Biometrika (2020), **xx**, x, *pp*. 1–20
Printed in Great Britain

Supplementary material for 'Finite time analysis of vector autoregressive models under linear restrictions'

BY YAO ZHENG
Department of Statistics, University of Connecticut, Storrs, Connecticut 06269, U.S.A.
yao.zheng@uconn.edu

GUANG CHENG
Department of Statistics, Purdue University West Lafayette, Indiana 47907, U.S.A.
chengg@purdue.edu

SUMMARY

This supplementary material contains all technical proofs of the main paper. § S1 gives the proofs of Theorem 1 and Proposition 1, which rely on three auxiliary lemmas, Lemmas S1–S3, whose proofs are relegated to § S1.4. § S2 contains the proofs of Lemmas 1–3. In § S3, we first verify equation (15) in the main paper and then prove Proposition 2 through four auxiliary lemmas, Lemmas S5–S8. § S4 contains the proof of Theorem 3. Lastly § S5 proves Theorem 4 and Corollary 1 after introducing two auxiliary lemmas, Lemmas S9 and S10.

S1. PROOFS OF THEOREM 1 AND PROPOSITION 1

S1.1. *Three Auxiliary Lemmas*

The proofs of Theorem 1 and Proposition 1 rely on three auxiliary lemmas, Lemmas S1–S3. Lemma S1 contains key results on covering and discretization. Lemma S2 gives a poinwise lower bound of $X^{\mathrm{T}}X$ through aggregation of all the $\lfloor n/k \rfloor$ blocks of size k using the Chernoff bound. Notice that the probability guarantee in Lemma S2 will degrade as k increases, since the probability guarantee of the Chernoff bound will degrade as the number of blocks decreases. Lemma S3 is a multivariate concentration bound for dependent data, respectively. We state these lemmas first and relegate their proofs to § S1.4.

The following notations will be used throughout our proofs: For any integer $d \geq 1$ and matrix $0 \prec \Gamma \in \mathbb{R}^{d\times d}$, let $\|\Gamma^{1/2}(\cdot)\|$ be the ellipsoidal vector norm associated to Γ, i.e., the mapping from $\omega \in \mathbb{R}^d$ to $(\omega^{\mathrm{T}}\Gamma\omega)^{1/2} \in (0,\infty)$. In addition, we denote the corresponding unit ball, or ellipsoid, by $\mathcal{S}_\Gamma = \{\omega \in \mathbb{R}^d : \|\Gamma^{1/2}\omega\| = 1\}$. For any set $\mathcal{S}$, we denote its cardinality, complement and volume by $|\mathcal{S}|$, $\mathcal{S}^c$ and $\mathrm{vol}(\mathcal{S})$, respectively.

LEMMA S1. *Suppose that $Z \in \mathbb{R}^{n\times m}$ and $0 \prec \Gamma_{\min} \preceq \Gamma_{\max} \in \mathbb{R}^{m\times m}$. Let $\mathcal{T}$ be a $1/4$-net of $\mathcal{S}_{\Gamma_{\min}}$ in the norm $\|\Gamma_{\max}^{1/2}(\cdot)\|$. Then, the following holds:*
(i) If $\Gamma_{\min}/2 \npreceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}$, then $\inf_{\omega\in\mathcal{T}} \omega^{\mathrm{T}}Z^{\mathrm{T}}Z\omega < 1$.
(ii) If $\mathcal{T}$ is a minimal $1/4$-net, then $\log|\mathcal{T}| \leq m\log 9 + (1/2)\log\det(\Gamma_{\max}\Gamma_{\min}^{-1})$.

(iii) If $\Gamma_{\min} \preceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}$, then for any $\nu \in \mathbb{R}^n$, we have

$$\sup_{\omega \in \mathcal{S}^{m-1}} \frac{\omega^{\mathrm{T}} Z^{\mathrm{T}} \nu}{\|Z\omega\|} = \sup_{\omega \in \mathcal{S}_{\Gamma_{\min}}} \frac{\omega^{\mathrm{T}} Z^{\mathrm{T}} \nu}{\|Z\omega\|} \leq 2 \max_{\omega \in \mathcal{T}} \frac{\omega^{\mathrm{T}} Z^{\mathrm{T}} \nu}{\|Z\omega\|}.$$

LEMMA S2. *Suppose that the process $\{X_t\}_{t=1}^n$ taking values in $\mathbb{R}^d$ satisfies the $(k, \Gamma_{\mathrm{sb}}, \alpha)$-BMSB condition. Let $X = (X_1, \ldots, X_n)^{\mathrm{T}}$. Then, for any $\omega \in \mathbb{R}^d$, we have*

$$\mathrm{pr}\left(\omega^{\mathrm{T}} X^{\mathrm{T}} X \omega \leq \frac{\alpha^2 k \lfloor n/k \rfloor}{8} \omega^{\mathrm{T}} \Gamma_{\mathrm{sb}} \omega\right) \leq \exp\left(-\frac{\alpha^2 \lfloor n/k \rfloor}{8}\right).$$

LEMMA S3. *Let $\{\mathcal{F}_t, t = 1, 2, \ldots\}$ be a filtration. Suppose that $\{x_t, t = 1, 2, \ldots\}$ and $\{\eta_t, t = 1, 2, \ldots\}$ are processes taking values in $\mathbb{R}^q$, and for each integer $t \geq 1$, x_t is $\mathcal{F}_t$-measurable, η_t is $\mathcal{F}_{t+1}$-measurable, and $\eta_t \mid \mathcal{F}_t$ is mean-zero and σ^2-sub-Gaussian. Then, for any constants $\beta_-, \beta_+, \gamma > 0$, we have*

$$\mathrm{pr}\left\{\frac{\sum_{t=1}^n x_t^{\mathrm{T}} \eta_t}{(\sum_{t=1}^n \|x_t\|^2)^{1/2}} \geq \gamma,\ \sum_{t=1}^n \|x_t\|^2 \in [\beta_-, \beta_+]\right\} \leq \frac{\beta_+}{\beta_-} \exp\left(-\frac{\gamma^2}{6\sigma^2}\right). \quad \text{(S1)}$$

S1.2. *Proof of Theorem* 1

Define the $m \times m$ matrices

$$\Gamma_{\max} = n\overline{\Gamma}_R, \quad \Gamma_{\min} = \frac{\alpha^2 k \lfloor n/k \rfloor}{8} \underline{\Gamma}_R, \quad \underline{\Gamma}_{\min} = \Gamma_{\min}/2, \quad \text{(S2)}$$

where $\overline{\Gamma}_R = R^{\mathrm{T}}(I_q \otimes \overline{\Gamma})R$ and $\underline{\Gamma}_R = R^{\mathrm{T}}(I_q \otimes \Gamma_{\mathrm{sb}})R$. Since R has full column rank, $\overline{\Gamma}_R$ and $\underline{\Gamma}_R$ are both positive definite matrices. Thus, $0 \prec \underline{\Gamma}_{\min} \prec \Gamma_{\min} \preceq \Gamma_{\max}$.

Consider the singular value decomposition $Z = \mathcal{U}\mathcal{D}\mathcal{V}^{\mathrm{T}}$, where $\mathcal{U} \in \mathbb{R}^{qn \times m}$, $\mathcal{D}, \mathcal{V} \in \mathbb{R}^{m \times m}$, and $\mathcal{U}^{\mathrm{T}}\mathcal{U} = I_m = \mathcal{V}^{\mathrm{T}}\mathcal{V}$. Let $Z^\dagger$ be the Moore-Penrose pseudoinverse of Z, i.e., $Z^\dagger = \mathcal{V}\mathcal{D}^-\mathcal{U}^{\mathrm{T}}$, where the diagonal matrix $\mathcal{D}^-$ is defined by taking the reciprocal of each nonzero diagonal entry of $\mathcal{D}$; in particular, $Z^\dagger = (Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}$ if $Z^{\mathrm{T}}Z \succ 0$. Then, we have $\widehat{\theta} - \theta_* = Z^\dagger \eta$. As a result,

$$\widehat{\beta} - \beta_* = R(\widehat{\theta} - \theta_*) = RZ^\dagger \eta = R\mathcal{V}\mathcal{D}^-\mathcal{U}^{\mathrm{T}}\eta.$$

Furthermore, since $\underline{\Gamma}_{\min} \succ 0$, it holds on the event $\{Z^{\mathrm{T}}Z \succeq \underline{\Gamma}_{\min}\}$ that

$$\begin{aligned}\|\widehat{\beta} - \beta_*\| \leq \|R\mathcal{V}\mathcal{D}^-\|_2 \|\mathcal{U}^{\mathrm{T}}\eta\| &= \left[\lambda_{\max}\{R(Z^{\mathrm{T}}Z)^{-1}R^{\mathrm{T}}\}\right]^{1/2} \|\mathcal{U}^{\mathrm{T}}\eta\| \\ &\leq \left\{\lambda_{\max}(R\underline{\Gamma}_{\min}^{-1}R^{\mathrm{T}})\right\}^{1/2} \|\mathcal{U}^{\mathrm{T}}\eta\|. \quad \text{(S3)}\end{aligned}$$

Note that (S3) exploits the self-cancellation effect inside the pseudoinverse $Z^\dagger$: the bound would not be as sharp if R, $Z^{\mathrm{T}}Z$ and $Z\eta$ were bounded separately.

By (S3) and Assumption 2, i.e., $\mathrm{pr}(Z^{\mathrm{T}}Z \npreceq \Gamma_{\max}) \leq \delta$, for any $K > 0$, we have

$$\begin{aligned}&\mathrm{pr}\left[\|\widehat{\beta} - \beta_*\| \geq K\left\{\lambda_{\max}(R\underline{\Gamma}_{\min}^{-1}R^{\mathrm{T}})\right\}^{1/2}\right] \\ &\leq \mathrm{pr}\left[\|\widehat{\beta} - \beta_*\| \geq K\left\{\lambda_{\max}(R\underline{\Gamma}_{\min}^{-1}R^{\mathrm{T}})\right\}^{1/2},\ Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right] + \delta \\ &\leq \mathrm{pr}\left[\|\widehat{\beta} - \beta_*\| \geq K\left\{\lambda_{\max}(R\underline{\Gamma}_{\min}^{-1}R^{\mathrm{T}})\right\}^{1/2},\ \underline{\Gamma}_{\min} \preceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right] \\ &\quad + \mathrm{pr}\left(\underline{\Gamma}_{\min} \npreceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right) + \delta \\ &\leq \mathrm{pr}\left(\|\mathcal{U}^{\mathrm{T}}\eta\| \geq K,\ \underline{\Gamma}_{\min} \preceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right) + \mathrm{pr}\left(\underline{\Gamma}_{\min} \npreceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right) + \delta. \quad \text{(S4)}\end{aligned}$$

Notice that condition (5) implies $k \leq n/10$, so that

$$k\lfloor n/k \rfloor \geq n - k \geq (9/10)n. \tag{S5}$$

As a result,

$$\left\{\lambda_{\max}(R\underline{\Gamma}_{\min}^{-1}R^{\mathrm{T}})\right\}^{1/2} \leq \frac{9}{2\alpha}\left\{\frac{\lambda_{\max}(R\underline{\Gamma}_R^{-1}R^{\mathrm{T}})}{n}\right\}^{1/2}. \tag{S6}$$

In view of (S4) and (S6), to prove this theorem, it remains to show that $Z^{\mathrm{T}}Z$ is bounded below and $\|\mathcal{U}^{\mathrm{T}}\eta\|$ is bounded above, with high probability. Specifically, we will prove that

$$\mathrm{pr}\left(\underline{\Gamma}_{\min} \npreceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right) \leq \delta \tag{S7}$$

if condition (5) of the theorem holds, and

$$\mathrm{pr}\left(\|\mathcal{U}^{\mathrm{T}}\eta\| \geq K,\ \underline{\Gamma}_{\min} \preceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right) \leq \delta \tag{S8}$$

if we choose

$$K = 2\sigma\left\{12m\log(14/\alpha) + 9\log\det(\overline{\Gamma}_R\underline{\Gamma}_R^{-1}) + 6\log(1/\delta)\right\}^{1/2}.$$

