{j=1}^s \left(\frac{\sum_{i=1}^n \widetilde{x}_{ij}}{n}\right)^2 = \frac{1}{n}\sum_{j=1}^s z_j^2, \quad \text{where } \; z_j = n^{-1/2}\sum_{i=1}^n \widetilde{x}_{ij}. \tag{1.15}$$

Noting that the $\{z_j\}_{j=1}^s$ are i.i.d. zero-mean sub-Gaussian random variables with parameter σ and variance no larger than one, we are in position to apply Lemma 1, which yields that for any $t \geq 0$

$$\mathbf{P}\left(\|w\|_2^2 > \frac{s}{n}(1+t)\right) \leq \exp\left(-c\min\left(\frac{t^2}{\sigma^4}, \frac{t}{\sigma^2}\right)s\right). \tag{1.16}$$

Combining (1.13), (1.14) and (1.16) and using Lemma 2 to control the term $s_{\max}\left(\frac{1}{1-\rho}\frac{\widetilde{X}_S^\top \widetilde{X}_S}{n} - I\right)$, we obtain that for any $t \geq 0$ and any $z \geq 0$

$$\begin{aligned}&\mathbf{P}\left(s_{\max}(\Sigma^*_{SS} - \Sigma_{SS}) > \max\left\{C\sqrt{\frac{s}{n}} + \frac{z}{\sqrt{n}}, \left(C\sqrt{\frac{s}{n}} + \frac{z}{\sqrt{n}}\right)^2\right\} + 2\sqrt{\frac{s^2(1+t)}{n}}\right)\\ &\leq \exp(-c_1 \min\{t, t^2\}s) - 2\exp(-c_2 z^2),\end{aligned} \tag{1.17}$$

where $C, c_1, c_2 > 0$ only depend on the sub-Gaussian parameter σ. Equipped with these auxiliary results, we now turn to the proofs of Proposition 3 and 4.

1.3. Proof of Proposition 3

Let us first recall the restricted eigenvalue condition.

Condition 2. Let $\mathcal{J}(s) = \{J \subseteq \{1, \dots, p\} : 1 \leq |J| \leq s\}$ and for $J \in \mathcal{J}(s)$ and $\alpha \geq 1$,

$$\mathcal{R}(J, \alpha) = \{\delta \in \mathbb{R}^p : \|\delta_{J^c}\|_1 \leq \alpha\|\delta_J\|_1\}.$$

We say that the design satisfies the (α, s)-**restricted eigenvalue condition** if there exists a constant $\phi(\alpha, s)$ so that

$$\min_{J \in \mathcal{J}(s)} \min_{\delta \in \mathcal{R}(J,\alpha)\setminus \mathbf{0}} \frac{\delta^\top \Sigma \delta}{\|\delta_J\|_2^2} \geq \phi(\alpha, s) > 0. \tag{1.18}$$

The proof of Proposition 3 relies on a recent result in [3]. In order to state that result, we need the following preliminaries concerning ψ_2-random variables taken from [1] (see Definition 1.1.1 and Theorem 1.1.5 therein).

Definition 1. *A random variable Z is said to be ψ_2 with parameter $\theta > 0$ if*

$$\inf\left\{a > 0 :\ \mathbf{E}\left[\exp(Z^2/a^2)\right] \leq e\right\} \leq \theta. \tag{1.19}$$

Lemma 4. *If a random variable Z has the property that there exist positive constants C, C' so that $\forall z \geq C'$*

$$\mathbf{P}\left(|Z| \geq z\right) \leq \exp\left(-z^2/C^2\right),$$

then Z is ψ_2 with parameter no more than $2\max(C, C')$.

The following statement is essentially a special case of Theorem 1.6 in [3]. We state it in simplified form that is sufficient for our purpose here.

Lemma 5. *Let $\Psi \in \mathbb{R}^{n\times p}$ be a matrix whose rows $\Psi^1, \ldots, \Psi^n$, are independent random vectors that are*

1. ***isotropic**, i.e. $\mathbf{E}[\langle \Psi^i, u\rangle^2] = 1$, $i = 1, \ldots, n$,*
2. ***ψ_2**, i.e. there exists $\theta > 0$ such that for every unit vector $u \in \mathbb{R}^p$*

$$\inf\left\{a > 0 :\ \mathbf{E}\left[\exp(\langle \Psi^i, u\rangle^2/a^2)\right] \leq e\right\} \leq \theta, \quad i = 1, \ldots, n. \tag{1.20}$$

Let further $R \in \mathbb{R}^{p\times p}$ be a positive definite matrix with minimum eigenvalue $\vartheta > 0$ and set $\Gamma = \frac{1}{n} R^\top \Psi^\top \Psi R$. Then, for any $\delta \in (0,1)$ and any $\alpha \in [1,\infty)$, there exist positive constants C_θ, $c > 0$ (the first depending on the ψ_2 parameter θ) so that if

$$n \geq \frac{C_\theta}{\delta^2} s \left(1 + \frac{16(3\alpha^2)(3\alpha+1)}{\vartheta^2\delta^2}\right) \log\left(c\frac{p}{s\delta}\right),$$

with probability at least $1 - 2\exp(-\delta^2 n/C_\theta)$, Γ satisfies the (α, s)-restricted eigenvalue condition with $\phi(\alpha, s) = \vartheta^2(1-\delta)^2$.

