Submitted to the Annals of Statistics

SUPPLEMENTARY MATERIAL FOR: HIGH-DIMENSIONAL REGRESSION WITH NOISY AND MISSING DATA: PROVABLE GUARANTEES WITH NON-CONVEXITY

By Po-Ling Loh, Martin J. Wainwright

University of California, Berkeley

APPENDIX A: PROOF OF COROLLARY 1

The proof of this corollary is based on two technical lemmas, one establishing that the lower- and upper-RE conditions hold with high probability, and the other proving a form of the deviation bounds (3.2).

Lemma 1 (RE conditions, i.i.d. with additive noise). *Under the conditions of Corollary 1, there are universal positive constants c_i such that the matrix $\widehat{\Gamma}_{add}$ satisfies the lower- and upper-RE conditions with parameters $\alpha_1 = \frac{\lambda_{\min}(\Sigma_x)}{2}, \alpha_2 = \frac{3}{2}\lambda_{\max}(\Sigma_x)$, and*

$$\tau(n,p) = c_0 \, \lambda_{\min}(\Sigma_x) \max\Big(\frac{(\sigma_x^2+\sigma_w^2)^2}{\lambda_{\min}^2(\Sigma_x)}, 1\Big) \frac{\log p}{n},$$

with probability at least $1 - c_1 \exp\big(-c_2 n \min\big(\frac{\lambda_{\min}^2(\Sigma_x)}{(\sigma_x^2+\sigma_w^2)^2}, 1\big)\big)$.

Proof. Using Lemma 13 in Appendix F, together with the substitutions

$$\widehat{\Gamma} - \Sigma_x = \frac{Z^T Z}{n} - \Sigma_z, \quad \text{and} \quad s := \frac{1}{c}\frac{n}{\log p} \min\left\{\frac{\lambda_{\min}^2(\Sigma_x)}{\sigma^4}, 1\right\}, \tag{A.1}$$

where $\sigma^2 = \sigma_x^2 + \sigma_w^2$ and c is chosen sufficiently small so $s \geq 1$, we see that it suffices to show that

$$\underbrace{\sup_{v \in \mathbb{K}(2s)} \Big|v^T\Big(\frac{Z^T Z}{n} - \Sigma_z\Big)v\Big|}_{D(s)} \leq \frac{\lambda_{\min}(\Sigma_x)}{54}$$

with high probability.

Note that the matrix Z is sub-Gaussian with parameters $(\Sigma_x + \Sigma_w, \sigma^2)$. Consequently, by Lemma 15 in Appendix G, we have

$$\mathbb{P}\big[D(s) \geq t\big] \leq 2\exp\big(-c'n\min\big(\frac{t^2}{\sigma^4}, \frac{t}{\sigma^2}\big) + 2s\log p\big),$$

for some universal constant $c' > 0$. Setting $t = \frac{\lambda_{\min}(\Sigma_x)}{54}$, we see that as long as the constant c in the definition (A.1) is chosen sufficiently small, we are guaranteed that

$$\mathbb{P}\big[D(s) \geq \frac{\lambda_{\min}(\Sigma_x)}{54}\big] \leq 2\exp\big(-c_2 n\min\big(\frac{\lambda^2_{\min}(\Sigma_x)}{(\sigma_x^2+\sigma_w^2)^2}, 1\big)\big), \tag{A.2}$$

which establishes the result.

□

LEMMA 2 (Deviation conditions, additive noise). *Under the conditions of Corollary 1, there are universal positive constants c_i such the deviation bound* (3.1) *holds with parameter*

$$\varphi(\mathbb{Q}, \sigma_\epsilon) = c_0\sigma_z(\sigma_w + \sigma_\epsilon)\|\beta^*\|_2,$$

with probability at least $1 - c_1\exp(-c_2\log p)$.

PROOF. Using the fact that $y = X\beta^* + \epsilon$, we may write

$$\begin{aligned}
\|\widehat{\gamma} - \widehat{\Gamma}\beta^*\|_\infty &= \big\|\frac{Z^Ty}{n} - \big(\frac{Z^TZ}{n} - \Sigma_w\big)\beta^*\big\|_\infty \\
&= \big\|\frac{Z^T(X\beta^*+\epsilon)}{n} - \big(\frac{Z^TZ}{n} - \Sigma_w\big)\beta^*\big\|_\infty \\
&\leq \big\|\frac{Z^T\epsilon}{n}\big\|_\infty + \big\|\big(\Sigma_w - \frac{Z^TW}{n}\big)\beta^*\big\|_\infty.
\end{aligned}$$

Hence, the conclusion follows easily from Lemma 14 in Appendix G.

□

Extension to unknown Σ_w. In the case when Σ_w is unknown, we first verify the deviation bound (3.1). Note that the form of $\widehat{\gamma}$ is the same as in the case when Σ_w is known, so it suffices to bound the quantity $\|(\widetilde{\Gamma} - \Sigma_x)\beta^*\|_\infty$ w.h.p. Furthermore,

$$\begin{aligned}
\|(\widetilde{\Gamma} - \Sigma_x)\beta^*\|_\infty &\leq \|(\widetilde{\Gamma} - \widehat{\Gamma})\beta^*\|_\infty + \|(\widehat{\Gamma} - \Sigma_x)\beta^*\|_\infty \\
&= \|(\widehat{\Sigma}_w - \Sigma_w)\beta^*\|_\infty + \|(\widehat{\Gamma} - \Sigma_x)\beta^*\|_\infty,
\end{aligned}$$

and the second term is bounded by $c\sigma_z^2\sqrt{\frac{\log p}{n}}$ w.h.p., by Lemma 14 in Appendix G. If we use the estimator $\widehat{\Sigma}_w = \frac{1}{n}W_0^T W_0$, then

$$\mathbb{P}\big(\|(\widehat{\Sigma}_w - \Sigma_w)\beta^*\|_\infty \leq c\sigma_w^2\sqrt{\frac{\log p}{n}}\big) \geq 1 - c_1\exp(-c_2\log p)$$

by the same sub-Gaussian tail bounds. Since $\sigma_w^2 \leq \sigma_z^2$, we conclude that

$$\|(\widetilde{\Gamma} - \Sigma_x)\beta^*\|_\infty \leq c\sigma_z^2\sqrt{\frac{\log p}{n}}$$

with probability at least $1 - c_1\exp(-c_2\log p)$, as wanted.

Turning to the RE conditions, we similarly write

$$\begin{aligned}|v^T(\widetilde{\Gamma} - \Sigma_x)v| &\leq |v^T(\widetilde{\Gamma} - \widehat{\Gamma})v| + |v^T(\widehat{\Gamma} - \Sigma_x)v| \\ &= |v^T(\widehat{\Sigma}_w - \Sigma_w)v| + |v^T(\widehat{\Gamma} - \Sigma_x)v|.\end{aligned}$$

Then applying Lemma 15 to both terms, followed by Lemma 13, yields the required bounds.

APPENDIX B: PROOF OF COROLLARY 2

We now turn to the proof of Corollary 2, which applies to the case of missing data, based on the general M-estimator using the pair $(\widehat{\Gamma}_{\mathrm{mis}}, \widehat{\gamma}_{\mathrm{mis}})$ defined in equation (2.11). We will establish that the RE conditions and deviation conditions (3.2) hold with high probability.

LEMMA 3 (RE conditions, i.i.d. with missing data). *Under the conditions of Corollary 2, there are universal positive constants c_i such that $\widehat{\Gamma}_{mis}$ satisfies the lower- and upper-RE conditions with parameters $\alpha_1 = \frac{\lambda_{\min}(\Sigma_x)}{2}$, $\alpha_2 = \frac{3}{2}\lambda_{\max}(\Sigma_x)$, and*

$$\tau(n,p) = c_0\lambda_{\min}(\Sigma_x)\max\big(\frac{1}{(1-\rho_{\max})^2}\frac{\sigma_x^4}{\lambda_{\min}^2(\Sigma_x)}, 1\big)\frac{\log p}{n},$$

with probability at least $1 - c_1\exp\big(-c_2 n\min\big((1-\rho_{\max})^4\cdot\frac{\lambda_{\min}^2(\Sigma_x)}{\sigma_x^4}, 1\big)\big)$.