Proof of (S7): Let $\mathcal{T}$ be a minimal $1/4$-net of $\mathcal{S}_{\Gamma_{\min}}$ in the norm $\|\Gamma_{\max}^{1/2}(\cdot)\|$. By Lemma S1(i), we have

$$\mathrm{pr}\left(\Gamma_{\min}/2 \npreceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right) \leq \mathrm{pr}\left(\inf_{\omega\in\mathcal{T}}\omega^{\mathrm{T}}Z^{\mathrm{T}}Z\omega < 1\right) \leq |\mathcal{T}| \sup_{\omega\in\mathcal{S}_{\Gamma_{\min}}}\mathrm{pr}\left(\omega^{\mathrm{T}}Z^{\mathrm{T}}Z\omega < 1\right). \tag{S9}$$

By Lemma S1(ii) and (S5), we have

$$\begin{aligned}\log|\mathcal{T}| &\leq m\log 9 + (1/2)\log\det(\Gamma_{\max}\Gamma_{\min}^{-1})\\ &= m\log 9 + (1/2)m\log\frac{8n}{k\lfloor n/k\rfloor\alpha^2} + (1/2)\log\det(\overline{\Gamma}_R\underline{\Gamma}_R^{-1})\\ &\leq m\log(27/\alpha) + (1/2)\log\det(\overline{\Gamma}_R\underline{\Gamma}_R^{-1}).\end{aligned} \tag{S10}$$

Note that $Z^{\mathrm{T}}Z = R^{\mathrm{T}}(I_q\otimes X^{\mathrm{T}}X)R = \sum_{i=1}^q R_i^{\mathrm{T}}X^{\mathrm{T}}XR_i$, where each R_i is a $d\times m$ block in $R = (R_1^{\mathrm{T}},\ldots,R_q^{\mathrm{T}})^{\mathrm{T}}$. Likewise, $\Gamma_{\min} = (1/8)\alpha^2k\lfloor n/k\rfloor\sum_{i=1}^q R_i^{\mathrm{T}}\Gamma_{\mathrm{sb}}R_i$. By a change of variables and Lemma S2, we have

$$\begin{aligned}\sup_{\omega\in\mathcal{S}_{\Gamma_{\min}}}\mathrm{pr}\left(\omega^{\mathrm{T}}Z^{\mathrm{T}}Z\omega < 1\right) &= \sup_{\omega\in\mathbb{R}^m}\mathrm{pr}\left(\omega^{\mathrm{T}}Z^{\mathrm{T}}Z\omega < \omega^{\mathrm{T}}\Gamma_{\min}\omega\right)\\ &= \sup_{\omega\in\mathbb{R}^m}\mathrm{pr}\left(\sum_{i=1}^q\omega^{\mathrm{T}}R_i^{\mathrm{T}}X^{\mathrm{T}}XR_i\omega < \frac{\alpha^2k\lfloor n/k\rfloor}{8}\sum_{i=1}^q\omega^{\mathrm{T}}R_i^{\mathrm{T}}\Gamma_{\mathrm{sb}}R_i\omega\right)\\ &\leq \sum_{i=1}^q\sup_{\omega\in\mathbb{R}^m}\mathrm{pr}\left(\omega^{\mathrm{T}}R_i^{\mathrm{T}}X^{\mathrm{T}}XR_i\omega < \frac{\alpha^2k\lfloor n/k\rfloor}{8}\omega^{\mathrm{T}}R_i^{\mathrm{T}}\Gamma_{\mathrm{sb}}R_i\omega\right)\\ &\leq q\exp\left(-\frac{\alpha^2\lfloor n/k\rfloor}{8}\right) \leq q\exp\left(-\frac{\alpha^2n}{9k}\right),\end{aligned}$$

where we used (S5) again in the last inequality. This, together with (S9) and (S10), yields

$$\mathrm{pr}\left(\Gamma_{\min}/2 \npreceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right) \leq \exp\left\{m\log\frac{27}{\alpha} + \frac{1}{2}\log\det(\overline{\Gamma}_R\underline{\Gamma}_R^{-1}) + \log q - \frac{\alpha^2n}{9k}\right\} \leq \delta,$$

as long as condition (5) of the theorem holds.

Proof of (S8): Recall that $Z = \mathcal{U}\mathcal{D}\mathcal{V}^{\mathrm{T}}$ and $\mathcal{U}^{\mathrm{T}}\mathcal{U} = I_m$. Thus, on the event $\{\underline{\Gamma}_{\min} \preceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\}$, we have

$$\|\mathcal{U}^{\mathrm{T}}\eta\| = \sup_{\omega\in\mathbb{R}^m\setminus\{0\}} \frac{\omega^{\mathrm{T}}\mathcal{U}^{\mathrm{T}}\eta}{\|\omega\|} = \sup_{\omega\in\mathbb{R}^m\setminus\{0\}} \frac{\omega^{\mathrm{T}}\underline{\Gamma}_{\min}^{-1/2}\mathcal{V}\mathcal{D}\mathcal{U}^{\mathrm{T}}\eta}{\|\mathcal{D}\mathcal{V}^{\mathrm{T}}\underline{\Gamma}_{\min}^{-1/2}\omega\|} = \sup_{\omega\in\mathbb{R}^m\setminus\{0\}} \frac{\omega^{\mathrm{T}}\underline{\Gamma}_{\min}^{-1/2}Z^{\mathrm{T}}\eta}{\|Z\underline{\Gamma}_{\min}^{-1/2}\omega\|}$$
$$= \sup_{\omega\in\mathcal{S}_{\underline{\Gamma}_{\min}}} \frac{\omega^{\mathrm{T}}Z^{\mathrm{T}}\eta}{\|Z\omega\|},$$

where the second equality uses the fact that $\mathcal{D}\mathcal{V}^{\mathrm{T}}\underline{\Gamma}_{\min}^{-1/2}$ is nonsingular if $Z^{\mathrm{T}}Z \succeq \underline{\Gamma}_{\min} \succ 0$. Then it follows from Lemma S1(iii) that, on the event $\{\underline{\Gamma}_{\min} \preceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\}$, we have

$$\|\mathcal{U}^{\mathrm{T}}\eta\| \leq 2\max_{\omega\in\underline{\mathcal{T}}} \frac{\omega^{\mathrm{T}}Z^{\mathrm{T}}\eta}{\|Z\omega\|},$$

where $\underline{\mathcal{T}}$ is a $1/4$-net of $\mathcal{S}_{\underline{\Gamma}_{\min}}$ in the norm $\|\Gamma_{\max}^{1/2}(\cdot)\|$. Therefore,

$$\begin{aligned}
&\mathrm{pr}\left(\|\mathcal{U}^{\mathrm{T}}\eta\| \geq K,\ \underline{\Gamma}_{\min} \preceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right)\\
&\quad\leq \mathrm{pr}\left(\max_{\omega\in\underline{\mathcal{T}}} \frac{\omega^{\mathrm{T}}Z^{\mathrm{T}}\eta}{\|Z\omega\|} \geq K/2,\ \underline{\Gamma}_{\min} \preceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right)\\
&\quad\leq |\underline{\mathcal{T}}| \sup_{\omega\in\mathcal{S}_{\underline{\Gamma}_{\min}}} \mathrm{pr}\left(\frac{\omega^{\mathrm{T}}Z^{\mathrm{T}}\eta}{\|Z\omega\|} \geq K/2,\ \underline{\Gamma}_{\min} \preceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right)\\
&\quad= |\underline{\mathcal{T}}| \sup_{\omega\in\mathcal{S}^{m-1}} \mathrm{pr}\left(\frac{\omega^{\mathrm{T}}\underline{\Gamma}_{\min}^{-1/2}Z^{\mathrm{T}}\eta}{\|Z\underline{\Gamma}_{\min}^{-1/2}\omega\|} \geq K/2,\ I_d \preceq \underline{\Gamma}_{\min}^{-1/2}Z^{\mathrm{T}}Z\underline{\Gamma}_{\min}^{-1/2} \preceq \underline{\Gamma}_{\min}^{-1/2}\Gamma_{\max}\underline{\Gamma}_{\min}^{-1/2}\right)\\
&\quad\leq |\underline{\mathcal{T}}| \sup_{\omega\in\mathcal{S}^{m-1}} \mathrm{pr}\left\{\frac{\omega^{\mathrm{T}}\underline{\Gamma}_{\min}^{-1/2}Z^{\mathrm{T}}\eta}{\|Z\underline{\Gamma}_{\min}^{-1/2}\omega\|} \geq K/2,\ 1 \leq \|Z\underline{\Gamma}_{\min}^{-1/2}\omega\|^2 \leq \lambda_{\max}(\underline{\Gamma}_{\min}^{-1/2}\Gamma_{\max}\underline{\Gamma}_{\min}^{-1/2})\right\}.
\end{aligned} \tag{S11}$$

Similarly to (S10), we can show that

$$\log|\underline{\mathcal{T}}| \leq m\log 9 + (1/2)\log\det(\Gamma_{\max}\underline{\Gamma}_{\min}^{-1}) \leq m\log(38/\alpha) + (1/2)\log\det(\overline{\Gamma}_R\underline{\Gamma}_R^{-1}). \tag{S12}$$

Now it remains to derive a pointwise upper bound on the probability in (S11) for any fixed $\omega \in \mathcal{S}^{m-1}$. Let $\eta_{i,t}$ be the ith element of η_t, and denote

$$\eta_{(i)} = (\eta_{i,1}, \ldots, \eta_{i,n})^{\mathrm{T}}.$$

Note that $\eta = (\eta_{(1)}^{\mathrm{T}}, \ldots, \eta_{(q)}^{\mathrm{T}})^{\mathrm{T}}$. Fixing $\omega \in \mathcal{S}^{m-1}$, define $x_t = (x_{1,t}, \ldots, x_{q,t})^{\mathrm{T}}$, where $x_{i,t} = X_t^{\mathrm{T}}R_i\underline{\Gamma}_{\min}^{-1/2}\omega$, and denote

$$x_{(i)} = (x_{i,1}, \ldots, x_{i,n})^{\mathrm{T}} = XR_i\underline{\Gamma}_{\min}^{-1/2}\omega.$$

Then, we have $\omega^{\mathrm{T}}\underline{\Gamma}_{\min}^{-1/2}Z^{\mathrm{T}} = \omega^{\mathrm{T}}\underline{\Gamma}_{\min}^{-1/2}R^{\mathrm{T}}(I_q \otimes X^{\mathrm{T}}) = (x_{(1)}^{\mathrm{T}}, \ldots, x_{(q)}^{\mathrm{T}})$. As a result,

$$\omega^{\mathrm{T}}\underline{\Gamma}_{\min}^{-1/2}Z^{\mathrm{T}}\eta = \sum_{i=1}^{q} x_{(i)}^{\mathrm{T}}\eta_{(i)} = \sum_{i=1}^{q}\sum_{t=1}^{n} x_{i,t}\eta_{i,t} = \sum_{t=1}^{n} x_t^{\mathrm{T}}\eta_t$$

and

$$\|Z\underline{\Gamma}_{\min}^{-1/2}\omega\|^2 = \sum_{i=1}^{q}\|x_{(i)}\|^2 = \sum_{t=1}^{n}\|x_t\|^2.$$

Applying Lemma S3 to $\{x_t\}$ and $\{\eta_t\}$, with $\beta_- = 1$ and $\beta_+ = \lambda_{\max}(\underline{\Gamma}_{\min}^{-1/2}\Gamma_{\max}\underline{\Gamma}_{\min}^{-1/2})$, we have

$$\begin{aligned}
&\mathrm{pr}\left\{\frac{\omega^{\mathrm{T}}\underline{\Gamma}_{\min}^{-1/2}Z^{\mathrm{T}}\eta}{\|Z\underline{\Gamma}_{\min}^{-1/2}\omega\|} \geq K/2,\ 1 \leq \|Z\underline{\Gamma}_{\min}^{-1/2}\omega\|^2 \leq \lambda_{\max}(\underline{\Gamma}_{\min}^{-1/2}\Gamma_{\max}\underline{\Gamma}_{\min}^{-1/2})\right\} \\
&\quad = \mathrm{pr}\left\{\frac{\sum_{t=1}^{n} x_t^{\mathrm{T}}\eta_t}{(\sum_{t=1}^{n}\|x_t\|^2)^{1/2}} \geq K/2,\ \sum_{t=1}^{n}\|x_t\|^2 \in [\beta_-, \beta_+]\right\} \leq \frac{\beta_+}{\beta_-}\exp(-\frac{K^2}{24\sigma^2}).
\end{aligned} \tag{S13}$$

Moreover, by a method similar to that for (S12), we can show that

$$\frac{\beta_+}{\beta_-} \leq \det(\underline{\Gamma}_{\min}^{-1/2}\Gamma_{\max}\underline{\Gamma}_{\min}^{-1/2}) = \det(\Gamma_{\max}\underline{\Gamma}_{\min}^{-1}) \leq \exp\left\{m\log\frac{9}{2\alpha} + \log\det(\overline{\Gamma}_R\underline{\Gamma}_R^{-1})\right\}. \tag{S14}$$

Combining (S11)–(S14), we have

$$\mathrm{pr}\left(\|\mathcal{U}^{\mathrm{T}}\eta\| \geq K,\ \underline{\Gamma}_{\min} \preceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right) \leq \exp\left[2m\log\frac{14}{\alpha} + \frac{3}{2}\log\det(\overline{\Gamma}_R\underline{\Gamma}_R^{-1}) - \frac{K^2}{24\sigma^2}\right] \leq \delta,$$

if we choose K as mentioned below (S8). This completes the proof of this theorem.