We now state and prove Proposition 3.

Proposition 3. Let X be a random matrix from Ens_+ (1.5) scaled such that $\Sigma^* = \mathbf{E}[\frac{1}{n}X^\top X] = (1-\rho)I + \rho\mathbf{1}\mathbf{1}^\top$ for some $\rho \in (0,1)$. Set $\delta \in (0,1)$. There exists constants $C, c > 0$ depending only on δ, ρ and the sub-Gaussian parameter of the centered entries of X so that if $n \geq C\, s\log(p \vee n)$, then, with probability at least $1 - \exp(-c\delta^2 n) - 6/(p\vee n)$, $\Sigma = X^\top X/n$ has the self-regularizing property with $\tau^2 = \rho/2$ and satisfies the $(3/\tau^2, s)$ restricted eigenvalue condition of Theorem 2 with $\phi(3/\tau^2, s) = (1-\rho)(1-\delta)^2$.

Proof. We first show that Σ satisfies the self-regularizing property with $\tau^2 \geq \rho/2$ with probability at least $1 - 6/(p \vee n)$. According to Eq.(6.4) in the paper, we have

$$\tau_0^2 = \min_{\lambda \in T^{p-1}} \lambda^\top \Sigma \lambda \geq \min_{\lambda \in T^{p-1}} \lambda^\top \Sigma^* \lambda - \max_{\lambda \in T^{p-1}} \lambda^\top \left(\Sigma^* - \Sigma\right)\lambda \geq \rho - \max_{j,k}\left|(\Sigma - \Sigma^*)_{jk}\right|.$$

Consequently, in virtue of (1.12), there exists a numerical constant C' depending on σ and ρ only so that if $n \geq C'\log(p \vee n)$, $\tau_0^2 \geq \frac{1}{2}\rho$ with the probability as claimed. In the sequel, it will be shown that conditional on the event $\{\tau_0^2 \geq \rho/2\}$, Lemma 5 can be applied with

$$\Gamma = \Sigma, \quad R = (\Sigma^*)^{1/2}, \quad \Psi = X(\Sigma^*)^{-1/2}, \quad \vartheta^2 = 1-\rho, \quad \alpha = \frac{3}{\tau^2} \leq \frac{6}{\rho}, \quad \theta = C_{\sigma,\rho},$$

where $(\Sigma^*)^{1/2}$ is the root of Σ^* and $C_{\sigma,\rho}$ is a constant depending only on σ and ρ. By construction, $\Psi = X(\Sigma^*)^{-1/2}$ has independent isotropic rows. It remains to establish that the rows satisfy condition (1.20) of Lemma 5. Since the rows of Ψ are i.i.d., it suffices to consider a single row. Let us write X^1 for the transpose of the first row of X and accordingly $\Psi^1 = (\Sigma^*)^{-1/2} X^1$ for the transpose of the first row of Ψ. Furthermore, we make use of the decomposition $X^1 = \widetilde{X}^1 + \sqrt{\rho}\mathbf{1}$, where the entries of $\widetilde{X}^1$ are i.i.d zero-mean sub-Gaussian random variables with parameter σ (cf. (1.8)). We then have for any unit vector u

$$\begin{aligned}\left\langle \Psi^1, u\right\rangle = \left\langle (\Sigma^*)^{-1/2} X^1, u\right\rangle &= \left\langle (\Sigma^*)^{-1/2} (\widetilde{X}^1 + \sqrt{\rho}\mathbf{1}), u\right\rangle \\ &= \left\langle \widetilde{X}^1, (\Sigma^*)^{-1/2} u\right\rangle + \sqrt{\frac{\rho}{(1-\rho)+p\rho}} \left\langle \mathbf{1}, u\right\rangle \\ &\le \left\langle \widetilde{X}^1, (\Sigma^*)^{-1/2} u\right\rangle + 1.\end{aligned}$$