PROOF. This proof parallels the proof of Lemma 1 for the additive noise case. We make use of Lemma 13. This time, we have

$$\widehat{\Gamma} - \Sigma_x = \frac{Z^T Z}{n} \oplus M - \Sigma_x = \big(\frac{Z^T Z}{n} - \Sigma_z\big) \oplus M,$$

with the parameter s defined as in equation (A.1), with $\sigma^2 = \frac{\sigma_x^2}{(1-\rho_{\max})^2}$. Note that for a vector $v \in \mathbb{R}^p$, we have

$$\begin{aligned}
\big|v^T(\widehat{\Gamma} - \Sigma_x)v\big| &= \big|v^T\big(\big(\frac{Z^T Z}{n} - \Sigma_z\big) \oplus M\big)v\big| \\
&\leq \frac{1}{\|M\|_{\min}}\big|v^T\big(\frac{Z^T Z}{n} - \Sigma_z\big)v\big| \\
&\leq \frac{1}{(1-\rho_{\max})^2}\big|v^T\big(\frac{Z^T Z}{n} - \Sigma_z\big)v\big|. \qquad \text{(B.1)}
\end{aligned}$$

Furthermore, Z is a sub-Gaussian matrix with parameters (Σ_z, σ_x^2), so applying Lemma 15 in Appendix G with $t = (1-\rho_{\max})^2 \frac{\lambda_{\min}(\Sigma_x)}{54}$ to the right-hand expression, we obtain the bound

$$\mathbb{P}\big[D(s) \geq \frac{\lambda_{\min}(\Sigma_x)}{54}\big] \leq 2\exp\big(-c_2 n \min\big((1-\rho_{\max})^4 \cdot \frac{\lambda_{\min}^2(\Sigma_x)}{\sigma_x^4}, 1\big)\big).$$

□

LEMMA 4 (Deviation conditions, missing data). *Under the conditions of Corollary 2, there are universal positive constants c_i such the deviation bounds* (3.2) *hold with parameter*

$$\varphi(\mathbb{Q}, \sigma_\epsilon) = c_0 \frac{\sigma_x}{1-\rho_{\max}} \big(\sigma_\epsilon + \frac{\sigma_x}{1-\rho_{\max}}\big),$$

with probability at least $1 - c_1\exp\big(-c_2\log p\big)$.

PROOF. The key idea is to note that the observed matrix Z is a sub-Gaussian matrix with parameter σ_x^2. Indeed, recalling that the hidden matrix X is sub-Gaussian with parameter σ_x^2, we see that for any unit vector $v \in \mathbb{R}^p$, and any missing value pattern of X_i, we have

$$\mathbb{E}\big[\exp(\lambda Z v) \mid \text{missing values}\big] = \mathbb{E}(\exp(\lambda X_i u)) \leq \exp\big(\frac{\sigma_x^2\lambda^2}{2}\big), \qquad \text{(B.2)}$$

where the vector $u \in \mathbb{R}^p$ has entries $u_i = v_i$ when entry i is observed, and $u_i = 0$ otherwise. By the tower property of conditional expectation, it follows that the moment generating function of Zv is upper-bounded by the same quantity, so Z is also a sub-Gaussian matrix with parameter at most σ_x^2.

Observe that

$$\begin{aligned}
\|\widehat{\gamma} - \Sigma_x \beta^*\|_\infty &= \big\|\big(\frac{1}{n}\big(Z^T y - \operatorname{cov}(Z_i, y)\big) \oplus (\mathbf{1} - \boldsymbol{\rho})\big)\beta^*\big\| \\
&\le \frac{1}{1-\rho_{\max}} \big\|\frac{1}{n}\big(Z^T y - \operatorname{cov}(z_i, y)\big)\beta^*\big\|_\infty \\
&\le \frac{1}{1-\rho_{\max}} \big(\underbrace{\big\|\frac{1}{n}\big(Z^T X - \operatorname{cov}(z_i, x_i)\big)\beta^*\big\|_\infty}_{T_1} + \underbrace{\big\|\frac{Z^T \epsilon}{n}\big\|_\infty}_{T_2}\big). \qquad \text{(B.3)}
\end{aligned}$$

Using the sub-Gaussianity of the matrices X, Z, and ϵ, and Lemma 14, the two terms may be bounded as

$$\mathbb{P}\Big[T_1 \ge c_0 \frac{\sigma_x^2}{(1-\rho_{\max})^2}\sqrt{\frac{\log p}{n}}\Big] \le c_1 \exp(-c_2 \log p), \qquad \text{(B.4a)}$$

$$\mathbb{P}\Big[T_2 \ge c_0 \frac{\sigma_x \sigma_\epsilon}{(1-\rho_{\max})}\sqrt{\frac{\log p}{n}}\Big] \le c_1 \exp(-c_2 \log p). \qquad \text{(B.4b)}$$

Now consider the quantity $\|(\widehat{\Gamma} - \Sigma_x)\beta^*\|_\infty$. By a similar manipulation, we have

$$\|(\widehat{\Gamma} - \Sigma_x)\beta^*\|_\infty = \big\|\big(\big(\frac{Z^T Z}{n} - \Sigma_z\big) \oplus M\big)\beta^*\big\|_\infty \qquad \text{(B.5)}$$

$$\le \frac{1}{(1-\rho_{\max})^2}\big\|\big(\frac{Z^T Z}{n} - \Sigma_z\big)\beta^*\big\|_\infty, \qquad \text{(B.6)}$$

so using Lemma 14 yields

$$\big\|\big(\widehat{\Gamma} - \Sigma_x\big)\beta^*\big\|_\infty \le c_0 \frac{\sigma_x^2}{(1-\rho_{\max})^2}\sqrt{\frac{\log p}{n}}, \qquad \text{(B.7)}$$

with probability at least $1 - c_1 \exp(-c_2 \log p)$. Combining bounds (B.4) and (B.7), we conclude that the deviation conditions (3.2) both hold with parameter

$$\varphi(\mathbb{Q}, \sigma_\epsilon) = c_0 \, \frac{\sigma_x}{1-\rho_{\max}}\big(\frac{\sigma_x}{1-\rho_{\max}} + \sigma_\epsilon\big),$$

with probability at least $1 - c_1 \exp(-c_2 \log p)$, as claimed. □

Extension to unknown ρ_j. We now consider the more challenging case when the missing probabilities ρ_j are unknown. Note that the estimates $\widehat{\rho}_j$ satisfy the deviation bound

$$\mathbb{P}\big(\max_j |\widehat{\rho}_j - \rho_j| \ge t\big) \le c_1 \exp(-c_2 n t^2 + \log p), \qquad \text{(B.8)}$$

by a Hoeffding bound for Bernoulli random variables, together with a union bound. In particular, taking $t = c_0\sqrt{\frac{\log p}{n}}$, we have

$$\|\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_\infty \le c_0\sqrt{\frac{\log p}{n}}, \tag{B.9}$$

with probability at least $1 - c_1\exp(-c_2\log p)$.

As long as $n \succsim \frac{\log p}{(1-\rho_{\max})}$, as required by our results, the deviation condition (B.9) implies that $|\widehat{\rho}_j - \rho_j| \le \frac{1-\rho_{\max}}{2}$ for each j, so

$$(1-\widehat{\rho}_j) \ge (1-\rho_j) - \frac{1-\rho_{\max}}{2} \ge \frac{1}{2}(1-\rho_j). \tag{B.10}$$

In particular, we obtain the bound

(B.11)
$$\max_j\left|\frac{1-\rho_j}{1-\widehat{\rho}_j} - 1\right| \le \max_j \frac{|\rho_j - \widehat{\rho}_j|}{(1-\rho_{\max})\,(1-\widehat{\rho}_j)} \;\le\; \frac{2}{(1-\rho_{\max})^2}\;\max_j|\rho_j - \widehat{\rho}_j|,$$

and since $\left|\frac{1-\rho_j}{1-\widehat{\rho}_j} - 1\right| \le 1$ by inequality (B.10), we also have

$$\begin{aligned}
\text{(B.12)}\quad \max_{i,j}&\left|\frac{(1-\rho_i)(1-\rho_j)}{(1-\widehat{\rho}_i)(1-\widehat{\rho}_j)} - 1\right| \\
&= \left|\Big(\frac{1-\rho_i}{1-\widehat{\rho}_i} - 1\Big)\Big(\frac{1-\rho_j}{1-\widehat{\rho}_j} - 1\Big) + \Big(\frac{1-\rho_i}{1-\widehat{\rho}_i} - 1\Big) + \Big(\frac{1-\rho_j}{1-\widehat{\rho}_j} - 1\Big)\right| \\
&\le 3\max_j\left|\frac{1-\rho_j}{1-\widehat{\rho}_j} - 1\right| \\
\text{(B.13)}\quad &\le \frac{6}{(1-\rho_{\max})^2}\max_j|\rho_j - \widehat{\rho}_j|,
\end{aligned}$$

using the triangle inequality and inequality (B.11).

