

**SUPPLEMENTARY MATERIAL FOR: ESTIMATION OF  
(NEAR) LOW-RANK MATRICES WITH NOISE AND  
HIGH-DIMENSIONAL SCALING**

BY SAHAND NEGAHBAN AND MARTIN J. WAINWRIGHT

*University of California, Berkeley*

APPENDIX A: INTRODUCTION

In this supplement we present many of the technical details from the main work [1]. Equation or theorem references made to the main document are relative to the numbering scheme of the document and will not contain letters.

APPENDIX B: PROOF OF LEMMA 1

Part (a) of the claim was proved in Recht et al. [2]; we simply provide a proof here for completeness. We write the SVD as  $\Theta^* = UDV^T$ , where  $U \in \mathbb{R}^{m_1 \times m_1}$  and  $V \in \mathbb{R}^{m_2 \times m_2}$  are orthogonal matrices, and  $D$  is the matrix formed by the singular values of  $\Theta^*$ . Note that the matrices  $U^r$  and  $V^r$  are given by the first  $r$  columns of  $U$  and  $V$  respectively. We then define the matrix  $\Gamma = U^T \Delta V \in \mathbb{R}^{m_1 \times m_2}$ , and write it in block form as

$$\Gamma = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}, \quad \text{where } \Gamma_{11} \in \mathbb{R}^{r \times r}, \text{ and } \Gamma_{22} \in \mathbb{R}^{(m_1-r) \times (m_2-r)}.$$

We now define the matrices

$$\Delta'' = U \begin{bmatrix} 0 & 0 \\ 0 & \Gamma_{22} \end{bmatrix} V^T, \quad \text{and } \Delta' = \Delta - \Delta''.$$

Note that we have

$$\text{rank}(\Delta') = \text{rank} \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & 0 \end{bmatrix} \leq \text{rank} \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ 0 & 0 \end{bmatrix} + \text{rank} \begin{bmatrix} \Gamma_{11} & 0 \\ \Gamma_{21} & 0 \end{bmatrix} \leq 2r,$$

which establishes Lemma 1(a). Moreover, we note for future reference that by construction of  $\Delta''$ , the nuclear norm satisfies the decomposition

$$(B.1) \quad \|\Pi_{\mathcal{A}^r}(\Theta^*) + \Delta''\|_1 = \|\Pi_{\mathcal{A}^r}(\Theta^*)\|_1 + \|\Delta''\|_1.$$

We now turn to the proof of Lemma 1(b). Recall that the error  $\Delta = \widehat{\Theta} - \Theta^*$  associated with any optimal solution must satisfy the inequality (30), which implies that

$$(B.2) \quad 0 \leq \frac{1}{N} \langle \vec{\varepsilon}, \mathfrak{X}(\Delta) \rangle + \lambda_N \{ \|\Theta^*\|_1 - \|\widehat{\Theta}\|_1 \} \leq \left\| \frac{1}{N} \mathfrak{X}^*(\vec{\varepsilon}) \right\|_{\text{op}} \|\Delta\|_1 + \lambda_N \{ \|\Theta^*\|_1 - \|\widehat{\Theta}\|_1 \},$$

where we have used the bound (31).

Note that we have the decomposition  $\Theta^* = \Pi_{\mathcal{A}^r}(\Theta^*) + \Pi_{\mathcal{B}^r}(\Theta^*)$ . Using this decomposition, the triangle inequality and the relation (B.1), we have

$$\begin{aligned} \|\widehat{\Theta}\|_1 &= \|(\Pi_{\mathcal{A}^r}(\Theta^*) + \Delta'') + (\Pi_{\mathcal{B}^r}(\Theta^*) + \Delta')\|_1 \\ &\geq \|(\Pi_{\mathcal{A}^r}(\Theta^*) + \Delta'')\|_1 - \|(\Pi_{\mathcal{B}^r}(\Theta^*) + \Delta')\|_1 \\ &\geq \|\Pi_{\mathcal{A}^r}(\Theta^*)\|_1 + \|\Delta''\|_1 - \{ \|\Pi_{\mathcal{B}^r}(\Theta^*)\|_1 + \|\Delta'\|_1 \}. \end{aligned}$$

Consequently, we have

$$\begin{aligned} \|\Theta^*\|_1 - \|\widehat{\Theta}\|_1 &\leq \|\Theta^*\|_1 - \{ \|\Pi_{\mathcal{A}^r}(\Theta^*)\|_1 + \|\Delta''\|_1 \} + \{ \|\Pi_{\mathcal{B}^r}(\Theta^*)\|_1 + \|\Delta'\|_1 \} \\ &= 2\|\Pi_{\mathcal{B}^r}(\Theta^*)\|_1 + \|\Delta'\|_1 - \|\Delta''\|_1. \end{aligned}$$

Substituting this inequality into the bound (B.2), we obtain

$$0 \leq \left\| \frac{1}{N} \mathfrak{X}^*(\vec{\varepsilon}) \right\|_{\text{op}} \|\Delta\|_1 + \lambda_N \{ 2\|\Pi_{\mathcal{B}^r}(\Theta^*)\|_1 + \|\Delta'\|_1 - \|\Delta''\|_1 \}.$$

Finally, since  $\left\| \frac{1}{N} \mathfrak{X}^*(\vec{\varepsilon}) \right\|_{\text{op}} \leq \lambda_N/2$  by assumption, we conclude that

$$0 \leq \lambda_N \{ 2\|\Pi_{\mathcal{B}^r}(\Theta^*)\|_1 + \frac{3}{2}\|\Delta'\|_1 - \frac{1}{2}\|\Delta''\|_1 \}.$$

Since  $\|\Pi_{\mathcal{B}^r}(\Theta^*)\|_1 = \sum_{j=r+1}^m \sigma_j(\Theta^*)$ , the bound (32) follows.

## APPENDIX C: PROOF OF COROLLARY 5

Recall that for this model, the observations are of the form  $y_i = \langle X_i, \Theta^* \rangle + \varepsilon_i$ , where  $\Theta^* \in \mathbb{R}^{m_1 \times m_2}$  is the unknown matrix, and  $\{\varepsilon_i\}_{i=1}^N$  is an associated noise sequence.

We now show how Proposition 1 implies the RSC property with an appropriate tolerance parameter  $\delta > 0$  to be defined. Observe that the bound (25) implies that for any  $\Delta \in \mathcal{C}$ , we have

$$(C.1) \quad \begin{aligned} \frac{\|\mathfrak{X}(\Delta)\|_2}{\sqrt{N}} &\geq \frac{\sqrt{\sigma_{\min}(\Sigma)}}{4} \|\Delta\|_F - 12\rho(\Sigma) \left( \sqrt{\frac{m_1}{N}} + \sqrt{\frac{m_2}{N}} \right) \|\Delta\|_1 \\ &= \frac{\sqrt{\sigma_{\min}(\Sigma)}}{4} \left\{ \|\Delta\|_F - \underbrace{\frac{48\rho(\Sigma)}{\sqrt{\sigma_{\min}(\Sigma)}} \left( \sqrt{\frac{m_1}{N}} + \sqrt{\frac{m_2}{N}} \right)}_{\tau} \|\Delta\|_1 \right\}, \end{aligned}$$

where we have defined the quantity  $\tau > 0$ . Following the arguments used in the proofs of Theorem 1 and Corollary 2, we find that

$$(C.2) \quad \|\Delta\|_1 \leq 4\|\Delta'\|_1 + 4 \sum_{j=r+1}^m \sigma_j(\Theta^*) \leq 4\sqrt{2R_q\tau^{-q}} \|\Delta'\|_F + 4R_q\tau^{1-q}.$$