For the second equality, we have used that $\mathbf{1}$ is an eigenvector of Σ^* with eigenvalue $1 + (p-1)\rho$, while the inequality results from Cauchy-Schwarz. We now estimate the moment-generating function of the random variable $\left\langle \Psi^1, u\right\rangle$ as follows. For any $t \ge 0$, we have

$$\begin{aligned}\mathbf{E}[\exp(t \left\langle \Psi^1, u\right\rangle)] &\le \exp(t)\, \mathbf{E}\left[\exp\left(t \left\langle \widetilde{X}^1, (\Sigma^*)^{-1/2} u\right\rangle\right)\right] \\ &\le \exp(t)\, \mathbf{E}\left[\exp\left(\frac{\sigma^2 t^2}{2} \|(\Sigma^*)^{-1/2} u\|_2^2\right)\right] \\ &\le \exp(t) \exp\left(\frac{\sigma^2 t^2}{2(1-\rho)}\right) \\ &\le e \exp\left(\frac{(\sigma^2+2)t^2}{2(1-\rho)}\right) = e \exp\left(\frac{\widetilde{\sigma}^2 t^2}{2}\right),\end{aligned}$$

where $\widetilde{\sigma} = \sqrt{(\sigma^2+2)/(1-\rho)}$. For the third equality, we have used that the maximum eigenvalue of $(\Sigma^*)^{-1}$ equals $(1-\rho)^{-1}$. Analogously, we obtain that

$$-\left\langle \Psi^1, u\right\rangle \le \left\langle -\widetilde{X}^1, (\Sigma^*)^{-1/2} u\right\rangle + 1, \quad \text{and} \quad \mathbf{E}[\exp(t \left\langle -\Psi^1, u\right\rangle)] \le e \exp\left(\frac{\widetilde{\sigma}^2 t^2}{2}\right) \;\; \forall t \ge 0.$$

From the Chernov bound, we hence obtain that for any $z \ge 0$

$$\mathbf{P}(|\left\langle \Psi^1, u\right\rangle| > z) \le 2e \exp\left(-\frac{z^2}{2\widetilde{\sigma}^2}\right).$$

Invoking Lemma 4 with $C' = \widetilde{\sigma}\sqrt{3\log(2e)}$ and $C = \sqrt{6}\widetilde{\sigma}$, it follows that the random variable $\left\langle \Psi^1, u\right\rangle$ is ψ_2 with parameter $2\sqrt{6}\widetilde{\sigma} =: C_{\sigma,\rho}$, and we conclude that the rows of the matrix Ψ indeed satisfy condition (1.20) with θ equal to that value of the parameter. $\square$

1.4. Proof of Proposition 4

Proposition 4. Let X be a random matrix from Ens_+ (1.5) scaled such that $\Sigma^* = \mathbf{E}[\frac{1}{n} X^\top X] = (1-\rho)I + \rho\mathbf{1}\mathbf{1}^\top$ for some $\rho \in (0,1)$. Fix $S \subset \{1, \ldots, p\}$, $|S| \le s$. Then there exists constants $c, c', C, C' > 0$ depending only on ρ and the sub-Gaussian parameter of the centered entries of X such that for all $n \ge Cs^2 \log(p \vee n)$,

$$\tau^2(S) \ge cs^{-1} - C'\sqrt{\frac{\log p}{n}}$$

with probability no less than $1 - 6/(p \vee n) - 3\exp(-c'(s \vee \log n))$.

Proof. The scaling of $\tau^2(S)$ is analyzed based on the representation

$$\tau^2(S) = \min_{\theta \in \mathbb{R}^s,\, \lambda \in T^{p-s-1}} \frac{1}{n} \|X_S\theta - X_{S^c}\lambda\|_2^2. \tag{1.21}$$

In the following, denote by $\mathbb{S}^{s-1} = \{u \in \mathbb{R}^s : \|u\|_2 = 1\}$ the unit sphere in $\mathbb{R}^s$. Expanding the square in (1.21), we have

$$\begin{aligned}
\tau^2(S) &= \min_{\theta \in \mathbb{R}^s,\ \lambda \in T^{p-s-1}} \ \theta^\top \Sigma_{SS} \theta - 2\theta^\top \Sigma_{SS^c} \lambda + \lambda^\top \Sigma_{S^c S^c} \lambda \\
&\geq \min_{r \geq 0,\ u \in \mathbb{S}^{s-1},\ \lambda \in T^{p-s-1}} \ r^2 u^\top \Sigma^*_{SS} u - r^2 s_{\max}\left(\Sigma_{SS} - \Sigma^*_{SS}\right) - \\
&\qquad - 2ru^\top \Sigma_{SS^c} \lambda + \lambda^\top \Sigma_{S^c S^c} \lambda \\
&\geq \min_{r \geq 0,\ u \in \mathbb{S}^{s-1},\ \lambda \in T^{p-s-1}} \ r^2 u^\top \Sigma^*_{SS} u - r^2 s_{\max}\left(\Sigma_{SS} - \Sigma^*_{SS}\right) \\
&\qquad - 2\rho r u^\top \mathbf{1} - 2ru^\top (\Sigma_{SS^c} - \Sigma^*_{SS^c})\lambda + \rho + \frac{1-\rho}{p-s} - \\
&\qquad - \max_{\lambda \in T^{p-s-1}} \left|\lambda^\top (\Sigma_{S^c S^c} - \Sigma^*_{S^c S^c})\lambda\right|.
\end{aligned} \tag{1.22}$$