We will use these bounds to verify the results of Lemmas 3 and 4 for the estimators (3.9). For the deviation bounds, we begin by writing

$$\|(\widetilde{\Gamma} - \Sigma_x)\beta^*\|_\infty \le \|(\widetilde{\Gamma} - \widehat{\Gamma})\beta^*\|_\infty + \|(\widehat{\Gamma} - \Sigma_x)\beta^*\|_\infty.$$

Note that we have already bounded $\|(\widehat{\Gamma} - \Sigma_x)\beta^*\|_\infty$ in inequality (B.7).

Furthermore,

$$\begin{aligned}\|(\widetilde{\Gamma}-\widehat{\Gamma})\beta^*\|_\infty &= \Big\|\Big(\frac{Z^TZ}{n}\oplus\widetilde{M}-\frac{Z^TZ}{n}\oplus M\Big)\beta^*\|_\infty \\ &= \Big\|\Big(\frac{Z^TZ}{n}\oplus M\Big)\odot\big(M\oplus\widetilde{M}-\mathbf{1}\mathbf{1}^T\big)\beta^*\Big\|_\infty \\ &\le \big\|M\oplus\widetilde{M}-\mathbf{1}\mathbf{1}^T\big\|_{\max}\Big\|\Big(\frac{Z^TZ}{n}\oplus M\Big)\beta^*\Big\|_\infty \\ &\le \frac{2}{(1-\rho_{\max})^2}\max_j|\rho_j-\widehat{\rho}_j|\,\big(\|\widehat{\Gamma}-\Sigma_x)\beta^*\|_\infty+\|\Sigma_x\beta^*\|_\infty\big),\end{aligned}$$

where we have used inequality (B.11) and the triangle inequality in the last inequality above. Noting that $\|\Sigma_x\beta^*\|_\infty \le \lambda_{\max}(\Sigma_x)\|\beta^*\|_2 \le c\sigma_x^2$ and using the bounds (B.7) and (B.9), we obtain

$$\|(\widetilde{\Gamma}-\Sigma_x)\beta^*\|_\infty \le \frac{c\sigma_x^2}{(1-\rho_{\max})^2}\sqrt{\frac{\log p}{n}}.$$

Combining the pieces, we conclude that the deviation conditions (3.2) are satisfied with $\varphi(\mathbb{Q},\sigma_\epsilon)=c_0\frac{\sigma_x}{1-\rho_{\max}}\left(\frac{\sigma_x}{1-\rho_{\max}}+\sigma_\epsilon\right)$, as claimed.

For the RE conditions, we use a similar argument. Note that by Lemma 13, we need to show that $|v^T(\widetilde{\Gamma}-\Sigma_x)v| \le \frac{\lambda_{\min}(\Sigma_x)}{54}$ for all $v\in\mathbb{K}(2s)$, with high probability. We write

$$|v^T(\widetilde{\Gamma}-\Sigma_x)v| \le |v^T(\widetilde{\Gamma}-\widehat{\Gamma})v| + |v^T(\widehat{\Gamma}-\Sigma_x)v|,$$

and note that we have already shown how to upper-bound $|v^T(\widetilde{\Gamma}-\widehat{\Gamma})v|$ by $c\lambda_{\min}(\Sigma_x)$ with high probability. Furthermore,

$$\begin{aligned}\big|v^T(\widetilde{\Gamma}-\widehat{\Gamma})v\big| &= \Big|v^T\Big(\frac{Z^TZ}{n}\oplus\widetilde{M}-\frac{Z^TZ}{n}\oplus M\Big)v\Big| \\ &= \Big|v^T\Big(\frac{Z^TZ}{n}\oplus M\Big)\odot\big(\widetilde{M}\oplus M-\mathbf{1}\mathbf{1}^T\big)v\Big| \\ &\le \big\|\widetilde{M}\oplus M-\mathbf{1}\mathbf{1}^T\big\|_{\max}\Big|v^T\Big(\frac{Z^TZ}{n}\oplus M\Big)v\Big| \\ &\le \big\|\widetilde{M}\oplus M-\mathbf{1}\mathbf{1}^T\big\|_{\max}\big(|v^T(\widehat{\Gamma}-\Sigma_x)v|+|v^T\Sigma_xv|\big).\end{aligned}$$

Making use of inequality (B.12) and the concentration bound (B.8) with $t=c\frac{\lambda_{\min}(\Sigma_x)}{\lambda_{\max}(\Sigma_x)}(1-\rho_{\max})^2$, we obtain

$$\|\widetilde{M}\oplus M-\mathbf{1}\mathbf{1}^T\|_{\max}|v^T\Sigma_xv| \le c\lambda_{\min}(\Sigma_x)$$

with probability at least

$$1-c_1 \exp\left[-c_2 n(1-\rho_{\max})^4 \frac{\lambda_{\min}^2(\Sigma_x)}{\lambda_{\max}^2(\Sigma_x)}\right] \geq 1-c_1 \exp\left[-c_2 n(1-\rho_{\max})^4 \frac{\lambda_{\min}^2(\Sigma_x)}{\sigma_x^4}\right].$$

Note that $t \leq c'$, so the earlier upper bound on $|v^T(\widehat{\Gamma}-\Sigma_x)v|$ is sufficient to ensure that $|v^T(\widetilde{\Gamma}-\Sigma_x)v| \leq \frac{\lambda_{\min}(\Sigma_x)}{54}$ with the required probability.

APPENDIX C: PROOF OF COROLLARY 3

We now need to establish the RE conditions and deviation bounds (3.2) for the Gaussian VAR case, which we summarize in the following:

LEMMA 5 (RE conditions, dependent case with missing data). *Under the conditions of Corollary 3, there are universal positive constants c_i such that $\widehat{\Gamma}_{mis}$ satisfies the lower- and upper-RE conditions with $\alpha_1 = \frac{\lambda_{\min}(\Sigma_x)}{2}$, $\alpha_2 = \frac{3}{2}\lambda_{\max}(\Sigma_x)$, and*

$$\tau(n,p) = c_0 \lambda_{\min}(\Sigma_x) \max\Big(\frac{\zeta^4}{\lambda_{\min}^2(\Sigma_x)} 1\Big) \frac{\log p}{n},$$

with probability at least $1 - c_1 \exp\big(-c_2 n \min\big(\frac{\lambda_{\min}^2(\Sigma_x)}{\zeta^4}, 1\big)\big)$.

PROOF. The proof is identical to the proof of Lemma 1, except we use Lemma 18 instead of Lemma 15 in Appendix G.

□

LEMMA 6 (Deviation conditions, VAR with additive noise). *Under the conditions of Corollary 3, there are universal positive constants c_i such the deviation bounds* (3.2) *hold with parameter*

$$\varphi(\mathbb{Q}, \sigma_\epsilon) = c_0 \zeta\left(\zeta + \sigma_\epsilon\right),$$

with probability at least $1 - c_1 \exp\big(-c_2 \log p\big)$.