Note that this corresponds to truncating the matrices at effective rank  $r = 2R_q\tau^{-q}$ . Combining this bound with the definition of  $\tau$ , we obtain

$$\tau \|\Delta\|_1 \leq 4\sqrt{2R_q\tau^{1-q/2}} \|\Delta'\|_F + 4R_q\tau^{2-q} \leq 4\sqrt{2R_q\tau^{1-q/2}} \|\Delta\|_F + 4R_q\tau^{2-q}.$$

Substituting this bound into equation (C.1) yields

$$\frac{\|\mathfrak{X}(\Delta)\|_2}{\sqrt{N}} \geq \frac{\sqrt{\sigma_{\min}(\Sigma)}}{4} \left\{ \|\Delta\|_F - 4\sqrt{2R_q\tau^{1-q/2}} \|\Delta'\|_F - 4R_q\tau^{2-q} \right\}.$$

As long  $N > c_0 R_q^{2/(2-q)} \frac{\rho^2(\Sigma)}{\sigma_{\min}(\Sigma)} (m_1 + m_2)$  for a sufficiently large constant  $c_0$ , we can ensure that  $4\sqrt{2R_q\tau^{1-q/2}} < 1/2$ , and hence that

$$\frac{\|\mathfrak{X}(\Delta)\|_2}{\sqrt{N}} \geq \frac{\sqrt{\sigma_{\min}(\Sigma)}}{4} \left\{ \frac{1}{2} \|\Delta\|_F - 4R_q\tau^{2-q} \right\}.$$

Consequently, if we define  $\delta := 16R_q\tau^{2-q}$ , then we are guaranteed that for all  $\|\Delta\|_F \geq \delta$ , we have  $4R_q\tau^{2-q} \leq \|\Delta\|_F/4$ , and hence

$$\frac{\|\mathfrak{X}(\Delta)\|_2}{\sqrt{N}} \geq \frac{\sqrt{\sigma_{\min}(\Sigma)}}{16} \|\Delta\|_F$$

for all  $\|\Delta\|_F \geq \delta$ . We have thus shown that  $\mathcal{C}(2R_q\tau^{-q}; \delta)$  with parameter  $\kappa(\mathfrak{X}) = \frac{\sigma_{\min}(\Sigma)}{256}$ .

The next step is to control the quantity  $\|\mathfrak{X}^*(\vec{\varepsilon})\|_{\text{op}}/N$ , required for specifying a suitable choice of  $\lambda_N$ .

LEMMA C.1. *If  $\|\vec{\varepsilon}\|_2 \leq 2\nu\sqrt{N}$ , then there are universal constants  $c_i$  such that*

$$(C.3) \quad \mathbb{P} \left[ \frac{\|\mathfrak{X}^*(\vec{\varepsilon})\|_{\text{op}}}{N} \geq c_0 \rho(\Sigma) \nu \left( \sqrt{\frac{m_1}{N}} + \sqrt{\frac{m_2}{N}} \right) \right] \leq c_1 \exp(-c_2(m_1 + m_2)).$$

PROOF. By the definition of the adjoint operator, we have  $Z = \frac{1}{N} \mathfrak{X}^*(\vec{\varepsilon}) = \frac{1}{N} \sum_{i=1}^N \varepsilon_i X_i$ . Since the observation matrices  $\{X_i\}_{i=1}^N$  are i.i.d. Gaussian, if the sequence  $\{\varepsilon_i\}_{i=1}^N$  is viewed as fixed (by conditioning as needed), then the random matrix  $Z$  is a sample from the  $\Gamma$ -ensemble with covariance matrix  $\Gamma = \frac{\|\vec{\varepsilon}\|^2}{N^2} \Sigma \preceq \frac{2\nu^2}{N} \Sigma$ . Therefore, letting  $\tilde{Z} \in \mathbb{R}^{m_1 \times m_2}$  be a random matrix drawn from the  $2\nu^2 \Sigma / N$ -ensemble, we have

$$\mathbb{P}[\|Z\|_{\text{op}} \geq t] \leq \mathbb{P}[\|\tilde{Z}\|_{\text{op}} \geq t].$$

Using Lemma H.1 from Appendix H, we have

$$\mathbb{E}[\|\tilde{Z}\|_{\text{op}}] \leq \frac{12\sqrt{2}\nu\rho(\Sigma)}{\sqrt{N}} (\sqrt{m_1} + \sqrt{m_2})$$

and

$$\mathbb{P}[\|\tilde{Z}\|_{\text{op}} \geq \mathbb{E}[\|\tilde{Z}\|_{\text{op}}] + t] \leq \exp\left(-c_1 \frac{Nt^2}{\nu^2 \rho^2(\Sigma)}\right)$$

for a universal constant  $c_1$ . Setting  $t^2 = \Omega\left(\frac{\nu^2 \rho^2(\Sigma)(\sqrt{m_1} + \sqrt{m_2})^2}{N}\right)$  yields the claim.  $\square$

#### APPENDIX D: PROOF OF COROLLARY 6

This corollary follows from a combination of Proposition 1 and Lemma 1. Let  $\hat{\Theta}$  be an optimal solution to the SDP (29), and let  $\Delta = \hat{\Theta} - \Theta^*$  be the error. Since  $\hat{\Theta}$  is optimal and  $\Theta^*$  is feasible for the SDP, we have  $\|\hat{\Theta}\|_1 = \|\Theta^* + \Delta\|_1 \leq \|\Theta^*\|_1$ . Using the decomposition  $\Delta = \Delta' + \Delta''$  from Lemma 1 and applying triangle inequality, we have

$$\|\Theta^* + \Delta' + \Delta''\|_1 \geq \|\Theta^* + \Delta''\|_1 - \|\Delta'\|_1.$$

From the properties of the decomposition in Lemma 1 (see Appendix B), we find that

$$\|\hat{\Theta}\|_1 = \|\Theta^* + \Delta' + \Delta''\|_1 \geq \|\Theta^*\|_1 + \|\Delta''\|_1 - \|\Delta'\|_1.$$

Combining the pieces yields that  $\|\Delta''\|_1 \leq \|\Delta'\|_1$ , and hence  $\|\Delta\|_1 \leq 2\|\Delta'\|_1$ . By Lemma 1(a), the rank of  $\Delta'$  is at most  $2r$ , so that we obtain  $\|\Delta\|_1 \leq 2\sqrt{2r}\|\Delta\|_F \leq 4\sqrt{r}\|\Delta\|_F$ .