For the last inequality, we have used that $\min_{\lambda \in T^{p-s-1}} \lambda^\top \Sigma^*_{S^c S^c} \lambda = \rho + \frac{1-\rho}{p-s}$. We further set

$$\Delta = s_{\max}\left(\Sigma_{SS} - \Sigma^*_{SS}\right), \tag{1.23}$$

$$\delta = \max_{u \in \mathbb{S}^{s-1}, \lambda \in T^{p-s-1}} \left|u^\top \left(\Sigma_{S^c S^c} - \Sigma^*_{S^c S^c}\right) \lambda\right|. \tag{1.24}$$

The random terms Δ and δ will be controlled uniformly over $u \in \mathbb{S}^{s-1}$ and $\lambda \in T^{p-s-1}$ below by invoking (1.12) and (1.17). For the moment, we treat these two terms as constants. We now minimize the lower bound in (1.22) w.r.t. u and r separately from λ. This minimization problem involving u and r only reads

$$\min_{r \geq 0,\ u \in \mathbb{S}^{s-1}} r^2 u^\top \Sigma^*_{SS} u - 2\rho r u^\top \mathbf{1} - r^2 \Delta - 2r\delta. \tag{1.25}$$

We first derive an expression for

$$\phi(r) = \min_{u \in \mathbb{S}^{s-1}} r^2 u^\top \Sigma^*_{SS} u - 2\rho r u^\top \mathbf{1}. \tag{1.26}$$

We decompose $u = u^{\|} + u^{\perp}$, where $u^{\|} = \left\langle \frac{\mathbf{1}}{\sqrt{s}}, u \right\rangle \frac{\mathbf{1}}{\sqrt{s}}$ is the projection of u on the unit vector $\mathbf{1}/\sqrt{s}$, which is an eigenvector of Σ^*_{SS} associated with its largest eigenvalue $1 + \rho(s-1)$. By Parseval's identity, we have $\|u^{\|}\|_2^2 = \gamma$, $\|u^{\perp}\|_2^2 = (1-\gamma)$ for some $\gamma \in [0,1]$. Inserting this decomposition into (1.26) and noting that the remaining eigenvalues of Σ^*_{SS} are all equal to $(1-\rho)$, we obtain that

$$\begin{aligned}
&\phi(r) = \min_{\gamma \in [0,1]} \Phi(\gamma, r), \\
&\text{with } \Phi(\gamma, r) = r^2 \gamma \underbrace{(1 + (s-1)\rho)}_{s_{\max}(\Sigma^*_{SS})} + r^2(1-\gamma) \underbrace{(1-\rho)}_{s_{\min}(\Sigma^*_{SS})} - 2\rho r \sqrt{\gamma}\sqrt{s},
\end{aligned} \tag{1.27}$$

where we have used that $\langle u^{\perp}, \mathbf{1} \rangle = 0$. Let us put aside the constraint $\gamma \in [0,1]$ for a moment. The function Φ in (1.27) is a convex function of γ, hence we may find an (unconstrained) minimizer $\widetilde{\gamma}$ by differentiating and setting the derivative equal to zero. This yields $\widetilde{\gamma} = \frac{1}{r^2 s}$, which coincides with the constrained minimizer if and only if $r \geq \frac{1}{\sqrt{s}}$. Otherwise, $\widetilde{\gamma} \in \{0, 1\}$. We can rule out the case $\widetilde{\gamma} = 0$, since for all $r < 1/\sqrt{s}$

$$\Phi(0, r) = r^2(1-\rho) > r^2(1 + (s-1)\rho) - 2\rho r\sqrt{s} = \Phi(1, r).$$

We have $\Phi(\frac{1}{r^2 s}, r) = r^2(1-\rho) - \rho$ and $\Phi(\frac{1}{r^2 s}, \frac{1}{\sqrt{s}}) = \Phi(1, \frac{1}{\sqrt{s}})$. Hence, the function $\phi(r)$ in (1.26) is given by

$$\phi(r) = \begin{cases} r^2 s_{\max}(\Sigma^*_{SS}) - 2\rho r\sqrt{s} & r \leq 1/\sqrt{s}, \\ r^2(1-\rho) - \rho & \text{otherwise.} \end{cases} \tag{1.28}$$

The minimization problem (1.25) to be considered eventually reads

$$\min_{r \geq 0} \psi(r), \quad \text{where } \ \psi(r) = \phi(r) - r^2 \Delta - 2r\delta. \tag{1.29}$$