S1.3. *Proof of Proposition* 1

Define the matrices $\Gamma_{\max}, \Gamma_{\min}$ and $\underline{\Gamma}_{\min}$ as in (S2), and consider the singular value decomposition of Z as in the proof of Theorem 1. Note that

$$\widehat{A} - A_* = \{I_q \otimes (\widehat{\theta} - \theta_*)^{\mathrm{T}}\}\widetilde{R} = \{I_q \otimes (Z^{\dagger}\eta)^{\mathrm{T}}\}\widetilde{R} = (I_q \otimes \eta^{\mathrm{T}}\mathcal{U})(I_q \otimes \mathcal{D}^{-}\mathcal{V}^{\mathrm{T}})\widetilde{R}.$$

Since $\underline{\Gamma}_{\min} \succ 0$, it holds on the event $\{Z^{\mathrm{T}}Z \succeq \underline{\Gamma}_{\min}\}$ that

$$\begin{aligned}
\|\widehat{A} - A_*\|_2 &\leq \|(I_q \otimes \mathcal{D}^{-}\mathcal{V}^{\mathrm{T}})\widetilde{R}\|_2\|\mathcal{U}^{\mathrm{T}}\eta\| = \left(\lambda_{\max}[\widetilde{R}^{\mathrm{T}}\{I_q \otimes (Z^{\mathrm{T}}Z)^{-1}\}\widetilde{R}]\right)^{1/2}\|\mathcal{U}^{\mathrm{T}}\eta\| \\
&\leq \left[\lambda_{\max}\{\widetilde{R}^{\mathrm{T}}(I_q \otimes \underline{\Gamma}_{\min}^{-1})\widetilde{R}\}\right]^{1/2}\|\mathcal{U}^{\mathrm{T}}\eta\|.
\end{aligned}$$

Consequently, by a method similar to that for (S4), under Assumption 2, we can show that

$$\begin{aligned}
&\mathrm{pr}\left(\|\widehat{A} - A_*\|_2 \geq K\left[\lambda_{\max}\{\widetilde{R}^{\mathrm{T}}(I_q \otimes \underline{\Gamma}_{\min}^{-1})\widetilde{R}\}\right]^{1/2}\right) \\
&\quad \leq \mathrm{pr}\left(\|\mathcal{U}^{\mathrm{T}}\eta\| \geq K,\ \underline{\Gamma}_{\min} \preceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right) + \mathrm{pr}\left(\underline{\Gamma}_{\min} \npreceq Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\right) + \delta
\end{aligned}$$

for any $K > 0$. Moreover, similarly to (S6), we have

$$\begin{aligned}
\left[\lambda_{\max}\{\widetilde{R}^{\mathrm{T}}(I_q \otimes \underline{\Gamma}_{\min}^{-1})\widetilde{R}\}\right]^{1/2} &\leq \frac{9}{2\alpha}\left[\frac{\lambda_{\max}\{\widetilde{R}^{\mathrm{T}}(I_q \otimes \underline{\Gamma}_R^{-1})\widetilde{R}\}}{n}\right]^{1/2} \\
&= \frac{9}{2\alpha}\left\{\frac{\lambda_{\max}\left(\sum_{i=1}^{q} R_i\underline{\Gamma}_R^{-1}R_i^{\mathrm{T}}\right)}{n}\right\}^{1/2}.
\end{aligned}$$

Then, along the same lines of the arguments for Theorem 1, we accomplish the proof of this proposition.

S1.4. *Proofs of Lemmas S1–S3*

The covering and discretization results in Lemma S1 are modified from Lemmas 4.1, D.1 and D.2 in Simchowitz et al. (2018). For clarity, we rewrite the proofs of Lemma S1(i)–(ii) to correct any typographical error in their proofs, and present our own proof of Lemma S1(iii). Lemma S2 establishes a pointwise lower bound on $X^{\mathrm{T}}X$ via the BMSB condition, which will be strengthened into a union bound in the proof of Theorem 1 via Lemma S1(i); see also Proposition 2.5 in the above paper. Finally, as a multivariate generalization of Lemma 4.2(b) in their paper, Lemma S3 gives a concentration bound on $\sum_{t=1}^n x_t^{\mathrm{T}}\eta_t/(\sum_{t=1}^n \|x_t\|^2)^{1/2}$. Note that it is crucial to bound this self-normalized process as a whole, instead of bounding the numerator $\sum_{t=1}^n x_t^{\mathrm{T}}\eta_t$ and the denominator $(\sum_{t=1}^n \|x_t\|^2)^{1/2}$ separately; otherwise, the bound would degrade for slower-mixing processes.

*Proof of Lemma S*1. Note that claim (i) will be used to cover $\mathcal{S}^{m-1}$ in terms of $\Gamma_{\min}$ and $\Gamma_{\max}$ for deriving the union upper bound on $Z^{\mathrm{T}}Z$ in the proof of Theorem 1. The corresponding covering number is given in claim (ii), which is larger when $\Gamma_{\max}$ is farther away from $\Gamma_{\min}$ as measured by $\log\det(\Gamma_{\max}\Gamma_{\min}^{-1})$. Claim (iii) is a discretization result for $\omega^{\mathrm{T}}Z\nu/\|Z^{\mathrm{T}}\omega\|$.

To prove (i), it is equivalent to show that

$$\mathcal{E} = \{\inf_{\omega\in\mathcal{T}} \omega^{\mathrm{T}}Z^{\mathrm{T}}Z\omega \geq 1\} \cap \{Z^{\mathrm{T}}Z \preceq \Gamma_{\max}\} \subseteq \{Z^{\mathrm{T}}Z \succeq \Gamma_{\min}/2\}. \tag{S15}$$

Since $\mathcal{T}$ is a $1/4$-net of $\mathcal{S}_{\Gamma_{\min}}$ in the norm $\|\Gamma_{\max}^{1/2}(\cdot)\|$, on the event $\mathcal{E}$, we have

$$\begin{aligned}
1/4 \geq \sup_{\omega\in\mathcal{S}_{\Gamma_{\min}}} \inf_{\upsilon\in\mathcal{T}} \|\Gamma_{\max}^{1/2}(\omega-\upsilon)\| &\geq \sup_{\omega\in\mathcal{S}_{\Gamma_{\min}}} \inf_{\upsilon\in\mathcal{T}} \|Z(\omega-\upsilon)\| \\
&\geq \sup_{\omega\in\mathcal{S}_{\Gamma_{\min}}} \inf_{\upsilon\in\mathcal{T}} (\|Z\upsilon\| - \|Z\omega\|) = \inf_{\omega\in\mathcal{T}} \|Z\omega\| - \inf_{\omega\in\mathcal{S}_{\Gamma_{\min}}} \|Z\omega\| \\
&\geq 1 - \inf_{\omega\in\mathcal{S}_{\Gamma_{\min}}} \|Z\omega\|,
\end{aligned}$$

where the second and last inequalities are due to $Z^{\mathrm{T}}Z \preceq \Gamma_{\max}$ and $\inf_{\omega\in\mathcal{T}} \omega^{\mathrm{T}}Z^{\mathrm{T}}Z\omega \geq 1$, respectively. As a result,

$$3/4 \leq \inf_{\omega\in\mathcal{S}_{\Gamma_{\min}}} \|Z\omega\| = \inf_{\omega\in\mathcal{S}^{m-1}} \|Z\Gamma_{\min}^{-1/2}\omega\| = \left\{\lambda_{\min}(\Gamma_{\min}^{-1/2}Z^{\mathrm{T}}Z\Gamma_{\min}^{-1/2})\right\}^{1/2}.$$

Therefore, $Z^{\mathrm{T}}Z \succeq (9/16)\Gamma_{\min} \succeq \Gamma_{\min}/2$, i.e., (S15) holds.

The proof of claim (ii) is basically the same as that in Simchowitz et al. (2018), except for some minor corrections. Note that $|\mathcal{T}|$ is equal to the covering number of the shell of the ellipsoid $E = \{\omega \in \mathbb{R}^m : \omega^{\mathrm{T}}\Gamma_{\max}^{-1/2}\Gamma_{\min}\Gamma_{\max}^{-1/2}\omega \leq 1\}$ in the Euclidean norm. Let $B = \{x \in \mathbb{R}^m : \|x\| \leq 1\}$ be the unit ball in $\mathbb{R}^m$, and denote by $+$ the Minkowski sum. If $\mathcal{T}$ is a minimal ϵ-net of $\mathcal{S}_{\Gamma_{\min}}$ in the norm $\|\Gamma_{\max}^{1/2}(\cdot)\|$, then it follows from a standard volumetric argument that

$$|\mathcal{T}| \le \frac{\text{vol}\{E + (\epsilon/2)B\}}{\text{vol}\{(\epsilon/2)B\}} \le \frac{\text{vol}\{(1+\epsilon/2)E\}}{\text{vol}\{(\epsilon/2)B\}} = \frac{(1+\epsilon/2)^m \text{vol}(E)}{(\epsilon/2)^m \text{vol}(B)}$$
$$= \frac{(1+\epsilon/2)^m}{(\epsilon/2)^m \left\{\det(\Gamma_{\max}^{-1/2}\Gamma_{\min}\Gamma_{\max}^{-1/2})\right\}^{1/2}}$$
$$= (2/\epsilon + 1)^m \left\{\det(\Gamma_{\min}^{-1}\Gamma_{\max})\right\}^{1/2}.$$

Taking $\epsilon = 1/4$ yields the result in (ii).

Finally, we prove (iii). First note that since $\Gamma_{\min} \succ 0$, we have

$$\sup_{\omega \in \mathcal{S}^{m-1}} \frac{\omega^{\mathrm{T}} Z^{\mathrm{T}} \nu}{\|Z\omega\|} = \sup_{\omega \in \mathbb{R}^m \setminus \{0\}} \frac{\omega^{\mathrm{T}} Z^{\mathrm{T}} \nu}{\|Z\omega\|} = \sup_{\omega \in \mathbb{R}^m \setminus \{0\}} \frac{\omega^{\mathrm{T}} \Gamma_{\min}^{-1/2} Z^{\mathrm{T}} \nu}{\|Z\Gamma_{\min}^{-1/2}\omega\|} = \sup_{\omega \in \mathcal{S}_{\Gamma_{\min}}} \frac{\omega^{\mathrm{T}} Z^{\mathrm{T}} \nu}{\|Z\omega\|}.$$

For a fixed $\nu \in \mathbb{R}^n$, define $\phi : \mathbb{R}^m \setminus \{0\} \to \mathbb{R}$ by

$$\phi(\omega) = \frac{\omega^{\mathrm{T}} Z^{\mathrm{T}} \nu}{\|Z\omega\|}.$$

To prove (iii), we will show that for any $\omega \in \mathcal{S}_{\Gamma_{\min}}$, there exist $\omega_0 \in \mathcal{T}$ and $u \in \mathbb{R}^d \setminus \{0\}$ such that

$$\phi(\omega) \le \phi(\omega_0) + (1/2)\phi(u). \tag{S16}$$

Let

$$u = \frac{\omega}{\|Z\omega\|} - \frac{\omega_0}{\|Z\omega_0\|}.$$

Then, $u \ne 0$ as long as $\omega \ne \omega_0$, and we have

$$\phi(\omega) - \phi(\omega_0) = u^{\mathrm{T}} Z^{\mathrm{T}} \nu = \|Zu\|\phi(u).$$