PROOF. We begin by bounding the term

$$\|(\widehat{\Gamma}-\Sigma_x)\beta^*\|_\infty = \max_{1\leq j\leq p} \Big|e_j^T\Big(\frac{1}{n}Z^T Z - \Sigma_z\Big)\beta^*\Big|.$$

Define the function $\Phi(u,v) := u^T\big(\frac{1}{n}Z^T Z - \Sigma_z\big)v$ and rewrite the term as $\max_{1\leq j\leq p}|\Phi(e_j,\beta^*)|$. For each fixed j, some simple algebra shows that

$$\Phi(e_j,\beta^*) = \frac{1}{2}\big\{\Phi(e_j+\beta^*, e_j+\beta^*) - \Phi(e_j,e_j) - \Phi(\beta^*,\beta^*)\big\}, \tag{C.1}$$

so it suffices to have a high-probability upper bound on the quantity $\Phi(v,v)$ for each fixed unit vector v. In particular, combining inequality (G.5) from Lemma 17 (see Appendix G) with the union bound and the relation (C.1), we conclude that

$$\mathbb{P}\big[\|(\widehat{\Gamma}-\Sigma_x)\beta^*\|_\infty \geq c_0\zeta^2\sqrt{\frac{\log p}{n}}\big] \leq c_1\exp(-c_2\log p). \tag{C.2}$$

We now turn to the quantity $\|\widehat{\gamma}-\Sigma_x\beta^*\|_\infty$, which by the triangle inequality may be upper-bounded as

$$\begin{aligned}\|\widehat{\gamma}-\Sigma_x\beta^*\|_\infty &\leq \big\|\big(\frac{1}{n}Z^TZ-\Sigma_z\big)\beta^*\big\|_\infty + \big\|\big(\frac{1}{n}W^TW-\Sigma_w\big)\beta^*\big\|_\infty \\ &\quad + \big\|\frac{1}{n}Z^T\epsilon\big\|_\infty + \big\|\frac{1}{n}X^TW\beta^*\big\|_\infty. \end{aligned} \tag{C.3}$$

We have already bounded the first term in inequality (C.2) above. As for the second term, the matrix W is sub-Gaussian, so that Lemma 14 can be used to control it (as in previous arguments). In order to upper-bound the third term on the RHS of equation (C.3), we first condition on Z. Under this conditioning, the third term may be written as $\max_{\ell=1,\ldots,p} v_\ell$, where $v_\ell := \frac{1}{n}\langle Ze_\ell,\, \epsilon\rangle$ is a zero-mean Gaussian variable with variance at most $\frac{\sigma_\epsilon^2}{n}\big(\frac{\|Ze_\ell\|_2}{\sqrt{n}}\big)^2$. Combining the union bound and the deviation bound (G.5) with $t=1$, we conclude that as long as $n \gtrsim \log p$, then

$$\mathbb{P}\big[\max_{\ell=1,\ldots,p}\big(\frac{\|Ze_\ell\|_2}{\sqrt{n}}\big)^2 \geq c\zeta^2\big] \leq c_1\exp\big(-c_2\log p\big).$$

Conditioning on this event and applying standard tail bounds to control $\{v_\ell\}$, we conclude that $\mathbb{P}\big[|v_\ell|\geq t\big] \leq \exp\big(-\frac{c_2}{\zeta^2\,\sigma_\epsilon^2}nt^2\big)$. Setting $t = c_0\sigma_\epsilon\zeta\sqrt{\frac{\log p}{n}}$ and then taking a union bound over $\ell \in \{1,\ldots,p\}$ yields the desired result. A similar analysis can be used to bound the fourth term, since the matrices X and W are independent. Combining the pieces yields the claim. ☐

APPENDIX D: PROOF OF COROLLARY 4

LEMMA 7 (RE conditions, dependent case with missing data). *Under the conditions of Corollary 4, there are universal positive constants c_i such that $\widehat{\Gamma}_{mis}$ satisfies the lower- and upper-RE conditions with $\alpha_1 = \frac{\lambda_{\min}(\Sigma_x)}{2}$, $\alpha_2 = \frac{3}{2}\lambda_{\max}(\Sigma_x)$, and*

$$\tau(n,p) = c_0\lambda_{\min}(\Sigma_x)\max\big(\frac{\zeta'^4}{\lambda_{\min}^2(\Sigma_x)},1\big)\frac{\log p}{n},$$

with probability at least $1-c_1\exp\big(-c_2 n\min\big(\frac{\lambda_{\min}^2(\Sigma_x)}{\zeta'^4},1\big)\big)$.

PROOF. Again, we simply substitute the bound of Lemma 18 for the bound of Lemma 15 in the proof of Lemma 3. ☐

The final step is verify the deviation bounds (3.2).

LEMMA 8 (Deviation conditions, VAR with missing data). *Under the conditions of Corollary 4, there are universal positive constants c_i such the deviation bounds* (3.2) *hold with parameter*

$$\varphi(\mathbb{Q}, \sigma_\epsilon) = c_0 \zeta' \left(\zeta' + \sigma_\epsilon\right),$$

with probability at least $1 - c_1 \exp(-c_2 \log p)$.

PROOF. To control the term $\|(\widehat{\Gamma} - \Sigma_x)\beta^*\|_\infty$, we use the same argument as in Lemma 6 to obtain inequality (C.2). For the term $\|\widehat{\gamma} - \Sigma_x \beta^*\|_\infty$, we use the expansion (B.3) from the i.i.d. case. We show how to bound the terms T_1 and T_2 appearing in the expansion.

For a vector $v \in \mathbb{R}^p$, write $\Psi(v) = \frac{\|v\|_2^2}{n} - \mathbb{E}\big(\frac{\|v\|_2^2}{n}\big)$, and note that

$$T_1 = \max_j \frac{1}{2}\big[\Psi(Ze_j + X\beta^*) - \Psi(Ze_j) - \Psi(X\beta^*)\big]. \tag{D.1}$$

By Lemma 17, we may upper-bound the last term in equation (D.1) by $C\zeta'^2(1-\rho_{\max})^2\sqrt{\frac{\log p}{n}}$, with probability at least $1 - c_1 \exp(-c_2 \log p)$. In order to bound the other two terms, we again use Lemma 17. Note that Ze_j is a mixture of Gaussians $N(0, Q_j)$, each with $|\!|\!|Q_j|\!|\!|_{\mathrm{op}} \le (1-\rho_{\max})^2 \zeta'^2$. Then $Ze_j + X\beta^*$ is also a mixture of Gaussians $N(0, Q_j')$, and we have the bound $|\!|\!|Q_i'|\!|\!|_{\mathrm{op}} \le 4\zeta'^2(1-\rho_{\max})^2$. Hence, by Lemma 17 and a union bound, we conclude that $T_1 \le c\zeta'^2(1-\rho_{\max})^2\sqrt{\frac{\log p}{n}}$, with probability at least $1 - c_1 \exp(-c_2 \log p)$. Turning to term T_2 in the expansion (B.3), we condition on Z. Repeating the argument in Lemma 6, we obtain the bound

$$T_2 \le c_0 \sigma_\epsilon \zeta'(1-\rho_{\max})\sqrt{\frac{\log p}{n}}.$$

Finally, plugging back into inequality (B.3), we arrive at the bound

$$\|\widehat{\gamma} - \Sigma_x \beta^*\|_\infty \le c(\zeta'^2(1-\rho_{\max}) + \sigma_\epsilon \zeta')\sqrt{\frac{\log p}{n}}.$$

Altogether, we have the form of φ given by $\varphi(\mathbb{Q}, \sigma_\epsilon) = c_0(\sigma_\epsilon \zeta' + \zeta'^2)$, as claimed. ☐

APPENDIX E: PROOF OF COROLLARY 5

First note that by Theorem 1, we have the bounds

$$\|\widehat{\theta}^j - \theta^j\|_1 \leq \frac{c\varphi(\mathbb{Q}, \sigma_\epsilon)}{\alpha_1} k\sqrt{\frac{\log p}{n}}, \tag{E.1}$$

$$\|\widehat{\theta}^j - \theta^j\|_2 \leq \frac{c\varphi(\mathbb{Q}, \sigma_\epsilon)}{\alpha_1} \sqrt{\frac{k \log p}{n}}. \tag{E.2}$$

We now establish the following lemma, which we will use to prove the theorem.