Note that  $\mathfrak{X}(\Delta) = 0$ , since both  $\widehat{\Theta}$  and  $\Theta^*$  agree with the observations. Consequently, from Proposition 1, we have that

$$\begin{aligned} 0 &= \frac{\|\mathfrak{X}(\Delta)\|_2}{\sqrt{N}} \geq \frac{1}{4}\|\Delta\|_F - 12\rho(\Sigma) \left( \sqrt{\frac{m_1}{N}} + \sqrt{\frac{m_2}{N}} \right) \|\Delta\|_1 \\ &\geq \|\Delta\|_F \left( \frac{1}{4} - 12\rho(\Sigma) \sqrt{\frac{rm_1}{N}} + 12\rho(\Sigma) \sqrt{\frac{rm_2}{N}} \right) \\ &\geq \frac{1}{20}\|\Delta\|_F \end{aligned}$$

where the final inequality as long as  $N > c_0\rho^2(\Sigma)r(m_1+m_2)$  for a sufficiently large constant  $c_0$ . We have thus shown that  $\Delta = 0$ , which implies that  $\widehat{\Theta} = \Theta^*$  as claimed.

#### APPENDIX E: CONSISTENCY IN OPERATOR NORM

In this appendix, we derive a bound on the operator norm error for both the low-rank multivariate regression and auto-regressive model estimation problems. In this statement, it is convenient to specify these models in the form  $Y = X\Theta^* + W$ , where  $Y \in \mathbb{R}^{n \times m_2}$  is a matrix of observations.

**PROPOSITION E.1 (Operator norm consistency).** *Consider the multivariate regression problem and the SDP under the conditions of Corollary 3. Then any solution  $\widehat{\Theta}$  to the SDP satisfies the bound*

$$(E.1) \quad \|\widehat{\Theta} - \Theta^*\|_{\text{op}} \leq c' \frac{\nu \sqrt{\sigma_{\max}(\Sigma)}}{\sigma_{\min}(\Sigma)} \sqrt{\frac{m_1 + m_2}{n}}.$$

We note that a similar bound applies to the auto-regressive model treated in Corollary 4.

**PROOF.** For any subgradient matrix  $Z \in \partial\|\widehat{\Theta}\|_1$ , we are guaranteed  $\|Z\|_{\text{op}} \leq 1$ . Furthermore, by the KKT conditions [3] for the nuclear norm SDP, any solution  $\widehat{\Theta}$  must satisfy the condition

$$\frac{1}{n}X^T X \widehat{\Theta} - \frac{X^T Y}{n} + \lambda_n Z = 0.$$

Hence, simple algebra and the triangle inequality yield that

$$\|\widehat{\Theta}\|_{\text{op}} \leq \left\| \left( \frac{1}{n}X^T X \right)^{-1} \right\|_{\text{op}} \left[ \|X^T W/n\|_{\text{op}} + \lambda_n \right].$$

Lemma 2 yields that  $\|(\frac{1}{n}X^T X)^{-1}\|_{\text{op}} \leq \frac{9}{\sigma_{\min}(\Sigma)}$  with high probability. Combining these inequalities yields

$$\|\hat{\Theta}\|_{\text{op}} \leq c_1 \frac{\lambda_n}{\sigma_{\min}(\Sigma)}.$$

We require that  $\lambda_n \geq 2\|X^T W\|_{\text{op}}/n$ . From Lemma 3, it suffices to set  $\lambda_n \geq c_0 \sqrt{\sigma_{\max}(\Sigma)} \nu \sqrt{\frac{m_1+m_2}{n}}$ . Combining the pieces yields the claim.  $\square$

### APPENDIX F: PROOF OF LEMMA 3

Let  $S^{m-1} = \{u \in \mathbb{R}^m \mid \|u\|_2 = 1\}$  denote the Euclidean sphere in  $m$ -dimensions. The operator norm of interest has the variational representation

$$\frac{1}{n}\|X^T W\|_{\text{op}} = \frac{1}{n} \sup_{u \in S^{m_1-1}} \sup_{v \in S^{m_2-1}} v^T X^T W u$$

For positive scalars  $a$  and  $b$ , define the (random) quantity

$$\Psi(a, b) := \sup_{u \in a S^{m_1-1}} \sup_{v \in b S^{m_2-1}} \langle Xv, Wu \rangle.$$

and note that our goal is to upper bound  $\Psi(1, 1)$ . Note moreover that  $\Psi(a, b) = ab\Psi(1, 1)$ , a relation which will be useful in the analysis.

Let  $\mathcal{A} = \{u^1, \dots, u^A\}$  and  $\mathcal{B} = \{v^1, \dots, v^B\}$  denote  $1/4$  coverings of  $S^{m_1-1}$  and  $S^{m_2-1}$ , respectively. We now claim that we have the upper bound

$$(F.1) \quad \Psi(1, 1) \leq 4 \max_{u^a \in \mathcal{A}, v^b \in \mathcal{B}} \langle Xv^b, Wu^a \rangle$$

To establish this claim, we note that since the sets  $\mathcal{A}$  and  $\mathcal{B}$  are  $1/4$ -covers, for any pair  $(u, v) \in S^{m_1-1} \times S^{m_2-1}$ , there exists a pair  $(u^a, v^b) \in \mathcal{A} \times \mathcal{B}$  such that  $u = u^a + \Delta u$  and  $v = v^b + \Delta v$ , with  $\max\{\|\Delta u\|_2, \|\Delta v\|_2\} \leq 1/4$ . Consequently, we can write

$$(F.2) \quad \langle Xv, Wu \rangle = \langle Xv^b, Wu^a \rangle + \langle Xv^b, W\Delta u \rangle + \langle X\Delta v, Wu^a \rangle + \langle X\Delta v, W\Delta u \rangle.$$

By construction, we have the bound  $|\langle Xv^b, W\Delta u \rangle| \leq \Psi(1, 1/4) = \frac{1}{4}\Psi(1, 1)$ , and similarly  $|\langle X\Delta v, Wu^a \rangle| \leq \frac{1}{4}\Psi(1, 1)$  as well as  $|\langle X\Delta v, W\Delta u \rangle| \leq \frac{1}{16}\Psi(1, 1)$ . Substituting these bounds into the decomposition (F.2) and taking suprema over the left and right-hand sides, we conclude that

$$\Psi(1, 1) \leq \max_{u^a \in \mathcal{A}, v^b \in \mathcal{B}} \langle Xv^b, Wu^a \rangle + \frac{9}{16}\Psi(1, 1),$$

from which the bound (F.1) follows.

We now apply the union bound to control the discrete maximum. It is known (e.g., [4, 5]) that there exists a  $1/4$  covering of  $S^{m_1-1}$  and  $S^{m_2-1}$  with at most  $A \leq 8^{m_1}$  and  $B \leq 8^{m_2}$  elements respectively. Consequently, we have

$$(F.3) \quad \mathbb{P}[|\Psi(1, 1)| \geq 4\delta n] \leq 8^{m_1+m_2} \max_{u^a, v^b} \mathbb{P} \left[ \frac{|\langle Xv^b, Wu^a \rangle|}{n} \geq \delta \right].$$

It remains to obtain a good bound on the quantity  $\frac{1}{n} \langle Xv, Wu \rangle = \frac{1}{n} \sum_{i=1}^n \langle v, X_i \rangle \langle u, W_i \rangle$ , where  $(u, v) \in S^{m_1-1} \times S^{m_2-1}$  are arbitrary but fixed. Since  $W_i \in \mathbb{R}^{m_1}$  has i.i.d.  $N(0, \nu^2)$  elements and  $u$  is fixed, we have  $Z_i := \langle u, W_i \rangle \sim N(0, \nu^2)$  for each  $i = 1, \dots, n$ . These variables are independent of one another, and of the random matrix  $X$ . Therefore, conditioned on  $X$ , the sum  $Z := \frac{1}{n} \sum_{i=1}^n \langle v, X_i \rangle \langle u, W_i \rangle$  is zero-mean Gaussian with variance

$$\alpha^2 := \frac{\nu^2}{n} \left( \frac{1}{n} \|Xv\|_2^2 \right) \leq \frac{\nu^2}{n} \|X^T X/n\|_{\text{op}}.$$