We argue that it suffices to consider the case $r \leq 1/\sqrt{s}$ in (1.28) provided

$$((1-\rho) - \Delta) > \delta\sqrt{s}, \tag{1.30}$$

a condition we will comment on below. If this condition is met, differentiating shows that ψ is increasing on $(\frac{1}{\sqrt{s}}, \infty)$. In fact, for all r in that interval,

$$\frac{d}{dr}\psi(r) = 2r(1-\rho) - 2r\Delta - 2\delta, \text{ and thus}$$

$$\frac{d}{dr}\psi(r) > 0 \text{ for all } r \in \left(\frac{1}{\sqrt{s}}, \infty\right) \;\Leftrightarrow\; \frac{1}{\sqrt{s}}((1-\rho) - \Delta) > \delta.$$

Considering the case $r \leq 1/\sqrt{s}$, we observe that $\psi(r)$ is convex provided

$$s_{\max}(\Sigma^*_{SS}) > \Delta, \tag{1.31}$$

a condition we shall comment on below as well. Provided (1.30) and (1.31) hold true, differentiating (1.29) and setting the result equal to zero, we obtain that the minimizer $\widehat{r}$ of (1.29) is given by $(\rho\sqrt{s} + \delta)/(s_{\max}(\Sigma^*_{SS}) - \Delta)$. Substituting this result back into (1.29) and in turn into the lower bound (1.22), one obtains after collecting terms

$$\begin{aligned} \tau^2(S) \geq & \rho \frac{(1-\rho) - \Delta}{(1-\rho) + s\rho - \Delta} - \frac{2\rho\sqrt{s}\delta + \delta^2}{s_{\max}(\Sigma^*_{SS}) - \Delta} + \frac{1-\rho}{p-s} - \\ & - \max_{\lambda \in T^{p-s-1}} \left|\lambda^\top (\Sigma_{S^c S^c} - \Sigma^*_{S^c S^c})\lambda\right|. \end{aligned} \tag{1.32}$$

In order to control Δ (1.23), we apply (1.17) with the choices

$$z = \sqrt{s \vee \log n}, \quad \text{and} \quad t = 1 \vee \frac{\log n}{s}.$$

Consequently, there exists a constant $C_1 > 0$ depending only on σ so that if $n > C_1(s \vee \log n)$, we have that

$$\begin{aligned} &\mathbf{P}(\mathcal{A}) \geq 1 - 3\exp(-c'(s \vee \log n)), \\ &\quad \text{where } \ \mathcal{A} = \left\{ \Delta \leq 2\sqrt{\frac{s^2(1 + 1 \vee (\log(n)/s))}{n}} + C'\sqrt{\frac{s \vee \log n}{n}} \right\} \end{aligned} \tag{1.33}$$

In order to control δ (1.24) and the last term in (1.32), we make use of (1.12), which yields that

$$\begin{aligned} &\mathbf{P}(\mathcal{B}) \geq 1 - \frac{6}{p \vee n}, \ \text{ where} \\ &\mathcal{B} = \left\{ \delta \leq C\sqrt{\frac{s\log(p \vee n)}{n}} \right\} \cap \left\{ \sup_{\lambda \in T^{p-s-1}} \left|\lambda^\top(\Sigma_{S^c S^c} - \Sigma^*_{S^c S^c})\lambda\right| \leq C\sqrt{\frac{\log(p \vee n)}{n}} \right\}. \end{aligned} \tag{1.34}$$

For the remainder of the proof, we work conditional on the two events $\mathcal{A}$ and $\mathcal{B}$. In view of (1.33) and (1.34), we first note that there exists $C_2 > 0$ depending only on σ and ρ such that if $n \geq C_2 s^2 \log(p \vee n)$ the two conditions (1.30) and (1.31) supposed to be fulfilled previously indeed hold. To conclude the proof, we re-write (1.32) as

$$\begin{aligned} \tau^2(S) \geq & \frac{\rho(1 - \Delta/(1-\rho))}{(1 - \Delta/(1-\rho)) + s\frac{\rho}{1-\rho}} + \frac{2\rho\frac{\sqrt{s}}{1+(s-1)\rho}\delta}{1 - \Delta/(1 + (s-1)\rho)} - \frac{\delta^2/(1 + (s-1)\rho)}{1 - \Delta/(1 + (s-1)\rho)} - \\ & - \max_{\lambda \in T^{p-s-1}} \left|\lambda^\top(\Sigma_{S^c S^c} - \Sigma^*_{S^c S^c})\lambda\right|. \end{aligned} \tag{1.35}$$

Conditional on $\mathcal{A} \cap \mathcal{B}$, there exists $C_3 > 0$ depending only on σ and ρ such that if $n \geq C_3(s^2 \vee (s \log n))$, when inserting the resulting scalings separately for each summand in (1.35), we have that

$$\begin{aligned} & c_1 s^{-1} - C_4 \sqrt{\frac{\log(p \vee n)}{n}} - C_5 \frac{\log(p \vee n)}{n} - C_6 \sqrt{\frac{\log(p \vee n)}{n}} \\ & = c_1 s^{-1} - C_7 \sqrt{\frac{\log(p \vee n)}{n}}. \end{aligned} \tag{1.36}$$

We conclude that if $n \geq \max\{C_1, C_2, C_3\} s^2 \log(p \vee n)$, (1.36) holds with probability no less than $1 - \frac{6}{p \vee n} - 3\exp(-c'(s \vee \log n))$.