Therefore, to prove (S16), it suffices to show that

$$\|Zu\| \le 1/2. \tag{S17}$$

Note that

$$Zu = \frac{Z(\omega - \omega_0)}{\|Z\omega\|} + \frac{Z\omega_0}{\|Z\omega_0\|}\frac{\|Z\omega\| - \|Z\omega_0\|}{\|Z\omega\|}.$$

As a result,

$$\|Zu\| \le \frac{2\|Z(\omega - \omega_0)\|}{\|Z\omega\|} \le \frac{2\|Z(\omega - \omega_0)\|}{\inf_{\omega \in \mathcal{S}_{\Gamma_{\min}}} \|Z\omega\|}. \tag{S18}$$

Since $0 \prec \Gamma_{\min} \preceq Z^{\mathrm{T}} Z$, we have

$$\inf_{\omega \in \mathcal{S}_{\Gamma_{\min}}} \|Z\omega\| = \inf_{\omega \in \mathcal{S}^{m-1}} \|Z\Gamma_{\min}^{-1/2}\omega\| = \left\{\lambda_{\min}(\Gamma_{\min}^{-1/2} Z^{\mathrm{T}} Z \Gamma_{\min}^{-1/2})\right\}^{1/2} \ge 1.$$

Moreover, since $Z^{\mathrm{T}} Z \preceq \Gamma_{\max}$ and $\mathcal{T}$ is a $1/4$-net of $\mathcal{S}_{\Gamma_{\min}}$ in $\|\Gamma_{\max}^{1/2}(\cdot)\|$, there is $\omega_0 \in \mathcal{T}$ such that

$$\|Z(\omega - \omega_0)\| \le \|\Gamma_{\max}^{1/2}(\omega - \omega_0)\| \le 1/4.$$

Combining the results above with (S18), we have (S17), and hence (S16). Taking the supremum with respect to $\omega \in \mathcal{S}_{\Gamma_{\min}}$ on both sides of (S16), we have

$$\sup_{\omega \in \mathcal{S}_{\Gamma_{\min}}} \frac{\omega^{\mathrm{T}} Z^{\mathrm{T}} \nu}{\|Z\omega\|} \leq \max_{\omega_0 \in \mathcal{T}} \frac{\omega_0^{\mathrm{T}} Z^{\mathrm{T}} \nu}{\|Z\omega_0\|} + \frac{1}{2} \sup_{u \in \mathbb{R}^d \setminus \{0\}} \frac{u^{\mathrm{T}} Z^{\mathrm{T}} \nu}{\|Zu\|} = \max_{\omega \in \mathcal{T}} \frac{\omega^{\mathrm{T}} Z^{\mathrm{T}} \nu}{\|Z\omega\|} + \frac{1}{2} \sup_{\omega \in \mathcal{S}_{\Gamma_{\min}}} \frac{\omega^{\mathrm{T}} Z^{\mathrm{T}} \nu}{\|Z\omega\|},$$

which yields the inequality in (iii). □

Proof of Lemma S2. This lemma directly follows from Proposition 2.5 of Simchowitz et al. (2018), where the Chernoff bound technique is applied to lower bound the Gram matrix via aggregating all the $\lfloor n/k \rfloor$ blocks of size k. □

Proof of Lemma S3. Along the lines of the proof of Lemma 4.2 in Simchowitz et al. (2018), we can show that the left-hand side of (S1) is bounded above by $\log\lceil \beta_+/\beta_- \rceil \exp\{-\gamma^2/(6\sigma^2)\}$. Then (S1) follows from the fact that $\log\lceil x \rceil < \log(1+x) \leq x$ for $x > 0$. □

S2. PROOFS OF LEMMAS 1–3

S2.1. *Proof of Lemma* 1

First note that for any $s \in \mathbb{Z}$ and positive integer t,

$$X_{s+t} = \eta_{s+t-1} + A_*\eta_{s+t-2} + \cdots + A_*^{t-1}\eta_s + A_*^t X_s = \sum_{\ell=0}^{t-1} A_*^{\ell} \eta_{s+t-\ell-1} + A_*^t X_s,$$

where, by Assumption 4(ii), $\sum_{\ell=0}^{t-1} A_*^{\ell} \eta_{s+t-\ell-1}$ is independent of $\mathcal{F}_s$, and $A_*^t X_s \in \mathcal{F}_s$. Then, for any $\omega \in \mathcal{S}^{d-1}$, we have

$$\frac{\omega^{\mathrm{T}} X_{s+t}}{\sigma(\omega^{\mathrm{T}} \Gamma_t \omega)^{1/2}} = S_\omega + c_\omega,$$

where $c_\omega = \omega^{\mathrm{T}} A_*^t X_s / \{\sigma(\omega^{\mathrm{T}} \Gamma_t \omega)^{1/2}\}$, and S_ω can be written as a weighted sum of real-valued independent random variables,

$$S_\omega = \frac{\omega^{\mathrm{T}} \sum_{\ell=0}^{t-1} A_*^{\ell} \eta_{s+t-\ell-1}}{\sigma(\omega^{\mathrm{T}} \Gamma_t \omega)^{1/2}} = \sum_{\ell=0}^{t-1} a_\ell e_\ell,$$

with

$$a_\ell = \left\{ \frac{\omega^{\mathrm{T}} A_*^{\ell} (A_*^{\mathrm{T}})^{\ell} \omega}{\omega^{\mathrm{T}} \Gamma_t \omega} \right\}^{1/2}, \quad e_\ell = \frac{\omega^{\mathrm{T}} A_*^{\ell} \eta_{s+t-\ell-1}}{\sigma\{\omega^{\mathrm{T}} A_*^{\ell} (A_*^{\mathrm{T}})^{\ell} \omega\}^{1/2}}.$$

Notice that $\sum_{\ell=0}^{t-1} a_\ell^2 = 1$, and $e_0, \ldots, e_{t-1}$ are real-valued independent random variables. Moreover, by Assumption 4(iii), the density of each e_ℓ is bounded by C_0 almost everywhere. Applying Theorem 1.2 in Rudelson & Vershynin (2015), it follows that the density of S_ω is bounded by $\sqrt{2}C_0$ almost everywhere. In addition, S_ω is independent of $\mathcal{F}_s$, and $c_\omega \in \mathcal{F}_s$. Therefore,

$$\begin{aligned} \mathrm{pr}\left\{ |\omega^{\mathrm{T}} X_{s+t}| \geq \sigma(\omega^{\mathrm{T}} \Gamma_t \omega)^{1/2} (4C_0)^{-1} \mid \mathcal{F}_s \right\} &= \mathrm{pr}\left\{ |S_\omega + c_\omega| \geq (4C_0)^{-1} \mid \mathcal{F}_s \right\} \\ &= 1 - \mathrm{pr}\left\{ |S_\omega + c_\omega| \leq (4C_0)^{-1} \mid \mathcal{F}_s \right\} \\ &\geq 1 - \sqrt{2}/2 > 0. \end{aligned} \tag{S19}$$

For any integer $k \geq 1$, by (S19) and the fact that $\Gamma_t \succeq \Gamma_k$ for $t \geq k$, we have

$$\begin{aligned}
&\frac{1}{2k}\sum_{t=1}^{2k} \mathrm{pr}\left\{|\omega^{\mathrm{T}} X_{s+t}| \geq \sigma(\omega^{\mathrm{T}}\Gamma_k\omega)^{1/2}(4C_0)^{-1} \mid \mathcal{F}_s\right\} \\
&\quad \geq \frac{1}{2k}\sum_{t=k}^{2k} \mathrm{pr}\left\{|\omega^{\mathrm{T}} X_{s+t}| \geq \sigma(\omega^{\mathrm{T}}\Gamma_k\omega)^{1/2}(4C_0)^{-1} \mid \mathcal{F}_s\right\} \\
&\quad \geq \frac{1}{2k}\sum_{t=k}^{2k} \mathrm{pr}\left\{|\omega^{\mathrm{T}} X_{s+t}| \geq \sigma(\omega^{\mathrm{T}}\Gamma_t\omega)^{1/2}(4C_0)^{-1} \mid \mathcal{F}_s\right\} \\
&\quad \geq \frac{(2k-k+1)(1-\sqrt{2}/2)}{2k} > \frac{1}{10}.
\end{aligned}$$

Choosing $\Gamma_{\mathrm{sb}} = \sigma^2\Gamma_k/(4C_0)^2$, we accomplish the proof of this lemma.

S2.2. *Proof of Lemma* 2

Since $E(X^{\mathrm{T}}X) = \sigma^2\sum_{t=1}^{n}\Gamma_t \preceq \sigma^2 n\Gamma_n$, we have

$$E(Z^{\mathrm{T}}Z) = R^{\mathrm{T}}\{I_d \otimes E(X^{\mathrm{T}}X)\}R \leq \sigma^2 nR^{\mathrm{T}}(I_d \otimes \Gamma_n)R.$$

Then, with $\overline{\Gamma}_R = (\sigma^2 m/\delta)R^{\mathrm{T}}(I_d \otimes \Gamma_n)R$, it follows from the Markov inequality that

$$\begin{aligned}
\mathrm{pr}(Z^{\mathrm{T}}Z \npreceq n\overline{\Gamma}_R) &= \mathrm{pr}\left[\lambda_{\max}\{(n\overline{\Gamma}_R)^{-1/2}Z^{\mathrm{T}}Z(n\overline{\Gamma}_R)^{-1/2}\} \geq 1\right] \\
&\leq E\left[\lambda_{\max}\{(n\overline{\Gamma}_R)^{-1/2}Z^{\mathrm{T}}Z(n\overline{\Gamma}_R)^{-1/2}\}\right] \\
&\leq \mathrm{tr}\left\{(n\overline{\Gamma}_R)^{-1/2}E(Z^{\mathrm{T}}Z)(n\overline{\Gamma}_R)^{-1/2}\right\} \\
&\leq \mathrm{tr}\{(\delta/m)I_m\} = \delta,
\end{aligned}$$

which completes the proof of this lemma. Note that the factor of m in the definition of $\overline{\Gamma}_R$ is a consequence of upper bounding $\lambda_{\max}(\cdot)$ by $\mathrm{tr}(\cdot)$.

S2.3. *Proof of Lemma* 3

Recall that $R = (R_1^{\mathrm{T}}, \ldots, R_d^{\mathrm{T}})^{\mathrm{T}}$, where each R_i is a $d \times m$ block. For simplicity, denote $Q_i = XR_i(R_i^{\mathrm{T}}R_i)^{-1/2} = (Q_{i,1}, \ldots, Q_{i,n})^{\mathrm{T}} \in \mathbb{R}^{n\times m}$ with $i = 1, \ldots, d$, where $Q_{i,t} = (R_i^{\mathrm{T}}R_i)^{-1/2}R_i^{\mathrm{T}}X_t \in \mathbb{R}^m$. To prove this lemma, it suffices to verify the following two results:

$$\mathrm{pr}(Z^{\mathrm{T}}Z \npreceq n\overline{\Gamma}_R) \leq \sum_{i=1}^{d} P\{\|Q_i^{\mathrm{T}}Q_i - E(Q_i^{\mathrm{T}}Q_i)\|_2 > n\sigma^2\xi\}, \tag{S20}$$

where $\overline{\Gamma}_R = \sigma^2 R^{\mathrm{T}}(I_d \otimes \Gamma_n)R + \sigma^2\xi R^{\mathrm{T}}R$, and

$$\mathrm{pr}\{\|Q_i^{\mathrm{T}}Q_i - E(Q_i^{\mathrm{T}}Q_i)\|_2 > n\sigma^2\xi\} \leq \delta/d \quad (i = 1, \ldots, d). \tag{S21}$$

We first prove (S20). Note that $Z^{\mathrm{T}}Z = R^{\mathrm{T}}\{I_d \otimes (X^{\mathrm{T}}X)\}R$ and $E(Z^{\mathrm{T}}Z) = R^{\mathrm{T}}\{I_d \otimes E(X^{\mathrm{T}}X)\}R = \sigma^2 R^{\mathrm{T}}\{I_d \otimes \sum_{t=1}^{n}\Gamma_t\}R$. Then, since $n\Gamma_n \succeq \sum_{t=1}^{n}\Gamma_t$, we have

$$n\overline{\Gamma}_R \succeq \sigma^2 R^{\mathrm{T}}\{I_d \otimes \sum_{t=1}^{n}\Gamma_t\}R + n\sigma^2\xi R^{\mathrm{T}}R = E(Z^{\mathrm{T}}Z) + n\sigma^2\xi R^{\mathrm{T}}R.$$