Lemma 9. *For each $1 \leq j \leq p$, we have*

$$\frac{1}{\lambda_{\max}(\Sigma)} \leq |a_j| \leq \frac{1}{\lambda_{\min}(\Sigma)} \quad \textit{and} \quad \|\theta^j\|_2 \leq \kappa(\Sigma). \tag{E.3}$$

Proof. Observe that $|a_j| \leq \max_j |\Theta_{jj}| \leq \lambda_{\max}(\Theta) = \frac{1}{\lambda_{\min}(\Sigma)}$, and similarly, $|a_j| \geq \frac{1}{\lambda_{\max}(\Sigma)}$, which establishes the first inequality (E.3). Next, note that the rows (and also columns) of Θ are bounded in ℓ_2-norm according to the inequality

$$\|\Theta_{j\cdot}\|_2 = \|\Theta e_j\|_2 \leq \lambda_{\max}(\Theta) = \frac{1}{\lambda_{\min}(\Sigma)},$$

which implies that $\|\theta^j\|_2 = \|\Theta_{\cdot j}/a_j\|_2 = \|\Theta_{j\cdot}\|_2/|a_j| \leq \kappa(\Sigma)$, as claimed. □

Moving forward, we establish the following deviation inequalities between a_j and $\Theta_{\cdot j}$ and their respective estimators.

Lemma 10. *For all j, we have the following deviation inequalities:*

$$|\widehat{a}_j - a_j| \leq \frac{c\kappa(\Sigma)}{\lambda_{\min}(\Sigma)} \Big(\frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\lambda_{\min}(\Sigma)} + \frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\alpha_1}\Big) \sqrt{\frac{k \log p}{n}}, \tag{E.4}$$

$$\|\widetilde{\Theta}_{\cdot j} - \Theta_{\cdot j}\|_1 \leq \frac{c\kappa^2(\Sigma)}{\lambda_{\min}(\Sigma)} \Big(\frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\lambda_{\min}(\Sigma)} + \frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\alpha_1}\Big) k\sqrt{\frac{\log p}{n}}. \tag{E.5}$$

Proof. We first derive inequality (E.4). Since the columns of Θ are k-sparse, $\|\theta^j\|_1 \leq \sqrt{k}\|\theta^j\|_2$. Then

$$\begin{aligned}
\|\widehat{\theta}^j\|_1 &\leq \|\theta^j\|_1 + \|\widehat{\theta}^j - \theta^j\|_1 \leq c\sqrt{k}\Big(\|\theta^j\|_2 + \frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\alpha_1}\sqrt{\frac{k \log p}{n}}\Big) \\
&\leq c\sqrt{k}\Big(\kappa(\Sigma) + \frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\alpha_1}\sqrt{\frac{\log p}{n}}\Big),
\end{aligned}$$

where we have used Lemma 9 and inequality (E.1). Under the assumed scaling $n \gtrsim k \log p$, this simplifies to the inequality

$$\|\widehat{\theta}^j\|_1 \le c\kappa(\Sigma)\sqrt{k}. \tag{E.6}$$

We now have

$$\begin{aligned} |\widehat{a}_j^{-1} - a_j^{-1}| &= \big|(\widehat{\Sigma}_{jj} - \widehat{\Sigma}_{j,-j}\widehat{\theta}^j) - (\Sigma_{jj} - \Sigma_{j,-j}\theta^j)\big| \\ &\le \underbrace{|\widehat{\Sigma}_{jj} - \Sigma_{jj}|}_{T_1} + \underbrace{|\widehat{\Sigma}_{j,-j}\widehat{\theta}^j - \Sigma_{j,-j}\theta^j|}_{T_2}. \end{aligned} \tag{E.7}$$

Using inequality (3.13), we have $T_1 \le c\varphi(\mathbb{Q}, \sigma_\epsilon)\sqrt{\frac{\log p}{n}}$. Furthermore,

$$\begin{aligned} T_2 &\le |(\widehat{\Sigma}_{j,-j} - \Sigma_{j,-j})\widehat{\theta}^j| + |\Sigma_{j,-j}(\widehat{\theta}^j - \theta^j)| \\ &\le \|\widehat{\Sigma} - \Sigma\|_{\max}\|\widehat{\theta}^j\|_1 + \|\Sigma_{j,-j}\|_2\|\widehat{\theta}^j - \theta^j\|_2 \\ &\le c\big(\varphi(\mathbb{Q}, \sigma_\epsilon)\kappa(\Sigma) + \lambda_{\max}(\Sigma_x)\frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\alpha_1}\big)\sqrt{\frac{k \log p}{n}}, \end{aligned}$$

using inequality (E.6) and inequality (E.2). Substituting back into inequality (E.7), we obtain

$$|\widehat{a}_j^{-1} - a_j^{-1}| \le c\big(\varphi(\mathbb{Q}, \sigma_\epsilon)\kappa(\Sigma) + \lambda_{\max}(\Sigma_x)\frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\alpha_1}\big)\sqrt{\frac{k \log p}{n}}$$

for all j. Hence,

$$\left|\frac{a_i}{\widehat{a}_i} - 1\right| = |a_i||\widehat{a}_i^{-1} - a_i^{-1}| \le c\kappa(\Sigma)\big(\frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\lambda_{\min}(\Sigma)} + \frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\alpha_1}\big)\sqrt{\frac{k \log p}{n}},$$

using Lemma 9, so $|\widehat{a}_j| \le 2a_j$ for $n \gtrsim k \log p$, and

$$|\widehat{a}_j - a_j| = |\widehat{a}_j|\left|\frac{a_j}{\widehat{a}_j} - 1\right| \le \frac{c\kappa(\Sigma)}{\lambda_{\min}(\Sigma)}\big(\frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\lambda_{\min}(\Sigma)} + \frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\alpha_1}\big)\sqrt{\frac{k \log p}{n}}, \tag{E.8}$$

which establishes inequality (E.4).

Turning to inequality (E.5), we have

$$\begin{aligned} \|\widetilde{\Theta}_{\cdot j} - \Theta_{\cdot j}\|_1 &= |\widehat{a}_j - a_j| + \|\widehat{a}_j\widehat{\theta}^j - a_j\theta^j\|_1 \\ &\le |\widehat{a}_j - a_j| + |a_j|\|\widehat{\theta}^j - \theta^j\|_1 + |\widehat{a}_j - a_j|\|\widehat{\theta}^j\|_1 \\ &\le c\big(\frac{1}{\lambda_{\min}(\Sigma)}\frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\alpha_1} + \frac{\kappa^2(\Sigma)}{\lambda_{\min}(\Sigma)}\big(\frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\lambda_{\min}(\Sigma)} + \frac{\varphi(\mathbb{Q}, \sigma_\epsilon)}{\alpha_1}\big)\big)k\sqrt{\frac{\log p}{n}}, \end{aligned}$$

by a combination of inequalities (E.1), (E.4), (E.6), and (E.8). Noting that $\kappa(\Sigma) > 1$, we arrive at inequality (E.5).

$\square$

Returning to the proof of Corollary 5, observe that since $\widehat{\Theta}$ and Θ are symmetric, we have $|||\widehat{\Theta} - \Theta|||_{\mathrm{op}} \leq |||\widehat{\Theta} - \Theta|||_1$. Furthermore, by the triangle inequality and the definition of $\widehat{\Theta}$,

$$|||\widehat{\Theta} - \Theta|||_1 \leq |||\widehat{\Theta} - \widetilde{\Theta}|||_1 + |||\widetilde{\Theta} - \Theta|||_1 \leq 2|||\widetilde{\Theta} - \Theta|||_1 = 2 \max_j \|\widetilde{\Theta}_{\cdot j} - \Theta_{\cdot j}\|_1,$$

so that the union bound and inequality (E.5) yield the claim.

APPENDIX F: RESTRICTED EIGENVALUE CONDITIONS

In this appendix, we provide the proofs for various lemmas used to establish restricted eigenvalue conditions for different classes of random matrices, depending on the observation model. We begin by establishing two auxiliary lemmas, and then proceed to the main lemma used directly in the proofs of the corollaries. Our first result shows how to bound the intersection of the ℓ_1-ball with the ℓ_2-ball in terms of a simpler set.

LEMMA 11. *For any constant $s \geq 1$, we have*

$$\mathbb{B}_1(\sqrt{s}) \cap \mathbb{B}_2(1) \subseteq 3 \ \mathrm{cl}\{\mathrm{conv}\left\{\mathbb{B}_0(s) \cap \mathbb{B}_2(1)\right\}\}, \tag{F.1}$$

where the balls are taken in p-dimensional space, and $\mathrm{cl}\{\cdot\}$ *and* $\mathrm{conv}\{\cdot\}$ *denote the topological closure and convex hull, respectively.*

PROOF. Note that when $s > p$, the containment is trivial, since the right-hand set equals $\mathbb{B}_2(3)$, and the left-hand set is contained in $\mathbb{B}_2(1)$. Hence, we will assume $1 \leq s \leq p$.