□

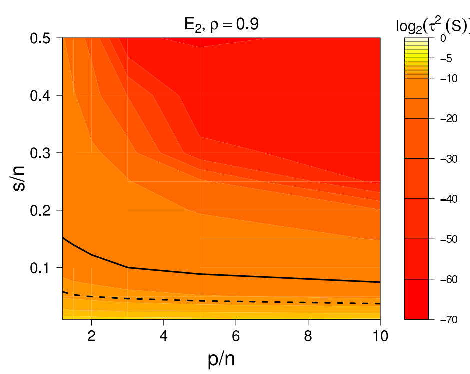
2. Empirical scaling of $\tau^2(S)$ for Ens_+

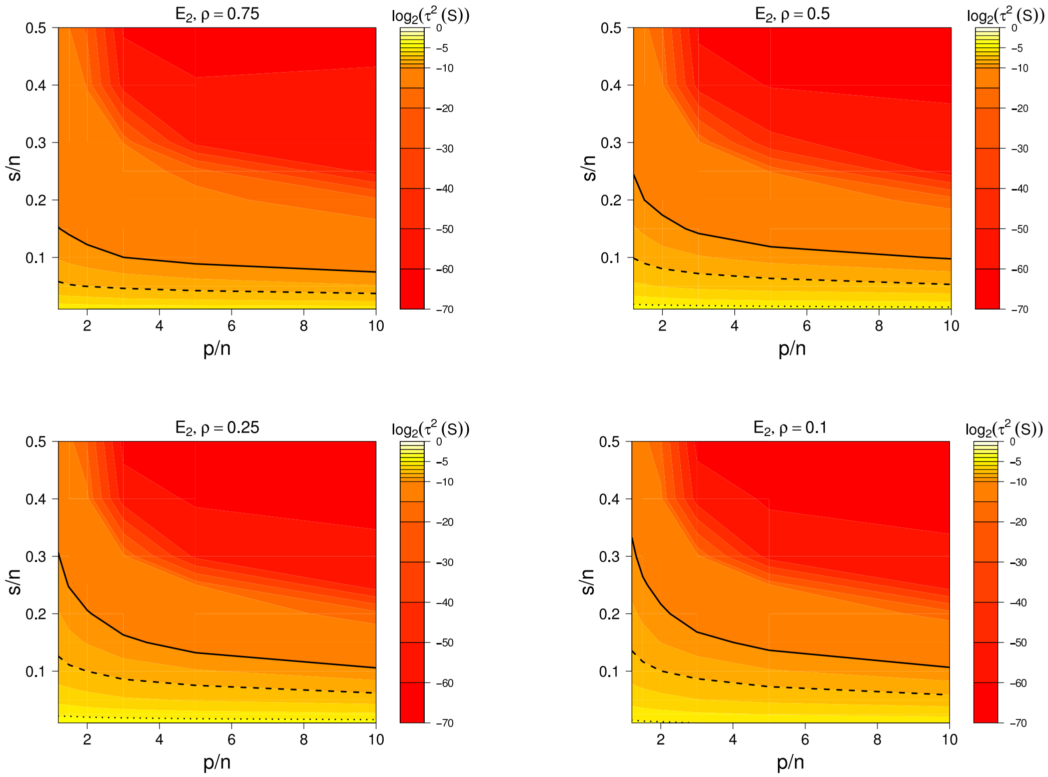
In Section 6.3, we have empirically investigated the scaling of $\tau^2(S)$ for the class (1.5) in a high-dimensional setting for the following designs.

E_1: $\{x_{ij}\} \overset{\text{i.i.d.}}{\sim}$ $a\,\text{uniform}([0, 1/\sqrt{3 \cdot a}]) + (1-a)\delta_0$, $a \in \{1, \frac{2}{3}, \frac{1}{3}, \frac{2}{15}\}$ $(\rho \in \{\frac{3}{4}, \frac{1}{2}, \frac{1}{3}, \frac{1}{10}\})$
E_2: $\{x_{ij}\} \overset{\text{i.i.d.}}{\sim}$ $\frac{1}{\sqrt{\pi}}\,\text{Bernoulli}(\pi)$, $\pi \in \{\frac{1}{10}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{9}{10}\}$ $(\rho \in \{\frac{1}{10}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{9}{10}\})$
E_3: $\{x_{ij}\} \overset{\text{i.i.d.}}{\sim}$ $|Z|$, $Z \sim a\,\text{Gaussian}(0,1) + (1-a)\delta_0$, $a \in \{1, \frac{\pi}{4}, \frac{\pi}{8}, \frac{\pi}{20}\}$ $(\rho \in \{\frac{2}{\pi}, \frac{1}{2}, \frac{1}{4}, \frac{1}{10}\})$
E_4: $\{x_{ij}\} \overset{\text{i.i.d.}}{\sim}$ $a\text{Poisson}(3/\sqrt{12\,a}) + (1-a)\delta_0$, $a \in \{1, \frac{2}{3}, \frac{1}{3}, \frac{2}{15}\}$ $(\rho \in \{\frac{3}{4}, \frac{1}{2}, \frac{1}{4}, \frac{1}{10}\})$