Define the event  $\mathcal{T} = \{\alpha^2 \leq \frac{9\nu^2 \|\Sigma\|_{\text{op}}}{n}\}$ . Using Lemma 2, we have  $\|X^T X/n\|_{\text{op}} \leq 9\sigma_{\max}(\Sigma)$  with probability at least  $1 - 2\exp(-n/2)$ , which implies that  $\mathbb{P}[\mathcal{T}^c] \leq 2\exp(-n/2)$ . Therefore, conditioning on the event  $\mathcal{T}$  and its complement  $\mathcal{T}^c$ , we obtain

$$\begin{aligned} \mathbb{P}[|Z| \geq t] &\leq \mathbb{P}[|Z| \geq t \mid \mathcal{T}] + \mathbb{P}[\mathcal{T}^c] \\ &\leq \exp\left(-n \frac{t^2}{2\nu^2(4 + \|\Sigma\|_{\text{op}})}\right) + 2\exp(-n/2). \end{aligned}$$

Combining this tail bound with the upper bound (F.3), we have

$$\mathbb{P}[|\psi(1, 1)| \geq 4\delta n] \leq 8^{m_1+m_2} \left\{ \exp\left(-n \frac{t^2}{18\nu^2 \|\Sigma\|_{\text{op}}}\right) + 2\exp(-n/2) \right\}.$$

Setting  $t^2 = 20\nu^2 \|\Sigma\|_{\text{op}} \frac{m_1+m_2}{n}$ , this probability vanishes as long as  $n > 16(m_1 + m_2)$ .

#### APPENDIX G: TECHNICAL DETAILS FOR COROLLARY 4

In this appendix, we collect the proofs of Lemmas 4 and 5.

**G.1. Proof of Lemma 4.** Recalling that  $S^{m-1}$  denotes the unit-norm Euclidean sphere in  $m$ -dimensions, we first observe that  $\|X\|_{\text{op}} = \sup_{u \in S^{m-1}} \|Xu\|_2$ . Our next step is to reduce the supremum to a maximization over a finite set, using a standard covering argument. Let  $\mathcal{A} = \{u^1, \dots, u^A\}$  denote a  $1/2$ -cover of it. By definition, for any  $u \in S^{m-1}$ , there is some  $u^a \in \mathcal{A}$  such that  $u = u^a + \Delta u$ , where  $\|\Delta u\|_2 \leq 1/2$ . Consequently, for any  $u \in S^{m-1}$ , the triangle inequality implies that

$$\|Xu\|_2 \leq \|Xu^a\|_2 + \|X\Delta u\|_2,$$

and hence that  $\|X\|_{\text{op}} \leq \max_{u^a \in \mathcal{A}} \|Xu^a\|_2 + \frac{1}{2}\|X\|_{\text{op}}$ . Re-arranging yields the useful inequality

$$(G.1) \quad \|X\|_{\text{op}} \leq 2 \max_{u^a \in \mathcal{A}} \|Xu^a\|_2.$$

Using inequality (G.1), we have

$$(G.2) \quad \begin{aligned} \mathbb{P}\left[\frac{1}{n}\|X^T X\|_{\text{op}} > t\right] &\leq \mathbb{P}\left[\max_{u^a \in \mathcal{A}} \frac{1}{n} \sum_{i=1}^n (\langle u^a, X_i \rangle)^2 > \frac{t}{2}\right] \\ &\leq 4^m \max_{u^a \in \mathcal{A}} \mathbb{P}\left[\frac{1}{n} \sum_{i=1}^n (\langle u^a, X_i \rangle)^2 > \frac{t}{2}\right]. \end{aligned}$$

where the last inequality follows from the union bound, and the fact [4, 5] that there exists a  $1/2$ -covering of  $S^{m-1}$  with at most  $4^m$  elements.

In order to complete the proof, we need to obtain a sharp upper bound on the quantity  $\mathbb{P}[\frac{1}{n} \sum_{i=1}^n (\langle u, X_i \rangle)^2 > \frac{t}{2}]$ , valid for any fixed  $u \in S^{m-1}$ . Define the random vector  $Y \in \mathbb{R}^n$  with elements  $Y_i = \langle u, X_i \rangle$ . Note that  $Y$  is zero mean, and its covariance matrix  $R$  has elements  $R_{ij} = \mathbb{E}[Y_i Y_j] = u^T \Sigma (\Theta^*)^{|j-i|} u$ . In order to bound the spectral norm of  $R$ , we note that since it is symmetric, we have  $\|R\|_{\text{op}} \leq \max_{i=1, \dots, m} \sum_{j=1}^m |R_{ij}|$ , and moreover

$$|R_{ij}| = |u^T \Sigma (\Theta^*)^{|j-i|} u| \leq (\|\Theta^*\|_{\text{op}})^{|j-i|} \Sigma \leq \gamma^{|j-i|} \|\Sigma\|_{\text{op}}.$$

Combining the pieces, we obtain

$$(G.3) \quad \|R\|_{\text{op}} \leq \max_i \sum_{j=1}^m |\gamma|^{|i-j|} \|\Sigma\|_{\text{op}} \leq 2\|\Sigma\|_{\text{op}} \sum_{j=0}^{\infty} |\gamma|^j \leq \frac{2\|\Sigma\|_{\text{op}}}{1-\gamma}.$$

Moreover, we have  $\text{trace}(R)/n = u^T \Sigma u \leq \|\Sigma\|_{\text{op}}$ . Applying Lemma I.2 with  $t = 5\sqrt{\frac{m}{n}}$ , we conclude that

$$\mathbb{P}\left[\frac{1}{n}\|Y\|_2^2 > \|\Sigma\|_{\text{op}} + 5\sqrt{\frac{m}{n}}\|R\|_{\text{op}}\right] \leq 2 \exp(-5m) + 2 \exp(-n/2)..$$

Combined with the bound (G.2), we obtain

$$(G.4) \quad \left\| \frac{1}{n} X^T X \right\|_{\text{op}} \leq \|\Sigma\|_{\text{op}} \left\{ 2 + \frac{20}{(1-\gamma)} \sqrt{\frac{m}{n}} \right\} \leq \frac{24\|\Sigma\|_{\text{op}}}{(1-\gamma)},$$

with probability at least  $1 - c_1 \exp(-c_2 m)$ , which establishes the upper bound (35)(a).