As a result,

$$\begin{aligned}
\mathrm{pr}(Z^{\mathrm{T}}Z \preceq n\overline{\Gamma}_R) &= \mathrm{pr}\{Z^{\mathrm{T}}Z - E(Z^{\mathrm{T}}Z) \preceq n\sigma^2\xi R^{\mathrm{T}}R\} \\
&= \mathrm{pr}\left[\sum_{i=1}^{d} R_i^{\mathrm{T}}\{X^{\mathrm{T}}X - E(X^{\mathrm{T}}X)\}R_i \preceq n\sigma^2\xi \sum_{i=1}^{d} R_i^{\mathrm{T}}R_i\right] \\
&\geq \mathrm{pr}\left\{\bigcap_{i=1}^{d} \left[R_i^{\mathrm{T}}\{X^{\mathrm{T}}X - E(X^{\mathrm{T}}X)\}R_i \preceq n\sigma^2\xi R_i^{\mathrm{T}}R_i\right]\right\} \\
&\geq 1 - \sum_{i=1}^{d} \mathrm{pr}\left[R_i^{\mathrm{T}}\{X^{\mathrm{T}}X - E(X^{\mathrm{T}}X)\}R_i \npreceq n\sigma^2\xi R_i^{\mathrm{T}}R_i\right]. \qquad \text{(S22)}
\end{aligned}$$

Moreover, for $i = 1, \ldots, d$, we have

$$\begin{aligned}
&\mathrm{pr}\left[R_i^{\mathrm{T}}\{X^{\mathrm{T}}X - E(X^{\mathrm{T}}X)\}R_i \preceq n\sigma^2\xi R_i^{\mathrm{T}}R_i\right] \\
&\quad = \mathrm{pr}\left[(R_i^{\mathrm{T}}R_i)^{-1/2}R_i^{\mathrm{T}}\{X^{\mathrm{T}}X - E(X^{\mathrm{T}}X)\}R_i(R_i^{\mathrm{T}}R_i)^{-1/2} \preceq n\sigma^2\xi I_m\right] \\
&\quad = \mathrm{pr}\{Q_i^{\mathrm{T}}Q_i - E(Q_i^{\mathrm{T}}Q_i) \preceq n\sigma^2\xi I_m\} \\
&\quad \geq \mathrm{pr}\{\|Q_i^{\mathrm{T}}Q_i - E(Q_i^{\mathrm{T}}Q_i)\|_2 \leq n\sigma^2\xi\},
\end{aligned}$$

which implies

$$\mathrm{pr}\left[R_i^{\mathrm{T}}\{X^{\mathrm{T}}X - E(X^{\mathrm{T}}X)\}R_i \npreceq n\sigma^2\xi R_i^{\mathrm{T}}R_i\right] \leq \mathrm{pr}\{\|Q_i^{\mathrm{T}}Q_i - E(Q_i^{\mathrm{T}}Q_i)\|_2 > n\sigma^2\xi\}. \qquad \text{(S23)}$$

Combining (S22) and (S23), we accomplish the proof of (S20).

Next we prove (S21). Let $\mathcal{T}_0$ be a minimal $(1/4)$-net of the sphere $\mathcal{S}^{m-1}$ in the Euclidean norm. It follows from a standard volumetric argument that $|\mathcal{T}_0| \leq 9^m$. Moreover, by Lemma 5.4 in Vershynin (2012), we have

$$\|Q_i^{\mathrm{T}}Q_i - E(Q_i^{\mathrm{T}}Q_i)\|_2 \leq 2\max_{\omega\in\mathcal{T}_0} |\omega^{\mathrm{T}}\{Q_i^{\mathrm{T}}Q_i - E(Q_i^{\mathrm{T}}Q_i)\}\omega|. \qquad \text{(S24)}$$

Furthermore, since $\{\eta_t\}$ are normal, $\mathrm{vec}(X^{\mathrm{T}}) = (X_1^{\mathrm{T}}, \ldots, X_n^{\mathrm{T}})^{\mathrm{T}}$ follows the multivariate normal distribution with mean zero and covariance matrix Σ_X. Then, for any $\omega \in \mathcal{S}^{m-1}$, we have

$$Q_i\omega = (Q_{i,1}^{\mathrm{T}}\omega, \ldots, Q_{i,n}^{\mathrm{T}}\omega)^{\mathrm{T}} = \left[I_n \otimes \{\omega^{\mathrm{T}}(R_i^{\mathrm{T}}R_i)^{-1/2}R_i^{\mathrm{T}}\}\right]\mathrm{vec}(X^{\mathrm{T}}) \sim N(0, \Sigma_\omega),$$

where

$$\Sigma_\omega = \left[I_n \otimes \{\omega^{\mathrm{T}}(R_i^{\mathrm{T}}R_i)^{-1/2}R_i^{\mathrm{T}}\}\right]\Sigma_X\left[I_n \otimes \{R_i(R_i^{\mathrm{T}}R_i)^{-1/2}\omega\}\right],$$

and hence there exists $z \sim N(0, I_n)$ such that $\omega^{\mathrm{T}}Q_i^{\mathrm{T}}Q_i\omega = z^{\mathrm{T}}\Sigma_\omega z$. As a result, it follows from the Hanson-Wright inequality (Vershynin, 2018) that, for every $\xi > 0$,

$$\begin{aligned}
\mathrm{pr}\left[|\omega^{\mathrm{T}}\{Q_i^{\mathrm{T}}Q_i - E(Q_i^{\mathrm{T}}Q_i)\}\omega| > n\sigma^2\xi/2\right] &= \mathrm{pr}\left[|z^{\mathrm{T}}\Sigma_\omega z - E(z^{\mathrm{T}}\Sigma_\omega z)| > n\sigma^2\xi/2\right] \\
&\leq 2\exp\left[-\frac{1}{C_1}\min\left\{\left(\frac{n\sigma^2\xi}{2\|\Sigma_\omega\|_F}\right)^2, \frac{n\sigma^2\xi}{2\|\Sigma_\omega\|_2}\right\}\right], \qquad \text{(S25)}
\end{aligned}$$

where $C_1 > 0$ is a universal constant. In view of (S24) and (S25), we have

$$\begin{aligned}\mathrm{pr}\{\|Q_i^{\mathrm{T}}Q_i - E(Q_i^{\mathrm{T}}Q_i)\|_2 > n\sigma^2\xi\} &\leq \mathrm{pr}\left[\max_{\omega\in\mathcal{T}_0}|\omega^{\mathrm{T}}\{Q_i^{\mathrm{T}}Q_i - E(Q_i^{\mathrm{T}}Q_i)\}\omega| > n\sigma^2\xi/2\right] \\ &\leq |\mathcal{T}_0|\ \mathrm{pr}\left[|\omega^{\mathrm{T}}\{Q_i^{\mathrm{T}}Q_i - E(Q_i^{\mathrm{T}}Q_i)\}\omega| > n\sigma^2\xi/2\right] \\ &\leq 2(9^m)\exp\left[-\frac{1}{C_1}\min\left\{\left(\frac{n\sigma^2\xi}{2\|\Sigma_\omega\|_F}\right)^2, \frac{n\sigma^2\xi}{2\|\Sigma_\omega\|_2}\right\}\right] \\ &\leq \delta/d\end{aligned}$$

as long as

$$\xi \geq \frac{2}{n\sigma^2}\left[\{\psi(m,d,\delta)\}^{1/2}\|\Sigma_\omega\|_F + \psi(m,d,\delta)\|\Sigma_\omega\|_2\right], \tag{S26}$$

where $\psi(m,d,\delta) = C_1\{m\log 9 + \log d + \log(2/\delta)\}$.

To prove (S21), it now remains to choose ξ such that (S26) holds. Note that

$$\|\Sigma_\omega\|_2 \leq \lambda_{\max}\left[I_n \otimes \{\omega^{\mathrm{T}}(R_i^{\mathrm{T}}R_i)^{-1/2}R_i^{\mathrm{T}}R_i(R_i^{\mathrm{T}}R_i)^{-1/2}\omega\}\right]\|\Sigma_X\|_2 = \|\Sigma_X\|_2. \tag{S27}$$

Moreover, since

$$\begin{aligned}\mathrm{tr}(\Sigma_\omega) &= \sigma^2\sum_{t=1}^{n}\omega^{\mathrm{T}}(R_i^{\mathrm{T}}R_i)^{-1/2}R_i^{\mathrm{T}}\Gamma_t R_i(R_i^{\mathrm{T}}R_i)^{-1/2}\omega \\ &\leq \sigma^2 n\lambda_{\max}(\Gamma_n)\lambda_{\max}\left\{(R_i^{\mathrm{T}}R_i)^{-1/2}R_i^{\mathrm{T}}R_i(R_i^{\mathrm{T}}R_i)^{-1/2}\right\} \\ &= \sigma^2 n\lambda_{\max}(\Gamma_n),\end{aligned}$$

in light of (S27), we have

$$\|\Sigma_\omega\|_F = \{\mathrm{tr}(\Sigma_\omega^2)\}^{1/2} \leq \{\|\Sigma_\omega\|_2\,\mathrm{tr}(\Sigma_\omega)\}^{1/2} \leq \left\{\sigma^2 n\lambda_{\max}(\Gamma_n)\|\Sigma_X\|_2\right\}^{1/2}. \tag{S28}$$

Replacing $\|\Sigma_\omega\|_2$ and $\|\Sigma_\omega\|_F$ in (S26) by their upper bounds in (S27) and (S28), respectively, it follows that (S21) holds if we choose ξ as in (11) in the main paper. The proof of this lemma is complete.

S3. PROOFS OF EQUATION (15) AND PROPOSITION 2

S3.1. *Proof of Equation* (15)

The proof of Equation (15) relies on the following lemma:

LEMMA S4. *Let A and B be $m \times m$ symmetric positive definite matrices such that $B^{1/2}AB^{1/2} \succeq I_m$. For any $\xi > 0$, it holds $\log\det(AB + \xi I_m) \leq m\log\{2\max(1,\xi)\} + \log\det(AB)$.*

Proof of Lemma S4. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 1$ be the eigenvalues of $B^{1/2}AB^{1/2}$. For any $\xi > 0$, it can be readily shown that $\lambda_i + \xi \leq 2\max(1,\xi)\lambda_i$ for all i. Moreover, by Theorem 1.3.20 in Horn & Johnson (1985), AB and $B^{1/2}AB^{1/2}$ have the same nonzero eigenvalues.

Thus,

$$\begin{aligned}\log\det(AB+\xi I_m) &= \sum_{i=1}^{m}\log(\lambda_i+\xi) \le \sum_{i=1}^{m}\left[\log\{2\max(1,\xi)\}+\log\lambda_i\right] \\ &= m\log\{2\max(1,\xi)\}+\log\det(AB).\end{aligned}$$

The proof of Lemma S4 is complete. □

Now we prove Equation (15). First note that

$$\log\det\{\overline{\Gamma}_R(\sigma^2R^{\mathrm{T}}R)^{-1}(4C_0)^2\} = \log\det\{\overline{\Gamma}_R(\sigma^2R^{\mathrm{T}}R)^{-1}\}+2m\log(4C_0). \tag{S29}$$

If $\overline{\Gamma}_R=\overline{\Gamma}_R^{(1)}=\sigma^2mR^{\mathrm{T}}(I_d\otimes\Gamma_n)R/\delta$, it is easy to see that

$$\log\det\{\overline{\Gamma}_R(\sigma^2R^{\mathrm{T}}R)^{-1}\} = m\log(m/\delta)+\kappa, \tag{S30}$$

where $\kappa=\log\det\left\{R^{\mathrm{T}}(I_d\otimes\Gamma_n)R(R^{\mathrm{T}}R)^{-1}\right\}$.

On the other hand, if $\overline{\Gamma}_R=\overline{\Gamma}_R^{(2)}=\sigma^2R^{\mathrm{T}}(I_d\otimes\Gamma_n)R+\sigma^2\xi R^{\mathrm{T}}R$, we have

$$\log\det\{\overline{\Gamma}_R(\sigma^2R^{\mathrm{T}}R)^{-1}\} = \log\det\left\{R^{\mathrm{T}}(I_d\otimes\Gamma_n)R(R^{\mathrm{T}}R)^{-1}+\xi I_m\right\},$$

Note that $(R^{\mathrm{T}}R)^{-1/2}R^{\mathrm{T}}(I_d\otimes\Gamma_n)R(R^{\mathrm{T}}R)^{-1/2}\succeq I_m$, since $\Gamma_n\succeq I_d$. Then, applying Lemma S4 with $A=R^{\mathrm{T}}(I_d\otimes\Gamma_n)R$ and $B=(R^{\mathrm{T}}R)^{-1}$, we have

$$\log\det\{\overline{\Gamma}_R(\sigma^2R^{\mathrm{T}}R)^{-1}\} \le m\log\{2\max(1,\xi)\}+\kappa. \tag{S31}$$

Combining (S29)–(S31), we accomplish the proof of Equation (15).