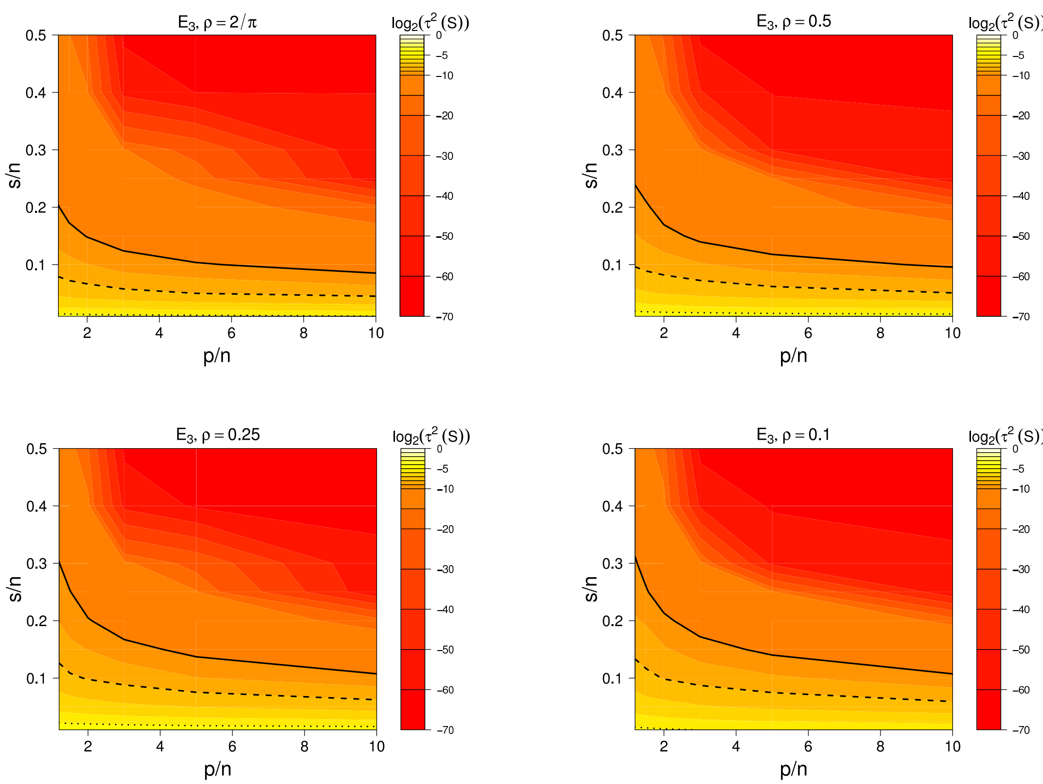
The results for E_1 are presented in the paper, and the results for E_2 to E_4 are displayed below.