Turning to the lower bound (35)(b), we let  $\mathcal{B} = \{v^1, \dots, v^B\}$  be an  $\epsilon$ -cover of  $S^{m-1}$  for some  $\epsilon \in (0, 1)$  to be chosen. Thus, for any  $v \in \mathbb{R}^m$ , there exists some  $v^b$  such that  $v = v^b + \Delta v$ , and  $\|\Delta v\|_2 \leq \epsilon$ . Define the function  $\Psi : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$  via  $\Psi(u, v) = u^T \left( \frac{1}{n} X^T X \right) v$ , and note that  $\Psi(u, v) = \Psi(v, u)$ . With this notation, we have

$$\begin{aligned} v^T \left( \frac{1}{n} X^T X \right) v &= \Psi(v, v) = \Psi(v^k, v^k) + 2\Psi(\Delta v, v) + \Psi(\Delta v, \Delta v) \\ &\geq \Psi(v^k, v^k) + 2\Psi(\Delta v, v), \end{aligned}$$

since  $\Psi(\Delta v, \Delta v) \geq 0$ . Since  $|\Psi(\Delta v, v)| \leq \epsilon \left\| \left( \frac{1}{n} X^T X \right) \right\|_{\text{op}}$ , we obtain the lower bound

$$\sigma_{\min} \left( \left( \frac{1}{n} X^T X \right) \right) = \inf_{v \in S^{m-1}} v^T \left( \frac{1}{n} X^T X \right) v \geq \min_{v^b \in \mathcal{B}} \Psi(v^b, v^b) - 2\epsilon \left\| \frac{1}{n} X^T X \right\|_{\text{op}}.$$

By the previously established upper bound (35)(a), have  $\left\| \frac{1}{n} X^T X \right\|_{\text{op}} \leq \frac{24\|\Sigma\|_{\text{op}}}{(1-\gamma)}$  with high probability. Hence, choosing  $\epsilon = \frac{(1-\gamma)\sigma_{\min}(\Sigma)}{200\|\Sigma\|_{\text{op}}}$  ensures that  $2\epsilon \left\| \frac{1}{n} X^T X \right\|_{\text{op}} \leq \sigma_{\min}(\Sigma)/4$ .

Consequently, it suffices to lower bound the minimum over the covering set. We first establish a concentration result for the function  $\Psi(v, v)$  that holds for any fixed  $v \in S^{m-1}$ . Note that we can write

$$\Psi(v, v) = \frac{1}{n} \sum_{i=1}^n (\langle v, X_i \rangle)^2,$$

As before, if we define the random vector  $Y \in \mathbb{R}^n$  with elements  $Y_i = \langle v, X_i \rangle$ , then  $Y \sim N(0, R)$  with  $\|R\|_{\text{op}} \leq \frac{2\|\Sigma\|_{\text{op}}}{1-\gamma}$ . Moreover, we have  $\text{trace}(R)/n = v^T \Sigma v \geq \sigma_{\min}(\Sigma)$ . Consequently, applying Lemma I.2 yields

$$\mathbb{P} \left[ \frac{1}{n} \|Y\|_2^2 < \sigma_{\min}(\Sigma) - \frac{8t\|\Sigma\|_{\text{op}}}{1-\gamma} \right] \leq 2 \exp \left( -n(t - 2/\sqrt{n})^2/2 \right) + 2 \exp \left( -\frac{n}{2} \right),$$

Note that this bound holds for any fixed  $v \in S^{m-1}$ . Setting  $t^* = \frac{(1-\gamma)\sigma_{\min}(\Sigma)}{16\|\Sigma\|_{\text{op}}}$  and applying the union bound yields that

$$\mathbb{P} \left[ \min_{v^b \in \mathcal{B}} \Psi(v^b, v^b) < \sigma_{\min}(\Sigma)/2 \right] \leq \left( \frac{4}{\epsilon} \right)^m \left\{ 2 \exp \left( -n(t^* - 2/\sqrt{n})^2/2 \right) + 2 \exp \left( -\frac{n}{2} \right) \right\},$$

which vanishes as long as  $n > \frac{4 \log(4/\epsilon)}{(t^*)^2} m$ .

**G.2. Proof of Lemma 5.** Let  $S^{m-1} = \{u \in \mathbb{R}^m \mid \|u\|_2 = 1\}$  denote the Euclidean sphere in  $m$ -dimensions, and for positive scalars  $a$  and  $b$ , define the random variable

$$\Psi(a, b) := \sup_{u \in a S^{m-1}} \sup_{v \in b S^{m-1}} \langle Xv, Wu \rangle.$$

Note that our goal is to upper bound  $\Psi(1, 1)$ . Let  $\mathcal{A} = \{u^1, \dots, u^A\}$  and  $\mathcal{B} = \{v^1, \dots, v^B\}$  denote  $1/4$  coverings of  $S^{m-1}$  and  $S^{m-1}$ , respectively. Following the same argument as in the proof of Lemma 3, we obtain the upper bound

$$(G.5) \quad \Psi(1, 1) \leq 4 \max_{u^a \in \mathcal{A}, v^b \in \mathcal{B}} \langle Xv^b, Wu^a \rangle$$

We now apply the union bound to control the discrete maximum. It is known (e.g., [4, 5]) that there exists a  $1/4$  covering of  $S^{m-1}$  with at most  $8^m$  elements. Consequently, we have

$$(G.6) \quad \mathbb{P}[|\psi(1, 1)| \geq 4\delta n] \leq 8^{2m} \max_{u^a, v^b} \mathbb{P}\left[\frac{|\langle Xv^b, Wu^a \rangle|}{n} \geq \delta\right].$$

It remains to obtain a tail bound on the quantity  $\mathbb{P}\left[\frac{|\langle Xv, Wu \rangle|}{n} \geq \delta\right]$ , for any fixed pair  $(u, v) \in \mathcal{A} \times \mathcal{B}$ .

For each  $i = 1, \dots, n$ , let  $X_i$  and  $W_i$  denote the  $i^{\text{th}}$  row of  $X$  and  $W$ . Following some simple algebra, we have the decomposition  $\frac{\langle Xv, Wu \rangle}{n} = T_1 - T_2 - T_3$ , where

$$\begin{aligned} T_1 &= \frac{1}{2n} \sum_{i=1}^n (\langle u, W_i \rangle + \langle v, X_i \rangle)^2 - \frac{1}{2}(u^T C u + v^T \Sigma v) \\ T_2 &= \frac{1}{2n} \sum_{i=1}^n (\langle u, W_i \rangle)^2 - \frac{1}{2} u^T C u \\ T_3 &= \frac{1}{2n} \sum_{i=1}^n (\langle v, X_i \rangle)^2 - \frac{1}{2} v^T \Sigma v \end{aligned}$$

We may now bound each  $T_j$  for  $j = 1, 2, 3$  in turn; in doing so, we make repeated use of Lemma I.2, which provides concentration bounds for a random variable of the form  $\|Y\|_2^2$ , where  $Y \sim N(0, Q)$  for some matrix  $Q \succeq 0$ .