Let $A, B \subseteq \mathbb{R}^p$ be closed convex sets, with support function given by $\phi_A(z) = \sup_{\theta \in A} \langle \theta, z \rangle$ and ϕ_B similarly defined. It is a well-known fact that $\phi_A(z) \leq \phi_B(z)$ if and only if $A \subseteq B$ (cf. Theorem 2.3.1 of Hug and Weil [1]). We now check this condition for the pair of sets $A = \mathbb{B}_1(\sqrt{s}) \cap \mathbb{B}_2(1)$ and $B = 3 \ \mathrm{cl}\left\{ \mathrm{conv}\left\{\mathbb{B}_0(s) \cap \mathbb{B}_2(1)\right\}\right\}$.

For any $z \in \mathbb{R}^p$, let $S \subseteq \{1, 2, \ldots, p\}$ be the subset that indexes the top $\lfloor s \rfloor$ elements of z in absolute value. Then $\|z_{S^c}\|_\infty \leq |z_j|$ for all $j \in S$, whence

$$\|z_{S^c}\|_\infty \ \leq \ \frac{1}{\lfloor s \rfloor} \|z_S\|_1 \leq \frac{1}{\sqrt{\lfloor s \rfloor}} \|z_S\|_2. \tag{F.2}$$

We now split the supremum over A into two parts, corresponding to the elements indexed by S and its complement S^c, thereby obtaining

$$\begin{aligned}
\phi_A(z) \;=\; \sup_{\theta \in A} \langle \theta,\, z \rangle &\le \sup_{\|\theta_S\|_2 \le 1} \langle \theta_S,\, z_S \rangle + \sup_{\|\theta_{S^c}\|_1 \le \sqrt{s}} \langle \theta_{S^c},\, z_{S^c} \rangle \\
&\le \|z_S\|_2 + \sqrt{s} \|z_{S^c}\|_\infty \\
&\stackrel{(i)}{\le} \big(1 + \sqrt{\frac{s}{\lfloor s \rfloor}}\big) \|z_S\|_2 \\
&\le 3 \|z_S\|_2,
\end{aligned}$$

where step (i) makes use of inequality (F.2). Finally, we recognize that

$$\phi_B(z) \;=\; \sup_{\theta \in B} \langle \theta,\, z \rangle = 3 \max_{|U| = \lfloor s \rfloor} \sup_{\|\theta_U\|_2 \le 1} \langle \theta_U,\, z_U \rangle \;=\; 3 \|z_S\|_2,$$

from which the claim follows. $\square$

For ease of notation, define the sparse set $\mathbb{K}(s) := \mathbb{B}_0(s) \cap \mathbb{B}_2(1)$ and the cone set $\mathbb{C}(s) := \{v : \|v\|_1 \le \sqrt{s} \|v\|_2\}$. Our next result builds on Lemma 11, showing how to control deviations uniformly over vectors in $\mathbb{R}^p$.

LEMMA 12. *For a fixed matrix $\Gamma \in \mathbb{R}^{p \times p}$, parameter $s \ge 1$, and tolerance $\delta > 0$, suppose we have the deviation condition*

$$|v^T \Gamma v| \le \delta \qquad \forall v \in \mathbb{K}(2s). \tag{F.3}$$

Then

$$\big|v^T \Gamma v\big| \le 27\, \delta \big(\|v\|_2^2 + \frac{1}{s} \|v\|_1^2\big) \qquad \forall v \in \mathbb{R}^p. \tag{F.4}$$

PROOF. We begin by establishing the inequalities

$$|v^T \Gamma v| \le 27\, \delta\, \|v\|_2^2 \qquad \forall v \in \mathbb{C}(s), \tag{F.5a}$$

$$|v^T \Gamma v| \le \frac{27\, \delta}{s}\, \|v\|_1^2 \qquad \forall v \notin \mathbb{C}(s). \tag{F.5b}$$

Inequality (F.4) then follows immediately.

By rescaling, inequality (F.5a) follows if we can show that

$$|v^T \Gamma v| \le 27 \delta \quad \text{for all } v \text{ such that } \|v\|_2 = 1 \text{ and } \|v\|_1 \le \sqrt{s}. \tag{F.6}$$

By Lemma 11 and continuity, we further reduce the problem to proving the bound (F.6) for all vectors $v \in 3\, \mathrm{conv}\, \big\{\mathbb{K}(s)\big\} = \mathrm{conv}\, \big\{\mathbb{B}_0(s) \cap \mathbb{B}_2(3)\big\}$. Consider a weighted linear combination of the form $v = \sum_i \alpha_i v_i$, with weights

$\alpha_i \geq 0$ such that $\sum_i \alpha_i = 1$, and $\|v_i\|_0 \leq s$ and $\|v_i\|_2 \leq 3$ for each i. Expanding, we can write

$$v^T \Gamma v = \big(\sum \alpha_i v_i\big)^T \Gamma \big(\sum \alpha_i v_i\big) = \sum_{i,j} \alpha_i \alpha_j \big(v_i^T \Gamma v_j\big).$$

Applying inequality (F.3) to the vectors $\frac{1}{3} v_i$, $\frac{1}{3} v_j$, and $\frac{1}{6}(v_i + v_j)$, we have

$$|v_i^T \Gamma v_j| = \frac{1}{2} |(v_i + v_j)^T \Gamma (v_i + v_j) - v_i^T \Gamma v_i - v_j^T \Gamma v_j| \leq \frac{1}{2}(36\delta + 9\delta + 9\delta) = 27\delta$$

for all i, j, and hence $|v^T \Gamma v| \leq \sum_{i,j} \alpha_i \alpha_j (27\delta) = 27\delta \|\alpha\|_2^2 = 27\delta$, establishing inequality (F.5a).

Turning to inequality (F.5b), first note that for $v \notin \mathbb{C}(s)$, we have

$$\frac{\big|v^T \Gamma v\big|}{\|v\|_1^2} \leq \frac{1}{s} \sup_{\substack{\|u\|_1 \leq \sqrt{s} \\ \|u\|_2 \leq 1}} \big|u^T \Gamma u\big| \leq \frac{27\,\delta}{s}, \tag{F.7}$$

where the first inequality follows by the substitution $u = \sqrt{s} \frac{v}{\|v\|_1}$, and the second follows by the same argument used to establish inequality (F.5a), since u is in the set appearing in Lemma 11. Rearranging inequality (F.7) yields inequality (F.5b). □

LEMMA 13 (RE conditions). *Suppose $s \geq 1$ and $\widehat{\Gamma}$ is an estimator of Σ_x satisfying the deviation condition*

$$\big|v^T (\widehat{\Gamma} - \Sigma_x) v\big| \leq \frac{\lambda_{\min}(\Sigma_x)}{54} \qquad \forall v \in \mathbb{K}(2s).$$

Then we have the lower-RE condition

$$v^T \widehat{\Gamma} v \geq \frac{\lambda_{\min}(\Sigma_x)}{2} \|v\|_2^2 - \frac{\lambda_{\min}(\Sigma_x)}{2s} \|v\|_1^2 \tag{F.8}$$

and the upper-RE condition

$$v^T \widehat{\Gamma} v \leq \frac{3}{2} \lambda_{\max}(\Sigma_x) \|v\|_2^2 + \frac{\lambda_{\min}(\Sigma_x)}{2s} \|v\|_1^2. \tag{F.9}$$

PROOF. This result follows easily from Lemma 12. Setting $\Gamma = \widehat{\Gamma} - \Sigma_x$ and $\delta = \frac{\lambda_{\min}(\Sigma_x)}{54}$, we have the bound

$$\big|v^T (\widehat{\Gamma} - \Sigma_x) v\big| \leq \frac{\lambda_{\min}(\Sigma_x)}{2} \big(\|v\|_2^2 + \frac{1}{s} \|v\|_1^2\big).$$

Then

$$v^T\widehat{\Gamma}v \geq v^T\Sigma_x v - \frac{\lambda_{\min}(\Sigma_x)}{2}\big(\|v\|_2^2 + \frac{1}{s}\|v\|_1^2\big) \quad \text{and}$$
$$v^T\widehat{\Gamma}v \leq v^T\Sigma_x v + \frac{\lambda_{\min}(\Sigma_x)}{2}\big(\|v\|_2^2 + \frac{1}{s}\|v\|_1^2\big),$$

so the inequalities follow from $\lambda_{\min}(\Sigma_x)\|v\|_2^2 \leq v^T\Sigma_x v \leq \lambda_{\max}(\Sigma_x)\|v\|_2^2$. □

APPENDIX G: DEVIATION BOUNDS

In this appendix, we state and prove some deviation bounds for various types of random matrices.