2.1. E_2

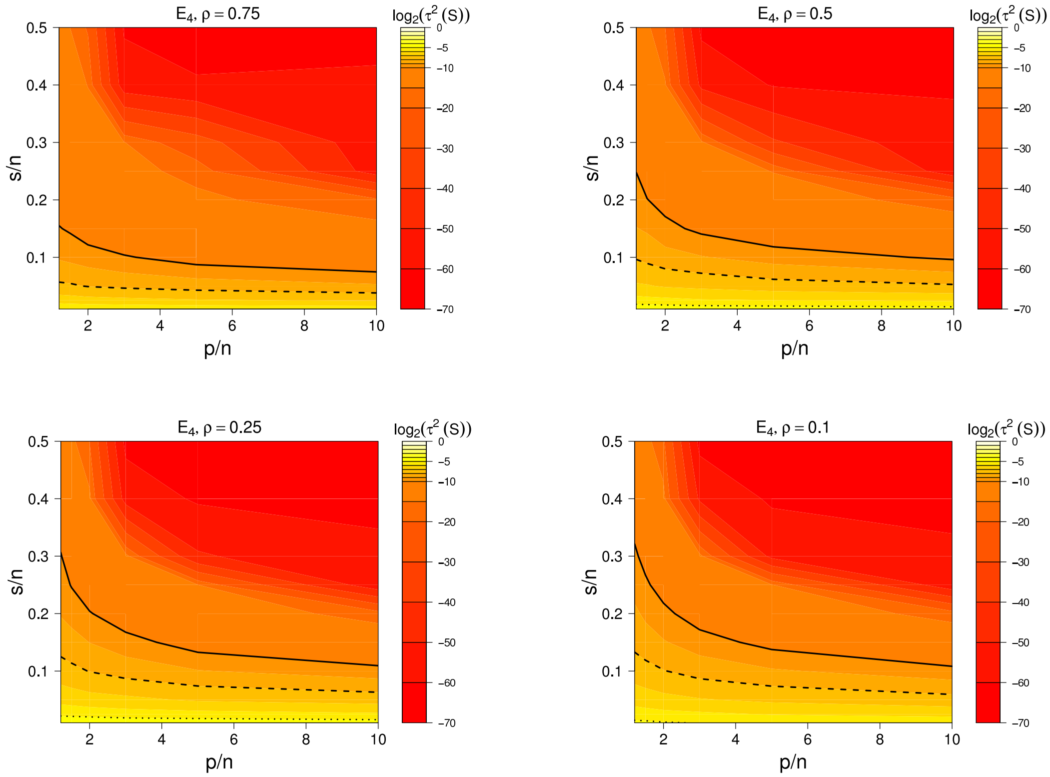



E2, ρ = 0.75
E2, ρ = 0.5
E2, ρ = 0.25
E2, ρ = 0.1
log2(τ2(S))
s/n
p/n

2.2. E_3


E3, ρ = 2/π
E3, ρ = 0.5
E3, ρ = 0.25
E3, ρ = 0.1
log2(τ2(S))
s/n
p/n

2.3. E_4



3. Additional empirical results on the ℓ_2-error in estimating β^*

The results in the two tables below are complementary to the experimental results in Section 6.2 of the paper. We here report $\|\widehat{\beta} - \beta^*\|_2$ (NNLS) and $\|\widehat{\beta}^{\ell_1,\succeq} - \beta^*\|_2$ (NNℓ_1) in correspondence to Tables 1 and 2 of the paper.

Design I

	p/n							
	2		3		5		10	
s/n	nnls	nnℓ_1	nnls	nnℓ_1	nnls	nnℓ_1	nnls	nnℓ_1
0.05	1.0±.01	1.0±.01	1.1±.01	1.1±.01	1.2±.01	1.2±.01	1.3±.01	1.3±.01
0.1	1.4±.01	1.4±.01	1.6±.01	1.6±.01	1.8±.02	1.8±.02	2.1±.02	2.1±.02
0.15	1.8±.01	1.8±.02	2.0±.02	2.0±.02	2.4±.02	2.4±.02	3.1±.04	3.4±.05
0.2	2.1±.02	2.2±.04	2.5±.02	2.6±.07	3.1±.03	3.3±.04	5.4±.10	6.9±.19
0.25	2.5±.02	2.6±.04	3.1±.03	3.7±.14	4.5±.07	7.2±.27	12.0±.2	15.3±.2
0.3	3.0±.03	3.4±.11	4.0±.05	5.5±.24	8.1±.19	12.8±.3	18.6±.1	19.8±.1

Design II

	p/n							
	2		3		5		10	
s/n	nnls	nnℓ_1	nnls	nnℓ_1	nnls	nnℓ_1	nnls	nnℓ_1
0.02	0.6±.01	0.7±.01	0.6±.01	0.7±.01	0.6±.01	0.7±.01	0.6±.01	0.7±.01
0.04	0.7±.01	1.0±.01	0.7±.01	1.0±.01	0.7±.01	1.0±.01	0.7±.01	1.0±.01
0.06	0.8±.01	1.2±.01	0.8±.01	1.2±.01	0.8±.01	1.2±.02	0.9±.01	1.2±.01
0.08	0.9±.01	1.3±.02	0.9±.01	1.3±.02	0.9±.01	1.3±.02	1.0±.01	1.4±.01
0.1	1.0±.01	1.4±.02	1.0±.01	1.5±.02	1.0±.01	1.5±.02	1.1±.01	1.5±.02

References

[1] D. Chafai, O. Guedon, G. Lecue, and A. Pajor. *Interactions between Compressed Sensing, Random Matrices and High-dimensional Geometry*. Lecture Notes, Université Paris Est Marne-la Vallée, 2012.

[2] P. Ravikumar, M. Wainwright, G. Raskutti, and B. Yu. High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *The Electronic Journal of Statistics*, 4:935–980, 2011.

[3] M. Rudelson and S. Zhou. Reconstruction from anisotropic random measurements. *IEEE Transactions on Information Theory*, to appear, 2012.

[4] R. Vershynin. *In: Compressed Sensing: Theory and Applications, Y. Eldar and G. Kutyniok (eds.)*, chapter 'Introduction to the non-asymptotic analysis of random matrices'. Cambridge University Press, 2012.