*Bound on  $T_3$ :* We can write the term  $T_3$  as a deviation of  $\|Y\|_2^2/n$  from its mean, where in this case the covariance matrix  $Q$  is no longer the identity. In concrete terms, let us define a random vector  $Y \in \mathbb{R}^n$  with elements  $Y_i := \langle v, X_i \rangle$ . As seen in the proof of Lemma 4 from Appendix G.1, the vector  $Y$  is zero-mean Gaussian with covariance matrix  $R$  such that  $\|R\|_{\text{op}} \leq \frac{2\|\Sigma\|_{\text{op}}}{1-\gamma}$  (see equation (G.3)). Since we have  $\text{trace}(R)/n = v^T R v$ , applying Lemma I.2 yields that

$$(G.7) \quad \mathbb{P}[|T_3| \geq \frac{8\|\Sigma\|_{\text{op}}}{1-\gamma}t] \leq 2 \exp\left(-\frac{n(t - 2/\sqrt{n})^2}{2}\right) + 2 \exp(-n/2).$$

*Bound on  $T_2$ :* We control the term  $T_2$  in a similar way. Define the random vector  $Y' \in \mathbb{R}^n$  with elements  $Y'_i := \langle u, W_i \rangle$ . Then  $Y'$  is a sample from the distribution  $N(0, (u^T C u) I_{n \times n})$ , so that  $\frac{2}{u^T C u} T_2$  is the difference between a rescaled  $\chi^2$  variable and its mean. Applying Lemma I.2 with  $Q = (u^T C u) I$ , we obtain

$$(G.8) \quad \mathbb{P}[|T_2| > 4(u^T C u) t] \leq 2 \exp\left(-\frac{n(t - 2/\sqrt{n})^2}{2}\right) + 2 \exp(-n/2).$$

*Bound on  $T_1$ :* To control this quantity, let us define a zero-mean Gaussian random vector  $Z \in \mathbb{R}^n$  with elements  $Z_i = \langle v, X_i \rangle + \langle u, W_i \rangle$ . This random vector has covariance matrix  $S$  with elements

$$S_{ij} = \mathbb{E}[Z_i Z_j] = (u^T C u) \delta_{ij} + (1 - \delta_{ij})(u^T C u) v^T (\Theta^*)^{|i-j|-1} u + v^T (\Theta^*)^{|i-j|} \Sigma v,$$

where  $\delta_{ij}$  is the Kronecker delta for the event  $\{i = j\}$ . As before, by symmetry of  $S$ , we have  $\|S\|_{\text{op}} \leq \max_{i=1, \dots, n} \sum_{j=1}^n |S_{ij}|$ , and hence

$$\begin{aligned} \|S\|_{\text{op}} &\leq (u^T C u) + \|\Sigma\|_{\text{op}} + \sum_{j=1}^{i-1} |(u^T C u) v^T (\Theta^*)^{|i-j|-1} u + v^T (\Theta^*)^{|i-j|} \Sigma v| \\ &\quad + \sum_{j=i+1}^n |(u^T C u) v^T (\Theta^*)^{|i-j|-1} u + v^T (\Theta^*)^{|i-j|} \Sigma v|. \end{aligned}$$

Since  $\|\Theta^*\|_{\text{op}} \leq \gamma < 1$ , and  $(u^T C u) \leq \|C\|_{\text{op}} \leq \|\Sigma\|_{\text{op}}$ , we have

$$\begin{aligned} \|S\|_{\text{op}} &\leq \|C\|_{\text{op}} + \|\Sigma\|_{\text{op}} + 2 \sum_{j=1}^{\infty} \|C\|_{\text{op}} \gamma^{j-1} + 2 \sum_{j=1}^{\infty} \|\Sigma\|_{\text{op}} \gamma^j \\ &\leq 4 \|\Sigma\|_{\text{op}} \left(1 + \frac{1}{1-\gamma}\right) \end{aligned}$$

Moreover, we have  $\frac{\text{trace}(S)}{n} = (u^T C u) + v^T \Sigma v \leq 2\|\Sigma\|_{\text{op}}$ , so that by applying Lemma I.2, we conclude that

$$(G.9) \quad \mathbb{P}\left[|T_1| > \left(\frac{24\|\Sigma\|_{\text{op}}}{1-\gamma}\right)t\right] \leq 2\exp\left(-\frac{n(t - 2/\sqrt{n})^2}{2}\right) + 2\exp(-n/2),$$

which completes the analysis of this term.

Combining the bounds (G.7), (G.8) and (G.9), we conclude that for all  $t > 0$ ,

$$(G.10) \quad \mathbb{P}\left[\frac{|\langle Xv, Wu \rangle|}{n} \geq \frac{40(\|\Sigma\|_{\text{op}} t)}{1-\gamma}\right] \leq 6\exp\left(-\frac{n(t - 2/\sqrt{n})^2}{2}\right) + 6\exp(-n/2).$$

Setting  $t = 10\sqrt{m/n}$  and combining with the bound (G.6), we conclude that

$$\mathbb{P}[|\psi(1, 1)| \geq \frac{1600\|\Sigma\|_{\text{op}}}{1-\gamma}\sqrt{\frac{m}{n}}] \leq 8^{2m} \{6\exp(-16m) + 6\exp(-n/2)\} \leq 12\exp(-m)$$

as long as  $n > ((4 \log 8) + 1)m$ .

## APPENDIX H: PROOF OF PROPOSITION 1

We begin by stating and proving a useful lemma. Recall the definition (22) of  $\rho(\Sigma)$ .

LEMMA H.1. *Let  $X \in \mathbb{R}^{m_1 \times m_2}$  be a random sample from the  $\Sigma$ -ensemble. Then we have*

$$(H.1) \quad \mathbb{E}[\|X\|_{\text{op}}] \leq 12\rho(\Sigma) [\sqrt{m_1} + \sqrt{m_2}]$$

and moreover

$$(H.2) \quad \mathbb{P}[\|X\|_{\text{op}} \geq \mathbb{E}[\|X\|_{\text{op}}] + t] \leq \exp\left(-\frac{t^2}{2\rho^2(\Sigma)}\right).$$

PROOF. We begin by making note of the variational representation

$$\|X\|_{\text{op}} = \sup_{(u,v) \in S^{m_1-1} \times S^{m_2-1}} u^T X v.$$

Since each variable  $u^T X v$  is zero-mean Gaussian, we thus recognize  $\|X\|_{\text{op}}$  as the supremum of a Gaussian process. The bound (H.2) thus follows from Theorem 7.1 in Ledoux [6].

We now use a simple covering argument establish the upper bound (H.1). Let  $\{v^1, \dots, v^{M_2}\}$  be a  $1/4$  covering of the sphere  $S^{m_2-1}$ . For an arbitrary  $v \in S^{m_2-1}$ , there exists some  $v^j$  in the cover such that  $\|v - v^j\|_2 \leq 1/4$ , whence

$$\|Xv\|_2 \leq \|Xv^j\|_2 + \|X(v - v^j)\|_2.$$

Taking suprema over both sides, we obtain that  $\|X\|_{\text{op}} \leq \max_{j=1, \dots, M_2} \|Xv^j\|_2 + \frac{1}{4}\|X\|_{\text{op}}$ . A similar argument using a  $1/4$ -covering  $\{u^1, \dots, u^{M_1}\}$  of  $S^{m_1-1}$  yields that

$$\|Xv^j\|_2 \leq \max_{i=1, \dots, M_1} \langle u^i, Xv^j \rangle + \frac{1}{4}\|X\|_{\text{op}}.$$

Combining the pieces, we conclude that

$$\|X\|_{\text{op}} \leq 2 \max_{\substack{i=1, \dots, M_1 \\ j=1, \dots, M_2}} \langle u^i, Xv^j \rangle.$$