G.1. Bounds in the i.i.d. setting. Given a zero-mean random variable Y, we refer to the quantity $\|Y\|_{\psi_1} := \sup_{\ell \geq 1} \ell^{-1}(\mathbb{E}|Y|^\ell)^{1/\ell}$ as its sub-exponential parameter. The finiteness of this quantity guarantees existence of all moments, and hence large-deviation bounds of the Bernstein type.

By Lemma 14 of Vershynin [5], if X is a zero-mean sub-Gaussian random variable with parameter σ, then the random variable $Y = X^2 - \mathbb{E}(X^2)$ is sub-exponential with $\|Y\|_{\psi_1} \leq 2\sigma^2$. It then follows that if $X_1, \ldots, X_n$ are zero-mean i.i.d. sub-Gaussian variables, we have the deviation inequality

$$\mathbb{P}\Big[\big|\frac{1}{n}\sum_{i=1}^n X_i^2 - \mathbb{E}[X_i^2]\big| \geq t\Big] \leq 2\exp\big(-c\min\big(\frac{nt^2}{4\sigma^4}, \frac{nt}{2\sigma^2}\big)\big) \quad \text{for all } t > 0,$$

where $c > 0$ is a universal constant (see Proposition 16 in Vershynin [5]). This deviation bound may be used to establish the following useful result:

LEMMA 14. *If $X \in \mathbb{R}^{n \times p_1}$ is a zero-mean sub-Gaussian matrix with parameters (Σ_x, σ_x^2), then for any fixed (unit) vector $v \in \mathbb{R}^{p_1}$, we have*

$$\mathbb{P}\big[\big|\|Xv\|_2^2 - \mathbb{E}\big[\|Xv\|_2^2\big]\big| \geq nt\big] \leq 2\exp\big(-cn\min\big(\frac{t^2}{\sigma_x^4}, \frac{t}{\sigma_x^2}\big)\big). \tag{G.1}$$

Moreover, if $Y \in \mathbb{R}^{n \times p_2}$ is a zero-mean sub-Gaussian matrix with parameters (Σ_y, σ_y^2), then

$$\mathbb{P}\big(\big\|\frac{Y^TX}{n} - \operatorname{cov}(y_i, x_i)\big\|_{\max} \geq t\big) \leq 6p_1p_2\exp\big(-cn\min\big(\frac{t^2}{(\sigma_x\sigma_y)^2}, \frac{t}{\sigma_x\sigma_y}\big)\big), \tag{G.2}$$

where X_i and Y_i are the i^{th} rows of X and Y, respectively. In particular, if $n \gtrsim \log p$, then

$$\mathbb{P}\big(\big\|\frac{Y^TX}{n} - \operatorname{cov}(y_i, x_i)\big\|_{\max} \geq c_0\sigma_x\,\sigma_y\sqrt{\frac{\log p}{n}}\big) \leq c_1\exp(-c_2\log p). \tag{G.3}$$

PROOF. Inequality (G.1) follows from the above discussion and the fact that Xv is a vector of i.i.d. sub-Gaussians with parameter σ. In order to prove inequality (G.2), we first note that if Z is a zero-mean sub-Gaussian variable with parameter σ_z, then the rescaled variable Z/σ_z is sub-Gaussian with parameter 1. Consequently, we may assume that $\sigma_x = \sigma_y = 1$ without loss of generality, rescaling as necessary. We then observe that

$$e_i^T \left\{ \frac{Y^T X}{n} - \operatorname{cov}(y_i, x_i) \right\} e_j = \frac{1}{2} \big[\Phi(Xe_j + Ye_i) - \Phi(Xe_j) - \Phi(Ye_i) \big],$$

where we have defined $\Phi(v) := \frac{\|v\|_2^2}{n} - \mathbb{E}\big(\frac{\|v\|_2^2}{n}\big)$. Since $Xe_j + Ye_i$ is sub-Gaussian with parameter at most 4, we may apply inequality (G.1) to each of the three terms, to obtain

$$\big| e_i^T \big(\frac{Y^T X}{n} - \operatorname{cov}(y_i, x_i) \big) e_j \big| \le \frac{3t}{2}$$

with probability at least $1 - 6 \exp\big(-cn \min\big\{t^2, t\big\}\big)$. Taking a union bound over all $1 \le i \le p_1$ and $1 \le j \le p_2$ yields inequality (G.2). Finally, setting $t = c_0\, \sigma_x\, \sigma_y \sqrt{\frac{\log p}{n}}$ and using the assumption $n \gtrsim \log p$ yields inequality (G.3). $\square$

We combine this lemma with a discretization argument and union bound to obtain the next result. For a parameter $s \ge 1$, recall the notation $\mathbb{K}(s) := \big\{ v \in \mathbb{R}^p \mid \|v\|_2 \le 1, \|v\|_0 \le s \big\}$.

LEMMA 15. *If $X \in \mathbb{R}^{n \times p}$ is a zero-mean sub-Gaussian matrix with parameters (Σ, σ^2), then there is a universal constant $c > 0$ such that*

(G.4)
$$\mathbb{P}\left[\sup_{v \in \mathbb{K}(2s)} \big| \frac{\|Xv\|_2^2}{n} - \mathbb{E}\big[\frac{\|Xv\|_2^2}{n}\big] \big| \ge t \right] \le 2 \exp\big(-cn \min\big(\frac{t^2}{\sigma^4}, \frac{t}{\sigma^2} \big) + 2s \log p \big).$$

PROOF. For each subset $U \subseteq \{1, \ldots, p\}$, we define the set S_U with $S_U = \{v \in \mathbb{R}^p : \|v\|_2 \le 1, \operatorname{supp}(v) \subseteq U\}$, and note that $\mathbb{K}(2s) = \bigcup_{|U| \le 2s} S_U$. If $\mathcal{A} = \{u_1, \ldots, u_m\}$ is a 1/3-cover of S_U, then for every $v \in S_U$, there is some $u_i \in \mathcal{A}$ such that $\|\Delta v\|_2 \le \frac{1}{3}$, where $\Delta v = v - u_i$. It is known [3] that we can construct $\mathcal{A}$ with $|\mathcal{A}| \le 9^{2s}$. If we define $\Phi(v_1, v_2) = v_1^T \big(\frac{X^T X}{n} - \Sigma \big) v_2$, we have

$$\sup_{v \in S_U} |\Phi(v, v)| \le \max_i |\Phi(u_i, u_i)| + 2 \sup_{v \in S_U} |\max_i \Phi(\Delta v, u_i)| + \sup_{v \in S_U} |\Phi(\Delta v, \Delta v)|.$$

Since $3\Delta v \in S_U$, it follows that

$$\sup_{v \in S_U} |\Phi(v,v)| \le \max_i |\Phi(u_i,u_i)| + \sup_{v \in S_U} \Big(\frac{2}{3}|\Phi(v,v)| + \frac{1}{9}|\Phi(v,v)|\Big),$$

hence $\sup_{v \in S_U} |\Phi(v,v)| \le \frac{9}{2} \max_i |\Phi(u_i,u_i)|$. By Lemma 14 and a union bound, we obtain

$$\mathbb{P}\Big(\sup_{v \in S_U} \Big|\frac{\|Xv\|_2^2}{n} - \mathbb{E}\Big(\frac{\|Xv\|_2^2}{n}\Big)\Big| \ge t\Big) \le 9^{2s} \cdot 2\exp\Big(-cn\min\Big(\frac{t^2}{\sigma^4}, \frac{t}{\sigma^2}\Big)\Big).$$