S3.2. *Proof of Proposition* 2

Proposition 2 is a direct consequence of Equations (14) and (15) in the main paper and the upper bounds of κ and ξ in Lemmas S5 and S6 below. Note that the proofs of Lemmas S5 and S6 rely crucially on the intermediate results on upper bounds of $\lambda_{\max}(\Gamma_n)$ and $\|\Sigma_X\|_2$ as given in Lemmas S7 and S8 below, respectively. The proofs of Lemmas S5–S8 are collected in § S3.3.

Recall that $\Sigma_X=[E(X_tX_s^{\mathrm{T}})]_{1\le t,s\le n}$, where $E(X_tX_s^{\mathrm{T}})=\sigma^2A_*^{t-s}\Gamma_s$ for $1\le s\le t\le n$ under Assumptions 4(i) and (ii), $\kappa=\log\det\left\{R^{\mathrm{T}}(I_d\otimes\Gamma_n)R(R^{\mathrm{T}}R)^{-1}\right\}$, and

$$\xi=\xi(m,d,n,\delta)=2\left\{\frac{\lambda_{\max}(\Gamma_n)\psi(m,d,\delta)\|\Sigma_X\|_2}{\sigma^2n}\right\}^{1/2}+\frac{2\psi(m,d,\delta)\|\Sigma_X\|_2}{\sigma^2n},$$

where $\psi(m,d,\delta)=C_1\{m\log 9+\log d+\log(2/\delta)\}$, and $C_1>0$ is a universal constant.

As in the main paper, let the Jordan decomposition of A_* be $A_*=SJS^{-1}$, where J has L blocks with sizes $1\le b_1,\dots,b_L\le d$, and both J and S are $d\times d$ complex matrices. Let $b_{\max}=\max_{1\le\ell\le L}b_\ell$, and denote the condition number of S by $\mathrm{cond}(S)=\{\lambda_{\max}(S^*S)/\lambda_{\min}(S^*S)\}^{1/2}$, where S^* is the conjugate transpose of S.

LEMMA S5. *For any $A_*\in\mathbb{R}^{d\times d}$, under Assumption* 5,

$$\kappa\lesssim m\left[\log\{d\,\mathrm{cond}(S)\}+b_{\max}\log n\right].$$

Moreover, if Assumption 6 *holds, then $\kappa\lesssim m$.*

LEMMA S6. *For any $A_*\in\mathbb{R}^{d\times d}$, under Assumption* 5,

$$\log\xi\lesssim\log\{d\,\mathrm{cond}(S)/\delta\}+b_{\max}\log n.$$

Moreover, if Assumption 6′ *holds and $n\gtrsim m+\log(d/\delta)$, then $\xi\lesssim 1$.*

LEMMA S7. *For any $A_* \in \mathbb{R}^{d\times d}$, under Assumption 5,*

$$\lambda_{\max}(\Gamma_n) \lesssim d b_{\max} n^{2b_{\max}-1}\{\mathrm{cond}(S)\}^2.$$

Moreover, if Assumption 6 holds, then $\lambda_{\max}(\Gamma_n) \lesssim 1$.

LEMMA S8. *For any $A_* \in \mathbb{R}^{d\times d}$, under Assumption 5,*

$$\|\Sigma_X\|_2 \lesssim dn\sigma^2\lambda_{\max}(\Gamma_n),$$

where Σ_X is the symmetric $dn \times dn$ matrix with its (t,s)th $d\times d$ block being $\sigma^2 A_*^{t-s}\Gamma_s$ for $1 \le s \le t \le n$. Moreover, if Assumption 6′ holds, then $\|\Sigma_X\|_2 \lesssim \sigma^2$.

S3.3. *Proofs of Lemmas S5–S8*

Proof of Lemma S5. Note that

$$\begin{aligned}\kappa &= \log\left[\det\{R^{\mathrm T}(I_d\otimes\Gamma_n)R\}\det\{(R^{\mathrm T}R)^{-1}\}\right]\\ &\le \log\left[\lambda_{\max}^m(\Gamma_n)\det(R^{\mathrm T}R)\det\{(R^{\mathrm T}R)^{-1}\}\right] = m\log\lambda_{\max}(\Gamma_n).\end{aligned}$$

Thus, the upper bound of κ follows directly from Lemma S7. □

Proof of Lemma S6. First consider the case under Assumption 5. Note that $m \le dn$; see the paragraph below (3) in the main paper. Then

$$\log\psi(m,d,\delta) \lesssim \log m + \log\log(d/\delta) \lesssim \log(d/\delta) + \log n.$$

This, together with Lemmas S8 and S5, leads to the upper bound of $\log\xi$ under Assumption 5.

Suppose that Assumption 6′ holds. Then it follows from Lemmas S8 and S5 that

$$\xi \lesssim \{\psi(m,d,\delta)/n\}^{1/2} + \psi(m,d,\delta)/n.$$

Moreover, if $n \gtrsim m + \log(d/\delta)$, then $\psi(m,d,\delta)/n \lesssim 1$, and consequently $\xi \lesssim 1$. The proof of this lemma is complete. □

Proof of Lemma S7. We first prove the conclusion under Assumption 5. By the Jordan normal form of A_*, we have

$$\Gamma_n = S\sum_{s=0}^{n-1} J^s S^{-1}(S^{-1})^*(J^*)^s S^* \preceq \{\lambda_{\min}(S^*S)\}^{-1} S\sum_{s=0}^{n-1} J^s(J^*)^s S^*.$$

Hence

$$\lambda_{\max}(\Gamma_n) \le \{\mathrm{cond}(S)\}^2\lambda_{\max}\left\{\sum_{s=0}^{n-1} J^s(J^*)^s\right\}. \tag{S32}$$

For $\ell = 1,\ldots,L$, denote by J_ℓ the ℓth block of J with size b_ℓ and diagonal entries λ_ℓ. Note that the ℓth block of the block diagonal matrix $\sum_{s=0}^{n-1} J^s(J^*)^s \in \mathbb{R}^{d\times d}$ is $B_\ell = \sum_{s=0}^{n-1} J_\ell^s(J_\ell^*)^s \in \mathbb{R}^{b_\ell\times b_\ell}$. Moreover, the (i,j)th entry of J_ℓ^s is

$$(J_\ell^s)_{ij} = \begin{cases}\binom{s}{j-i}\lambda_\ell^{s-(j-i)}, & \text{if} \quad 1\le i\le j\le \min(i+s, b_\ell)\\ 0, & \text{otherwise}\end{cases},$$

where $\binom{0}{0} = 1$. Then the ith diagonal entry of B_ℓ is

$$(B_\ell)_{ii} = \sum_{s=0}^{n-1}\sum_{j=1}^{b_\ell}(J_\ell^s)_{ij}^2 = \sum_{s=0}^{n-1}\sum_{j=i}^{\min(i+s,b_\ell)}\left\{\binom{s}{j-i}|\lambda_\ell|^{s-(j-i)}\right\}^2. \tag{S33}$$

By Assumption 5, $|\lambda_\ell| \le \rho(A_*) \le 1 + c/n$ for $c > 0$. Thus

$$|\lambda_\ell|^{2\{s-(j-i)\}} \le \left(1 + \frac{c}{n}\right)^{2n}.$$

Note that $(1 + c/n)^{2n}$ monotonically increases to $\exp(2c)$ as $n \to \infty$, which implies that $|\lambda_\ell|^{2\{s-(j-i)\}}$ is uniformly bounded by a universal constant $C_2 > 0$. Moreover, for $j - i \le b_\ell - 1$ and $s < n$, $\binom{s}{j-i}$ in (S33) is uniformly bounded above by $n^{b_\ell - 1}$. As a result, for any $1 \le i \le b_\ell$ and $1 \le \ell \le L$, we have

$$(B_\ell)_{ii} \le C_2 b_\ell n^{2b_\ell - 1}.$$

Notice that the diagonal entries of $\sum_{s=0}^{n-1} J^s (J^*)^s$ are $\{(B_\ell)_{ii}\}_{1 \le i \le b_\ell, 1 \le \ell \le L}$. Therefore

$$\lambda_{\max}\left\{\sum_{s=0}^{n-1} J^s (J^*)^s\right\} \le d \max_{1 \le i \le b_\ell, 1 \le \ell \le L} (B_\ell)_{ii} \le C_2 d b_{\max} n^{2b_{\max} - 1}. \tag{S34}$$

Combining (S32) and (S34), we obtain the upper bound of $\lambda_{\max}(\Gamma_n)$ under Assumption 5 as stated in this lemma.

Next we verify the conclusion under Assumption 6. Since $\rho(A_*) \le \bar{\rho} < 1$, we have $\Gamma_n \preceq \Gamma_\infty = \sum_{s=0}^{\infty} A_*^s (A_*^{\mathrm{T}})^s < \infty$. Note that $\rho(A_*) = \lim_{s \to \infty} \|A_*^s\|_2^{1/s}$. Thus, for any $\epsilon > 0$, there exists a positive integer $n_0 = n_0(\epsilon)$ such that $\|A_*^s\|_2^{1/s} < \rho(A_*) + \epsilon$ for all $s \ge n_0$. Taking $\epsilon = \{1 - \rho(A_*)\}/2$, we have $\rho(A_*) + \epsilon = (1 + \bar{\rho})/2 < 1$. As a result,

$$\begin{aligned}\lambda_{\max}(\Gamma_n) \le \lambda_{\max}(\Gamma_\infty) \le \sum_{s=0}^{\infty} \|A_*^s\|_2^2 &\le \sum_{s=0}^{n_0 - 1} \|A_*\|_2^{2s} + \sum_{s=n_0}^{\infty} \left(\frac{1 + \bar{\rho}}{2}\right)^{2s} \\ &\le \sum_{s=0}^{n_0 - 1} C^{2s} + \left\{1 - \left(\frac{1 + \bar{\rho}}{2}\right)^2\right\}^{-1},\end{aligned}$$

where the last upper bound is a fixed constant. The proof of this lemma is complete. □

Proof of Lemma S8. The result under Assumption 5 is straightforward, since

$$\|\Sigma_X\|_2 = \lambda_{\max}(\Sigma_X) \le \mathrm{tr}(\Sigma_X) = \sigma^2 \sum_{t=1}^{n} \mathrm{tr}(\Gamma_t) \le n\sigma^2 \, \mathrm{tr}(\Gamma_n) \le dn\sigma^2 \lambda_{\max}(\Gamma_n).$$

However, showing that $\|\Sigma_X\|_2$ is bounded by a fixed constant proportional to σ^2 under Assumption $6'$ requires a much more delicate argument. This is largely because $\|\Sigma_X\|_2$ is affected by not only the growing diagonal blocks $\sigma^2\Gamma_1, \ldots, \sigma^2\Gamma_n$ but also the growing off-diagonal blocks; note that for any $1 \le t, s \le n$, the (t, s)th block of Σ_X is

$$E(X_t X_s^{\mathrm{T}}) = \begin{cases} \sigma^2 \Gamma_t (A_*^{\mathrm{T}})^{s-t}, & \text{if } t < s \\ \sigma^2 A_*^{t-s} \Gamma_s, & \text{if } t \ge s \end{cases}. \tag{S35}$$

To overcome this difficulty, under Assumption $6'$, we consider the following 'coupled' stable VAR(1) process $\{\widetilde{X}_t\}$ with independent and identically distributed innovations $\{\eta_t\}$ such that $E(\eta_t) = 0$ and $\mathrm{var}(\eta_t) = \sigma^2 I_d$, but assuming that $\widetilde{X}_t$ starts from $t = -\infty$:

$$\widetilde{X}_{t+1} = A_* \widetilde{X}_t + \eta_t, \quad t \in \mathbb{Z}. \tag{S36}$$