By construction, each variable  $\langle u^i, Xv^j \rangle$  is zero-mean Gaussian with variance at most  $\rho(\Sigma)$ , so that by standard bounds on Gaussian maxima, we obtain

$$\mathbb{E}[\|X\|_{\text{op}}] \leq 4\rho(\Sigma)\sqrt{\log(M_1M_2)} \leq 4\rho(\Sigma)[\sqrt{\log M_1} + \sqrt{\log M_2}].$$

There exist  $1/4$ -coverings of  $S^{m_1-1}$  and  $S^{m_2-1}$  with  $\log M_1 \leq m_1 \log 8$  and  $\log M_2 \leq m_2 \log 8$ , from which the bound (H.1) follows.  $\square$

We now return to the proof of Proposition 1. To simplify the proof, let us define an operator  $T_\Sigma : \mathbb{R}^{m_1 \times m_2} \rightarrow \mathbb{R}^{m_1 \times m_2}$  such that  $\text{vec}(T_\Sigma(\Theta)) = \sqrt{\Sigma} \text{vec}(\Theta)$ . Let  $\mathfrak{X}' : \mathbb{R}^{m_1 \times m_2} \rightarrow \mathbb{R}^N$  be a random Gaussian operator formed with  $X'_i$  sampled with i.i.d.  $N(0, 1)$  entries. By construction, we then have  $\mathfrak{X}(\Theta) = \mathfrak{X}'(T_\Sigma(\Theta))$  for all  $\Theta \in \mathbb{R}^{m_1 \times m_2}$ . Now by the variational characterization of the  $\ell_2$ -norm, we have

$$\|\mathfrak{X}'(T_\Sigma(\Theta))\|_2 = \sup_{u \in S^{N-1}} \langle u, \mathfrak{X}'(T_\Sigma(\Theta)) \rangle.$$

Since the original claim (25) is invariant to rescaling, it suffices to prove it for matrices such that  $\|T_\Sigma(\Theta)\|_F = 1$ . Letting  $t \geq 1$  be a given radius, we seek lower bounds on the quantity

$$Z^*(t) := \inf_{\Theta \in \mathcal{R}(t)} \sup_{u \in S^{N-1}} \langle u, \mathfrak{X}'(T_\Sigma(\Theta)) \rangle,$$

$$\text{where } \mathcal{R}(t) = \{\Theta \in \mathbb{R}^{m_1 \times m_2} \mid \|T_\Sigma(\Theta)\|_F = 1, \|\Theta\|_1 \leq t\}.$$

In particular, our goal is to prove that for any  $t \geq 1$ , the lower bound

$$(H.3) \quad \frac{Z^*(t)}{\sqrt{N}} \geq \frac{1}{4} - 12 \rho(\Sigma) \left[ \frac{m_1 + m_2}{N} \right]^{1/2} t$$

holds with probability at least  $1 - c_1 \exp(-c_2 N)$ . By a standard peeling argument (see Raskutti et al. [7] for details), this lower bound implies the claim (25).

We establish the lower bound (H.3) using Gaussian comparison inequalities [4] and concentration of measure (see Lemma I.1). For each pair  $(u, \Theta) \in S^{N-1} \times \mathcal{R}(t)$ , consider the random variable  $Z_{u,\Theta} = \langle u, \mathfrak{X}'(T_\Sigma(\Theta)) \rangle$ , and note that it is Gaussian with zero mean. For any two pairs  $(u, \Theta)$  and  $(u', \Theta')$ , some calculation yields

$$(H.4) \quad \mathbb{E}[(Z_{u,\Theta} - Z_{u',\Theta'})^2] = \|u \otimes T_\Sigma(\Theta) - u' \otimes T_\Sigma(\Theta')\|_F^2.$$

We now define a second Gaussian process  $\{Y_{u,\Theta} \mid (u, \Theta) \in S^{N-1} \times \mathcal{R}(t)\}$  via

$$Y_{u,\Theta} := \langle g, u \rangle + \langle\langle G, T_\Sigma(\Theta) \rangle\rangle,$$

where  $g \in \mathbb{R}^N$  and  $G \in \mathbb{R}^{m_1 \times m_2}$  are independent with i.i.d.  $N(0, 1)$  entries. By construction,  $Y_{u,\Theta}$  is zero-mean, and moreover, for any two pairs  $(u, \Theta)$  and  $(u', \Theta')$ , we have

$$(H.5) \quad \mathbb{E}[(Y_{u,\Theta} - Y_{u',\Theta'})^2] = \|u - u'\|_2^2 + \|T_\Sigma(\Theta) - T_\Sigma(\Theta')\|_F^2.$$

For all pairs  $(u, \Theta), (u', \Theta') \in S^{N-1} \times \mathcal{R}(t)$ , we have  $\|u\|_2 = \|u'\|_2 = 1$ , and moreover  $\|T_\Sigma(\Theta)\|_F = \|T_\Sigma(\Theta')\|_F = 1$ . Using this fact, some algebra yields that

$$(H.6) \quad \|u \otimes T_\Sigma(\Theta) - u' \otimes T_\Sigma(\Theta')\|_F^2 \leq \|u - u'\|_2^2 + \|T_\Sigma(\Theta) - T_\Sigma(\Theta')\|_F^2.$$

Moreover, equality holds whenever  $\Theta = \Theta'$ . The conditions of the Gordon-Slepian inequality [4] are satisfied, so that we are guaranteed that

$$(H.7) \quad \mathbb{E} \left[ \inf_{\Theta \in \mathcal{R}(t)} \|\mathfrak{X}'(T_\Sigma(\Theta))\|_2 \right] = \mathbb{E} \left[ \inf_{\Theta \in \mathcal{R}(t)} \sup_{u \in S^{N-1}} Z_{u,\Theta} \right] \geq \mathbb{E} \left[ \inf_{\Theta \in \mathcal{R}(t)} \sup_{u \in S^{N-1}} Y_{u,\Theta} \right]$$

We compute

$$\begin{aligned} \mathbb{E} \left[ \inf_{\Theta \in \mathcal{R}(t)} \sup_{u \in S^{N-1}} Y_{u,\Theta} \right] &= \mathbb{E} \left[ \sup_{u \in S^{N-1}} \langle g, u \rangle \right] + \mathbb{E} \left[ \inf_{\Theta \in \mathcal{R}(t)} \langle\langle G, T_\Sigma(\Theta) \rangle\rangle \right] \\ &= \mathbb{E}[\|g\|_2] - \mathbb{E} \left[ \sup_{\Theta \in \mathcal{R}(t)} \langle\langle G, T_\Sigma(\Theta) \rangle\rangle \right] \\ &\geq \frac{1}{2} \sqrt{N} - t \mathbb{E}[\|T_\Sigma(G)\|_{\text{op}}], \end{aligned}$$

where we have used the fact that  $T_\Sigma$  is self-adjoint, and Hölder's inequality (involving the operator and nuclear norms). Since  $T_\Sigma(G)$  is a random matrix from the  $\Sigma$ -ensemble, Lemma H.1 yields the upper bound  $\mathbb{E}[\|T_\Sigma(G)\|_{\text{op}}] \leq 12\rho(\Sigma)(\sqrt{m_1} + \sqrt{m_2})$ . Putting together the pieces, we conclude that

$$\mathbb{E}\left[\inf_{\Theta \in \mathcal{R}(t)} \frac{\|\mathfrak{X}'(T_\Sigma(\Theta))\|_2}{\sqrt{N}}\right] \geq \frac{1}{2} - 12\rho(\Sigma) \left(\frac{\sqrt{m_1} + \sqrt{m_2}}{\sqrt{N}}\right) t.$$