Finally, taking a union bound over the $\binom{p}{\lfloor 2s \rfloor} \le p^{2s}$ choices of U yields

$$\mathbb{P}\Big(\sup_{v \in \mathbb{K}(2s)} \Big|\frac{\|Xv\|_2^2}{n} - \mathbb{E}\Big(\frac{\|Xv\|_2^2}{n}\Big)\Big| \ge t\Big) \le 2\exp\Big(-cn\min\Big(\frac{t^2}{\sigma^4}, \frac{t}{\sigma^2}\Big) + 2s\log p\Big),$$

as claimed. □

G.2. Bounds for autoregressive processes. We base our analysis of Gaussian autoregressive matrices on the following lemma:

LEMMA 16. *Suppose $Y \in \mathbb{R}^m$ is a mixture of multivariate Gaussians $Y_j \sim N(0, Q_j)$, and let $\sigma^2 = \sup_j |||Q_j|||_{op}$. Then for all $t > \frac{2}{\sqrt{m}}$, we have*

$$\mathbb{P}\Big[\frac{1}{n}\big|\|Y\|_2^2 - \mathbb{E}(\|Y\|_2^2)\big| > 4t\sigma^2\Big] \le 2\exp\Big(-\frac{m\big(t - \frac{2}{\sqrt{m}}\big)^2}{2}\Big) + 2\exp(-m/2).$$

PROOF. This result is a generalization of Lemma I.2 in the paper [4]. By definition, the random vector Y is a mixture of random vectors of the form $\sqrt{Q_j}X_j$, where $X_j \sim N(0, I_m)$. For each index j, the function $f_j(x) = \|\sqrt{Q_j}x\|_2/\sqrt{m}$ is Lipschitz with constant $|||\sqrt{Q}|||_{\text{op}}/\sqrt{m}$. Since each X_j is Gaussian, it follows from the concentration for Lipschitz functions of Gaussians [2] that $f_j(X_j)$ is a sub-Gaussian random variable with parameter $\sigma_j^2 = |||Q_j|||_{\text{op}}/m$. Therefore, the mixture $\|Y\|_2/\sqrt{n}$ is sub-Gaussian with parameter $\sigma^2 = \frac{1}{m}\sup_j |||Q_j|||_{\text{op}}$. The remainder of the proof proceeds as in the paper [4].

□

We now specialize the preceding lemma to the cases of additive noise and missing data appearing in our paper.

LEMMA 17. *Let $X \in \mathbb{R}^{n \times p}$ be a Gaussian random matrix, with rows x_i generated according to a vector autoregression* (3.10) *with driving matrix A. Let $v \in \mathbb{R}^p$ be a fixed vector with unit norm. Then for all $t > \frac{2}{\sqrt{n}}$,*

$$\mathbb{P}\Big[\big|v^T\big(\widehat{\Gamma} - \Sigma_x\big)v\big| \geq 4t\zeta^2\Big] \leq 2\exp\Big(-\frac{n\big(t - \frac{2}{\sqrt{n}}\big)^2}{2}\Big) + 2\exp(-n/2), \tag{G.5}$$

where

$$\zeta^2 := \begin{cases} |||\Sigma_w|||_{op} + \frac{2|||\Sigma_x|||_{op}}{1-|||A|||_{op}} & \textit{(additive noise case)}, \\ \frac{1}{(1-\rho_{\max})^2}\frac{2|||\Sigma_x|||_{op}}{1-|||A|||_{op}} & \textit{(missing data case)}. \end{cases}$$

PROOF. First consider the additive noise case, where $\widehat{\Gamma} - \Sigma_x = \frac{Z^TZ}{n} - \Sigma_z$. For any fixed vector with $\|v\|_2 = 1$, the variable $Zv \in \mathbb{R}^n$ is a zero-mean Gaussian random variable with covariance matrix, say $Q \succeq 0$. In order to apply Lemma 16, we need to upper-bound the spectral norm of Q, which we do using the elementary upper bound $|||Q|||_{\mathrm{op}} \leq \max_{1 \leq i \leq n} \sum_{\ell=1}^n |Q_{i\ell}|$. For each pair $i, \ell \in \{1, 2, \ldots, n\}$, we have

$$|Q_{i\ell}| = |\operatorname{cov}(e_i^T Zv, e_\ell^T Zv)| = |v^T \operatorname{cov}(Z_i, Z_\ell)v|,$$

where Z_i and Z_ℓ are the i^{th} and ℓ^{th} rows of Z, and $\|v\|_2 = 1$. For $i \neq \ell$, we have

$$|v^T \operatorname{cov}(Z_i, Z_\ell)v| = |v^T \operatorname{cov}(X_i, X_\ell)v| = |v^T A^{|i-\ell|}\Sigma_x v| \leq |||\Sigma_x|||_{\mathrm{op}} |||A|||_{\mathrm{op}}^{|i-\ell|},$$

and for $i = \ell$, we have $|v^T \operatorname{cov}(Z^i, Z^i)v| \leq |||\Sigma_z|||_{\mathrm{op}} \leq |||\Sigma_w|||_{\mathrm{op}} + |||\Sigma_x|||_{\mathrm{op}}$. Putting together the pieces, we conclude that $|||Q|||_{\mathrm{op}} \leq \zeta^2$, with ζ as defined in the lemma statement. Consequently, the bound (G.5) follows from Lemma 16.

In the missing data case, the variable Zv is a zero-mean mixture of Gaussians, conditioned on the positions of the missing data. Suppose Z' is the random matrix Z corresponding to a given positioning scheme (with 0's in the missing positions). We claim that

$$|||Q_j|||_{\mathrm{op}} \leq \frac{2|||\Sigma_x|||_{\mathrm{op}}}{1 - |||A|||_{\mathrm{op}}}, \tag{G.6}$$

where $Q_j = \operatorname{Cov}(Z'v)$. Indeed, we write $|||Q_j|||_{\mathrm{op}} \leq \max_i \sum_{\ell=1}^n |Q_{j,i\ell}|$, and for each pair (i, ℓ),

$$|Q_{j,i\ell}| = |\operatorname{cov}(e_i^T Z'v, e_\ell^T Z'v)| = |\operatorname{cov}(Z'^i v, Z'^\ell v)| = |\operatorname{cov}(Z^i v_1, Z^\ell v_2)|,$$

where v_1 and v_2 are the vector v with 0's in the positions corresponding to the 0's of Z'^i and Z'^ℓ, respectively. Since $|\operatorname{cov}(Z^i v_1, Z^\ell v_2)| \le |||\Sigma_x|||\, |||A|||^{|i-\ell|}$ for $i \neq \ell$ and

$$|\operatorname{cov}(Z^i v_1, Z^i v_2)| \le |||\Sigma_w|||_{\mathrm{op}} + |||\Sigma_x|||_{\mathrm{op}}$$

by a similar argument as before, the claim (G.6) follows. By the bounding technique (B.1) earlier in the paper, together with Lemma 16, we arrive at inequality (G.5).

□

LEMMA 18. *Let X be a Gaussian matrix with rows generated from a vector autoregression with driving matrix A. Let $s \geq 1$. Then for all $t > \frac{2}{\sqrt{n}}$,*
(G.7)

$$\mathbb{P}\left[\sup_{v \in \mathbb{K}(2s)} \left|v^T\left(\widehat{\Gamma} - \Sigma_x\right)v\right| \geq 4t\zeta^2\right] \leq 4\exp\left(-cn\min\left(\left(t - \frac{2}{\sqrt{n}}\right)^2, 1\right) + 2s\log p\right),$$

with ζ as defined in Lemma 17.

PROOF. We use the single-deviation bounds from Lemma 17, together with a discretization argument identical to that of Lemma 15.

□

REFERENCES

[1] D. Hug and W. Weil. A course on convex geometry, 2010. Lecture notes available at www.math.kit.edu/iag4/lehre/convgeo2010w/media/cg.pdf.

[2] M. Ledoux. *The concentration of measure phenomenon.* Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2001.

[3] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes.* Springer-Verlag, New York, NY, 1991.

[4] S. Negahban and M. J. Wainwright. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Annals of Statistics*, 39(2):1069–1097, 2011.

[5] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *Compressed Sensing: Theory and Applications*, To appear. Available at http://www-personal.umich.edu/~romanv/papers/non-asymptotic-rmt-plain.pdf.

Department of Statistics
University of California, Berkeley
Berkeley, California 94720
USA
E-mail: ploh@berkeley.edu
wainwrig@stat.berkeley.edu