Unlike $\{X_t\}_{t\geq 0}$ in the main paper, this process is weakly stationary. Indeed, for any $t \in \mathbb{Z}$, it holds $E(\widetilde{X}_t) = 0$ and

$$E(\widetilde{X}_t \widetilde{X}_{t+k}^{\mathrm{T}}) = \begin{cases} \sigma^2 \Gamma_\infty (A_*^{\mathrm{T}})^k, & \text{if } k > 0 \\ \sigma^2 A_*^k \Gamma_\infty, & \text{if } k \leq 0 \end{cases}, \tag{S37}$$

where $\Gamma_\infty = \lim_{n\to\infty} \Gamma_n = \sum_{s=0}^{\infty} A_*^s (A_*^{\mathrm{T}})^s < \infty$. Analogously to Σ_X, let $\widetilde{\Sigma}_X$ be the symmetric $dn \times dn$ matrix with its (t, s)th $d \times d$ block being $E(\widetilde{X}_t \widetilde{X}_s^{\mathrm{T}})$ for $1 \leq t, s \leq n$. In other words, $\widetilde{\Sigma}_X$ is the covariance matrix of the $dn \times 1$ vector $\mathrm{vec}(\widetilde{X}^{\mathrm{T}}) = (\widetilde{X}_1^{\mathrm{T}}, \ldots, \widetilde{X}_n^{\mathrm{T}})^{\mathrm{T}}$. Note that in contrast to Σ_X, the blocks of $\widetilde{\Sigma}_X$ do not grow in the diagonal direction, in the sense that all $E(\widetilde{X}_t \widetilde{X}_s^{\mathrm{T}})$'s share the same factor matrix Γ_∞. By Basu & Michailidis (2015), for the weakly stationary VAR(1) process $\{\widetilde{X}_t\}$ in (S36), it holds

$$\|\widetilde{\Sigma}_X\|_2 \leq \frac{\sigma^2}{\mu_{\min}(\mathcal{A})} \leq \frac{\sigma^2}{\mu_1}, \tag{S38}$$

where $\mu_{\min}(\mathcal{A}) \geq \mu_1 > 0$ is defined as in Assumption $6'$.

In view of (S38) and the triangle inequality

$$\|\Sigma_X\|_2 \leq \|\widetilde{\Sigma}_X\|_2 + \|\widetilde{\Sigma}_X - \Sigma_X\|_2, \tag{S39}$$

it remains to prove that $\|\widetilde{\Sigma}_X - \Sigma_X\|_2 \lesssim \sigma^2$. To this end, for any $1 \leq t, s \leq n$, consider the difference between the (t, s)th blocks of $\widetilde{\Sigma}_X$ and Σ_X:

$$\begin{aligned} E(\widetilde{X}_t \widetilde{X}_s^{\mathrm{T}}) - E(X_t X_s^{\mathrm{T}}) &= \begin{cases} \sigma^2 (\Gamma_\infty - \Gamma_t)(A_*^{\mathrm{T}})^{s-t}, & \text{if } t < s \\ \sigma^2 A_*^{t-s} (\Gamma_\infty - \Gamma_s), & \text{if } t \geq s \end{cases} \\ &= \sigma^2 A_*^t \Gamma_\infty (A_*^{\mathrm{T}})^s. \end{aligned} \tag{S40}$$

Note that under Assumption $6'$, $\|\Gamma_\infty\|_2 \leq \sum_{t=0}^{\infty} \|A_*^t\|_2^2 \leq C^2 \sum_{t=0}^{\infty} \varrho^{2t} = \frac{C^2}{1-\varrho^2}$, where $\varrho \in (0, 1)$. This, together with (S40), implies that for any $1 \leq t, s \leq n$,

$$\|E(\widetilde{X}_t \widetilde{X}_s^{\mathrm{T}}) - E(X_t X_s^{\mathrm{T}})\|_2 \leq \sigma^2 \|A_*^t\|_2 \|\Gamma_\infty\|_2 \|A_*^s\|_2 \leq \frac{C^2 \sigma^2 \varrho^{t+s}}{1 - \varrho^2}.$$

Consequently, for any $u = (u_1^{\mathrm{T}}, \ldots, u_n^{\mathrm{T}})^{\mathrm{T}} \in \mathcal{S}^{dn-1}$ with $u_t \in \mathbb{R}^d$, we have

$$\begin{aligned} u^{\mathrm{T}} (\widetilde{\Sigma}_X - \Sigma_X) u &= \sum_{t=1}^n \sum_{s=1}^n u_t^{\mathrm{T}} \{E(\widetilde{X}_t \widetilde{X}_s^{\mathrm{T}}) - E(X_t X_s^{\mathrm{T}})\} u_s \\ &\leq \sum_{t=1}^n \sum_{s=1}^n \frac{u_t^{\mathrm{T}} \{E(\widetilde{X}_t \widetilde{X}_s^{\mathrm{T}}) - E(X_t X_s^{\mathrm{T}})\} u_s}{\|u_t\| \|u_s\|} \\ &\leq \sum_{t=1}^n \sum_{s=1}^n \|E(\widetilde{X}_t \widetilde{X}_s^{\mathrm{T}}) - E(X_t X_s^{\mathrm{T}})\|_2 \\ &\leq \frac{C^2 \sigma^2}{1 - \varrho^2} \sum_{t=1}^n \sum_{s=1}^n \varrho^{t+s} \leq \frac{C^2 \sigma^2 \varrho^2}{(1 - \varrho^2)(1 - \varrho)^2}. \end{aligned}$$

Thus,

$$\|\widetilde{\Sigma}_X - \Sigma_X\|_2 \leq \frac{C^2 \sigma^2 \varrho^2}{(1 - \varrho^2)(1 - \varrho)^2}. \tag{S41}$$

Combining (S38), (S39) and (S41), the proof of this lemma is complete. □

S4. PROOF OF THEOREM 3

We will prove claim (i) of Theorem 3 only, as claims (ii) and (iii) can be proved by a method similar to that for (i).

First, by an argument similar to that in Lütkepohl (2005, p. 199), we can show that

$$R\left\{R^{\mathrm{T}}(I_d\otimes\Gamma_k)R\right\}^{-1}R^{\mathrm{T}}\preceq I_d\otimes\Gamma_k^{-1}.$$

In addition, note that

$$\lambda_{\min}(\Gamma_k)\geq\sum_{s=0}^{k-1}\lambda_{\min}\{A_*^s(A_*^{\mathrm{T}})^s\}=\sum_{s=0}^{k-1}\sigma_{\min}^s(A_*A_*^{\mathrm{T}})\geq\sum_{s=0}^{k-1}\sigma_{\min}^{2s}(A_*).$$

As a result,

$$\lambda_{\max}\left[R\left\{R^{\mathrm{T}}(I_d\otimes\Gamma_k)R\right\}^{-1}R^{\mathrm{T}}\right]\leq\lambda_{\max}(\Gamma_k^{-1})=\frac{1}{\lambda_{\min}(\Gamma_k)}\leq\frac{1}{\sum_{s=0}^{k-1}\sigma_{\min}^{2s}(A_*)}.\tag{S42}$$

Now we prove the rate in (S.1) under condition (A.1). By the existence condition of k in (18), we can choose

$$k=\frac{c_0n}{m\left[\log\{d\,\mathrm{cond}(S)/\delta\}+b_{\max}\log n\right]},\tag{S43}$$

where $c_0>0$ is a universal constant. Then, (A.1) can be written as

$$\sigma_{\min}(A_*)\leq 1-c_2/k,\tag{S44}$$

where $c_2=c_1c_0>0$. Since

$$\frac{1}{\sum_{s=0}^{k-1}\sigma_{\min}^{2s}(A_*)}=\frac{1-\sigma_{\min}^2(A_*)}{1-\sigma_{\min}^{2k}(A_*)},$$

by Theorem 2(i) and (S42), to prove the rate for $\|\widehat{\beta}-\beta_*\|$ in (S.1), it suffices to show that there exists a universal constant $c_3\in(0,1)$ such that

$$1-\sigma_{\min}^{2k}(A_*)\geq c_3.\tag{S45}$$

Moreover, by (S44), we can show that (S45) is satisfied if

$$-2k\log(1-c_2/k)\geq-\log(1-c_3).\tag{S46}$$

Note that the function $f(k)=-2k\log(1-c_2/k)$ monotonically deceases to $2c_2$ as $k\to\infty$. Thus, by choosing c_3 such that $-\log(1-c_3)=2c_2$, i.e., $c_3=1-\exp(-2c_2)\in(0,1)$, we accomplish the proof of (S.1).

Next we prove the rate in (F.1) when the opposite of (A.1) is true, i.e., when

$$\sigma_{\min}(A_*)\geq 1-\frac{c_1m\left[\log\{d\,\mathrm{cond}(S)/\delta\}+b_{\max}\log n\right]}{n}.\tag{S47}$$

Again, we choose k in (S43), and then (S47) becomes

$$\sigma_{\min}(A_*)\geq 1-c_2/k,$$

where c_2 is defined as in (S44). Thus,

$$\sum_{s=0}^{k-1} \sigma_{\min}^{2s}(A_*) \geq \sum_{s=0}^{k-1} (1 - c_2/k)^{2s} \geq k(1 - c_2/k)^{2k}. \tag{S48}$$

In view of Theorem 2(i), (S42), (S43) and (S48), to prove the rate for $\|\widehat{\beta} - \beta_*\|$ in (F.1), we only need to show that there exists a universal constant $c_4 \in (0, 1)$ such that

$$(1 - c_2/k)^{2k} \geq c_4. \tag{S49}$$

By the choice of k in (S43), we have $k > c_2$. Hence, there exists $\epsilon > 0$ such that $k \geq c_2 + \epsilon$. Moreover, notice that the function $g(k) = (1 - c_2/k)^{2k}$ is monotonically increasing in k. As a result, by choosing $c_4 = g(c_2 + \epsilon)$, we complete the proof of (F.1).

S5. Proofs of Theorem 4 and Corollary 1

S5.1. *Two Auxiliary Lemmas*

The proof of Theorem 4 is based upon Lemmas S9 and S10 below. Denote by $\mathrm{KL}(\mathbb{Q}, \mathbb{P})$ the Kullback-Leibler divergence between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ on the same measurable space.

LEMMA S9. *Fix $\delta \in (0, 1/2)$, $\epsilon > 0$ and $R \in \mathbb{R}^{N \times m}$. Suppose that $\mathcal{N}$ is a finite subset of $\mathbb{R}^m$ such that $\|R(\theta_1 - \theta_2)\| \geq 2\epsilon, \forall \theta_1 \neq \theta_2 \in \mathcal{N}$. If*

$$\inf_{\widehat{\theta}} \sup_{\theta \in \mathcal{N}} \mathrm{pr}_\theta^{(n)} \left\{ \|R(\widehat{\theta} - \theta)\| \geq \epsilon \right\} \leq \delta, \tag{S50}$$

where the infimum is taken over all estimators of θ which are $\mathcal{F}_{n+1}$-measurable, then

$$\inf_{\theta_0 \in \mathcal{N}} \sup_{\theta \in \mathcal{N} \setminus \{\theta_0\}} \mathrm{KL}(\mathrm{pr}_\theta^{(n)}, \mathrm{pr}_{\theta_0}^{(n)}) \geq (1 - 2\delta) \log \frac{|\mathcal{N}| - 1}{2\delta}.$$

Proof of Lemma S9. For any $\mathcal{F}_{n+1}$-measurable estimator $\widehat{\theta}$, let $\mathcal{E}_\theta = \{\|R(\widehat{\theta} - \theta)\| < \epsilon\}$ for $\theta \in \mathcal{N}$. Since $\mathcal{N}$ is a 2ϵ-packing of $\mathbb{R}^m$, the events $\mathcal{E}_\theta$'s with $\theta \in \mathcal{N}$ are pairwise disjoint in $\mathcal{F}_{n+1}$. By (S50), there exists a $\widehat{\theta}$ such that $\sup_{\theta \in \mathcal{N}} \mathrm{pr}_\theta^{(n)}(\mathcal{E}_\theta^c) \leq \delta < 1/2$, i.e., $\inf_{\theta \in \mathcal{N}} \mathrm{pr}_\theta^{(n)}(\mathcal{E}_\theta) \geq 1 - \delta > 1/2$. Applying Birgé's inequality (Boucheron et al., 2013, Theorem 4.21) and an argument similar to that for Lemma F.1 in Simchowitz et al. (2018), we can readily prove that for any $\theta_0 \in \mathcal{N}$,

$$\sup_{\theta \in \mathcal{N} \setminus \{\theta_0\}} \mathrm{KL}(\mathrm{pr}_\theta^{(n)}, \mathrm{pr}_{\theta_0}^{(n)}) \geq (1 - 2\delta) \log \frac{(1 - \delta)(|\mathcal{N}| - 1)}{\delta} \geq (1 - 2\delta) \log \frac{|\mathcal{N}| - 1}{2\delta}.$$

Taking the infimum over $\theta_0 \in \mathcal{N}$, we accomplish the proof of this lemma. □

LEMMA S10. *For the linearly restricted vector autoregressive model, under the conditions of Theorem 4, for any $\theta, \theta_0 \in \mathbb{R}^m$, we have*

$$\mathrm{KL}(\mathrm{pr}_\theta^{(n)}, \mathrm{pr}_{\theta_0}^{(n)}) = \frac{1}{2}(\theta - \theta_0)^{\mathrm{T}} \Gamma_{R,n}(\theta)(\theta - \theta_0),$$

where $\Gamma_{R,n}(\theta) = \sum_{i=1}^d R_i^{\mathrm{T}} \sum_{t=1}^n \Gamma_t(\theta) R_i = R^{\mathrm{T}}\{I_d \otimes \sum_{t=1}^n \Gamma_t(\theta)\}R$.