Finally, we need to establish sharp concentration around the mean. Since  $\|T_\Sigma(\Theta)\|_F = 1$  for all  $\Theta \in \mathcal{R}(t)$ , the function  $f(\mathfrak{X}) := \inf_{\Theta \in \mathcal{R}(t)} \|\mathfrak{X}'(T_\Sigma(\Theta))\|_2/\sqrt{N}$  is Lipschitz with constant  $1/\sqrt{N}$ , so that Lemma I.1 implies that

$$\mathbb{P}\left[\inf_{\Theta \in \mathcal{R}(t)} \frac{\|\mathfrak{X}(\Theta)\|_2}{\sqrt{N}} \leq \frac{1}{2} - 12\rho(\Sigma) \left(\frac{\sqrt{m_1} + \sqrt{m_2}}{\sqrt{N}}\right) t - \delta\right] \leq 2\exp(-N\delta^2/2)$$

for all  $\delta > 0$ . Setting  $\delta = 1/4$  yields the claim.

#### APPENDIX I: SOME USEFUL CONCENTRATION RESULTS

The following lemma is classical [4, 8], and yields sharp concentration of a Lipschitz function of Gaussian random variables around its mean.

LEMMA I.1. *Let  $X \in \mathbb{R}^n$  have i.i.d.  $N(0, 1)$  entries, and let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be Lipschitz with constant  $L$  (i.e.,  $|f(x) - f(y)| \leq L\|x - y\|_2 \forall x, y \in \mathbb{R}^n$ ). Then for all  $t > 0$ , we have*

$$\mathbb{P}[|f(X) - Ef(X)| > t] \leq 2\exp\left(-\frac{t^2}{2L^2}\right).$$

By exploiting this lemma, we can prove the following result, which yields concentration of the squared  $\ell_2$ -norm of an arbitrary Gaussian vector:

LEMMA I.2. *Given a Gaussian random vector  $Y \sim N(0, Q)$ , for all  $t > 2/\sqrt{n}$ , we have*

$$(I.1) \quad \mathbb{P}\left[\frac{1}{n}|\|Y\|_2^2 - \text{trace } Q| > 4t\|Q\|_{\text{op}}\right] \leq 2\exp\left(-\frac{n(t - \frac{2}{\sqrt{n}})^2}{2}\right) + 2\exp(-n/2).$$

PROOF. Let  $\sqrt{Q}$  be the symmetric matrix square root, and consider the function  $f(x) = \|\sqrt{Q}x\|_2/\sqrt{n}$ . Since it is Lipschitz with constant  $\|\sqrt{Q}\|_{\text{op}}/\sqrt{n}$ , Lemma I.1 implies that

$$(I.2) \quad \mathbb{P}[|\|\sqrt{Q}X\|_2 - E\|\sqrt{Q}X\|_2| > \sqrt{n}\delta] \leq 2\exp\left(-\frac{n\delta^2}{2\|Q\|_{\text{op}}}\right) \quad \text{for all } \delta > 0.$$

By integrating this tail bound, we find that the variable  $Z = \|\sqrt{Q}X\|_2/\sqrt{n}$  satisfies the bound  $\text{var}(Z) \leq 4\|Q\|_{\text{op}}/n$ , and hence conclude that

(I.3)

$$|\sqrt{\mathbb{E}[Z^2]} - |\mathbb{E}[Z]|| = |\sqrt{\text{trace}(Q)/n} - \mathbb{E}[\|\sqrt{Q}X\|_2/\sqrt{n}]| \leq \frac{2\sqrt{\|Q\|_{\text{op}}}}{\sqrt{n}}.$$

Combining this bound with the tail bound (I.2), we conclude that

(I.4)

$$\mathbb{P}\left[\frac{1}{\sqrt{n}}|\|\sqrt{Q}X\|_2 - \sqrt{\text{trace}(Q)}| > \delta + 2\sqrt{\frac{\|Q\|_{\text{op}}}{n}}\right] \leq 2\exp\left(-\frac{n\delta^2}{2\|Q\|_{\text{op}}}\right) \quad \text{for all } \delta > 0.$$

Setting  $\delta = (t - 2/\sqrt{n})\sqrt{\|Q\|_{\text{op}}}$  in the bound (I.4) yields that

(I.5)

$$\mathbb{P}\left[\frac{1}{\sqrt{n}}|\|\sqrt{Q}X\|_2 - \sqrt{\text{trace}(Q)}| > t\sqrt{\|Q\|_{\text{op}}}\right] \leq 2\exp\left(-\frac{n(t - 2/\sqrt{n})^2}{2}\right).$$

Similarly, setting  $\delta = \sqrt{\|Q\|_{\text{op}}}$  in the tail bound (I.4) yields that with probability greater than  $1 - 2\exp(-n/2)$ , we have

$$(I.6) \quad \left|\frac{\|Y\|_2}{\sqrt{n}} + \sqrt{\frac{\text{trace}(Q)}{n}}\right| \leq \sqrt{\frac{\text{trace}(Q)}{n}} + 3\sqrt{\|Q\|_{\text{op}}} \leq 4\sqrt{\|Q\|_{\text{op}}}.$$

Using these two bounds, we obtain

$$\left|\frac{\|Y\|_2^2}{n} - \frac{\text{trace}(Q)}{n}\right| = \left|\frac{\|Y\|_2}{\sqrt{n}} - \sqrt{\frac{\text{trace}(Q)}{n}}\right| \left|\frac{\|Y\|_2}{\sqrt{n}} + \sqrt{\frac{\text{trace}(Q)}{n}}\right| \leq 4t\|Q\|_{\text{op}}$$

with the claimed probability.  $\square$

## REFERENCES

- [1] S. Negahban and M. J. Wainwright. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. Technical Report <http://arxiv.org/abs/0912.5100>, UC Berkeley, Department of Statistics, December 2009.
- [2] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 2007. to appear.
- [3] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, UK, 2004.
- [4] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Springer-Verlag, New York, NY, 1991.

- [5] J. Matousek. *Lectures on discrete geometry*. Springer-Verlag, New York, 2002.
- [6] M. Ledoux. *The Concentration of Measure Phenomenon*. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2001.
- [7] G. Raskutti, M. J. Wainwright, and B. Yu. Minimax rates of estimation for high-dimensional linear regression over  $\ell_q$ -balls. Technical Report arXiv:0910.2042, UC Berkeley, Department of Statistics, 2009. Presented in part at Allerton Conference, Sep. 2009.
- [8] P. Massart. *Concentration Inequalities and Model Selection*. Ecole d'Été de Probabilités, Saint-Flour. Springer, New York, 2003.

DEPARTMENT OF ELECTRICAL ENGINEERING  
UNIVERSITY OF CALIFORNIA, BERKELEY  
BERKELEY, CA 94720  
USA  
E-MAIL: sahand\_n@eecs.berkeley.edu

DEPARTMENT OF STATISTICS  
UNIVERSITY OF CALIFORNIA, BERKELEY  
BERKELEY, CALIFORNIA 94720  
USA  
E-MAIL: wainwrig@eecs.berkeley.edu