Proof of Lemma S10. Without loss of generality, we assume that $\gamma = 0$, so that $\beta = R\theta$. Let $X_{i,t}$ be the ith entry of X_t, and denote $Z_{i,t} = R_i^{\mathrm{T}} X_t$. For any $\theta \in \mathbb{R}^m$, under $\mathrm{pr}_\theta^{(n)}$ we have $X_{i,t+1} \mid \mathcal{F}_t \sim N(\theta^{\mathrm{T}} Z_{i,t}, \sigma^2)$, where $0 \le t \le n$ and $\mathcal{F}_0 = \emptyset$. Hence, the log-likelihood of $(X_1, \ldots, X_{n+1})$ under $\mathrm{pr}_\theta^{(n)}$ is

$$\begin{aligned}
&\log \prod_{t=0}^{n} \prod_{i=1}^{d} \frac{1}{(2\pi)^{1/2}\sigma} \exp\left\{-\frac{(X_{i,t+1} - \theta^{\mathrm{T}} Z_{i,t})^2}{2\sigma^2}\right\} \\
&= -(n+1)d \log((2\pi)^{1/2}\sigma) - \frac{1}{2\sigma^2} \sum_{t=0}^{n} \sum_{i=1}^{d} (X_{i,t+1} - \theta^{\mathrm{T}} Z_{i,t})^2.
\end{aligned}$$

As a result,

$$\begin{aligned}
\mathrm{KL}(\mathrm{pr}_\theta^{(n)}, \mathrm{pr}_{\theta_0}^{(n)}) = E_{\mathrm{pr}_\theta^{(n)}} \left(\log \frac{d\mathrm{pr}_\theta^{(n)}}{d\mathrm{pr}_{\theta_0}^{(n)}}\right) &= \frac{1}{2} \sum_{t=0}^{n} \sum_{i=1}^{d} E_{\mathrm{pr}_\theta^{(n)}} \left[\{\eta_{i,t} + (\theta - \theta_0)^{\mathrm{T}} Z_{i,t}\}^2 - \eta_{i,t}^2\right] \\
&= \frac{1}{2} (\theta - \theta_0)^{\mathrm{T}} \sum_{t=1}^{n} \sum_{i=1}^{d} E_{\mathrm{pr}_\theta^{(n)}} (Z_{i,t} Z_{i,t}^{\mathrm{T}}) (\theta - \theta_0) \\
&= \frac{1}{2} (\theta - \theta_0)^{\mathrm{T}} \Gamma_{R,n}(\theta) (\theta - \theta_0),
\end{aligned}$$

where the last equality is because of $E_{\mathrm{pr}_\theta^{(n)}}(Z_{i,t} Z_{i,t}^{\mathrm{T}}) = R_i^{\mathrm{T}} \Gamma_t(\theta) R_i$. The proof is complete. □

S5.2. *Proof of Theorem* 4

Without loss of generality, we assume that $\gamma = 0$, so that $\beta = R\theta$. Define the ellipsoid $E = \{\theta \in \mathbb{R}^m : \|R\theta\| \le \bar{\rho}\} = \{(R^{\mathrm{T}} R)^{-1/2} \omega : \omega \in B(0, \bar{\rho})\}$, where $B(0, r)$ denotes the Euclidean ball in $\mathbb{R}^m$ with center zero and radius r. Since $\rho\{A(\theta)\} \le \|A(\theta)\|_F = \left(\sum_{i=1}^d \|R_i \theta\|^2\right)^{1/2} = \|R\theta\|$, we have $E \subseteq \Theta(\bar{\rho})$.

For any $\epsilon \in (0, \bar{\rho}/4]$, let $\mathcal{N}_1$ be a maximal 2ϵ-packing of $B(0, 4\epsilon)$ in $\mathbb{R}^m$, and define $\mathcal{N} = \{(R^{\mathrm{T}} R)^{-1/2} \omega : \omega \in \mathcal{N}_1\}$. Then, $\mathcal{N}$ is a 2ϵ-packing of E in the norm $\|(R^{\mathrm{T}} R)^{1/2} (\cdot)\|$. As a result, $2\epsilon \le \|R(\theta - \theta_0)\| \le 8\epsilon$ for all $\theta \ne \theta_0 \in \mathcal{N}$. In addition, by a standard volumetric argument, we have $|\mathcal{N}| = |\mathcal{N}_1| \ge 2^m$. By Lemma S9, for any $\delta \in (0, 1/2)$, this theorem holds if

$$\inf_{\theta_0 \in \mathcal{N}} \sup_{\theta \in \mathcal{N} \setminus \{\theta_0\}} \mathrm{KL}(\mathrm{pr}_\theta^{(n)}, \mathrm{pr}_{\theta_0}^{(n)}) < (1 - 2\delta) \log \frac{|\mathcal{N}| - 1}{2\delta}. \tag{S51}$$

Since $\sum_{t=1}^n \Gamma_t(\theta) \preceq n\Gamma_n(\theta)$ for any $\theta \in \mathbb{R}^m$, and

$$\sup_{\theta \in \Theta(\bar{\rho})} \lambda_{\max}\{\Gamma_n(\theta)\} \le \sum_{s=0}^{n-1} \lambda_{\max}[A^s(\theta)\{A^{\mathrm{T}}(\theta)\}^s] \le \sum_{s=0}^{n-1} \bar{\rho}^{2s} = \gamma_n(\bar{\rho}),$$

it follows from Lemma S10 that

$$\begin{aligned}
\max_{\theta, \theta_0 \in \mathcal{N}} \mathrm{KL}(\mathrm{pr}_\theta^{(n)}, \mathrm{pr}_{\theta_0}^{(n)}) &\le \frac{1}{2} \max_{\theta, \theta_0 \in \mathcal{N}} (\theta - \theta_0)^{\mathrm{T}} \Gamma_{R,n}(\theta) (\theta - \theta_0) \\
&\le \frac{n}{2} \max_{\theta, \theta_0 \in \mathcal{N}} (\theta - \theta_0)^{\mathrm{T}} R^{\mathrm{T}} \{I_d \otimes \Gamma_n(\theta)\} R (\theta - \theta_0) \\
&\le \frac{n}{2} \max_{\theta, \theta_0 \in \mathcal{N}} \|R(\theta - \theta_0)\|^2 \sup_{\theta \in \Theta(\bar{\rho})} \lambda_{\max}\{\Gamma_n(\theta)\} \\
&\le 32\epsilon^2 n \gamma_n(\bar{\rho})
\end{aligned}$$

As a result, a sufficient condition for (S51) is

$$n\gamma_n(\bar\rho) < \frac{(1-2\delta)}{32\epsilon^2}\log\frac{2^m}{4\delta}. \tag{S52}$$

In particular, for any $\delta \in (0, 1/4)$, we can show that there exists a universal constant $c > 0$ such that the right-hand side of (S52) is bounded below by $c\{m + \log(1/\delta)\}/\epsilon^2$, i.e., the conclusion of this theorem follows.

S5.3. *Proof of Corollary* 1

Under the conditions of Theorem 4, we have

$$\inf_{\widehat\beta}\sup_{\theta\in\Theta(\bar\rho)} \mathrm{pr}_\theta^{(n)}\left[\|\widehat\beta-\beta\| \geq C\left\{\frac{m+\log(1/\delta)}{n\gamma_n(\bar\rho)}\right\}^{1/2}\right] \geq \delta,$$

where $C > 0$ is fixed. It then suffices to derive lower bounds of $1/\gamma_n(\bar\rho)$ for $\bar\rho \in (0, \infty)$.

First, suppose that $\bar\rho \in (0, 1)$. Then we have $\gamma_n(\bar\rho) = \sum_{s=0}^{n-1}\bar\rho^{2s} = (1-\bar\rho^{2n})/(1-\bar\rho^2) < \min\{n, (1-\bar\rho^2)^{-1}\}$, and therefore

$$\frac{1}{\gamma_n(\bar\rho)} > \begin{cases} 1-\bar\rho^2, & \text{if} \quad \bar\rho \in (0, (1-1/n)^{1/2}) \\ 1/n, & \text{if} \quad \bar\rho \in [(1-1/n)^{1/2}, 1) \end{cases}. \tag{S53}$$

Next, suppose that $\bar\rho \in [1, 1 + c/n]$ for a fixed $c > 0$. Then

$$\frac{\gamma_n(\bar\rho)}{n} = \frac{1}{n}\sum_{s=0}^{n-1}\bar\rho^{2s} \leq \frac{1}{n}\sum_{s=0}^{n-1}(1+c/n)^{2s} \leq (1+c/n)^{2n}.$$

Since $(1 + c/n)^{2n}$ monotonically increases to $\exp(2c)$ as $n \to \infty$, there exists a constant $C_2 > 0$ free of n such that $\gamma_n(\bar\rho)/n$ is uniformly bounded above by C_2, i.e.,

$$\frac{1}{\gamma_n(\bar\rho)} \geq \frac{1}{C_2}n^{-1} \quad \text{if} \quad \bar\rho \in [1, 1+c/n]. \tag{S54}$$

Moreover, for any $\bar\rho \in (1, \infty)$, we have

$$\frac{1}{\gamma_n(\bar\rho)} = \frac{\bar\rho^2-1}{\bar\rho^{2n}-1} > \frac{\bar\rho^2-1}{\bar\rho^{2n}}. \tag{S55}$$

Combining (S53)–(S55), we have

$$\left\{\frac{m}{n\gamma_n(\bar\rho)}\right\}^{1/2} \geq \begin{cases} \left\{(1-\bar\rho^2)m/n\right\}^{1/2}, & \text{if} \quad \bar\rho \in (0, (1-1/n)^{1/2}) \\ m^{1/2}/n, & \text{if} \quad \bar\rho \in [(1-1/n)^{1/2}, 1+c/n] \\ \bar\rho^{-n}\left\{(\bar\rho^2-1)m/n\right\}^{1/2}, & \text{if} \quad \bar\rho \in (1+c/n, \infty) \end{cases},$$

and this completes the proof of Corollary 1.

References

BASU, S. & MICHAILIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics* **43**, 1535–1567.

BOUCHERON, S., LUGOSI, G. & MASSART, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford: Oxford University Press.

HORN, R. A. & JOHNSON, C. R. (1985). *Matrix Analysis*. New York: Cambridge University Press.

LÜTKEPOHL, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag Berlin Heidelberg.

RUDELSON, M. & VERSHYNIN, R. (2015). Small ball probabilities for linear images of high-dimensional distributions. *International Mathematics Research Notices* **2015**, 9594–9617.

SIMCHOWITZ, M., MANIA, H., TU, S., JORDAN, M. & RECHT, B. (2018). Learning without mixing: Towards a sharp analysis of linear system identification. In *Proceedings of Machine Learning Research*, vol. 75. 31st Annual Conference on Learning Theory.

VERSHYNIN, R. (2012). Introduction to the nonasymptotic analysis of random matrices. In *Compressed Sensing: Theory and Applications*, Y. Eldar & G. Kutyniok, eds., chap. 5. Cambridge University Press, pp. 210–268.

VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge: Cambridge University Press.