Supplementary Material to Robust Estimation of Transition Matrices in High Dimensional Heavy-tailed Vector Autoregressive Processes

Huitong Qiu HQIU7@JHU.EDU
Johns Hopkins University, 615 N. Wolfe St., Baltimore, MD 21210 USA

Sheng Xu SHXU@JHU.EDU
Johns Hopkins University, 615 N. Wolfe St., Baltimore, MD 21210 USA

Fang Han FHAN@JHU.EDU
Johns Hopkins University, 615 N. Wolfe St., Baltimore, MD 21210 USA

Han Liu HANLIU@PRINCETON.EDU
Princeton University, 98 Charlton Street, Princeton, NJ 08544 USA

Brian Caffo BCAFFO@JHU.EDU
Johns Hopkins University, 615 N. Wolfe St., Baltimore, MD 21210 USA

A. Technical Proofs

A.1. Supporting Lemmas

Lemma A.1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be an absolutely continuous stationary process with ϕ-mixing coefficient $\phi(n)$. Define U-statistic*

$$U_T(K_u) := \frac{2}{T(T-1)}\sum_{1\leqslant s<t\leqslant T} K_u(X_s, X_t), \qquad \text{(A.1)}$$

for kernel function $K_u(x,y) := I(|x-y| \leqslant u)$. Let $\widetilde{X}_1$ be an independent copy of X_1, and $G(u) := \mathbb{P}(|X_1 - \widetilde{X}_1| \leqslant u)$ be the distribution function of $|X_1 - \widetilde{X}|$. Then, we have

$$\left|\mathbb{E}U_T(K_u) - G(u)\right| \leqslant \frac{2\phi_T}{T},$$

for any $u > 0$, where $\phi_T = \sum_{n=1}^T \phi(n)$.

Proof. Denote $G_{st}(u) := \mathbb{P}(|X_s - X_t| \leqslant u)$ to be the distribution function of $|X_s - X_t|$ for $s < t$. Let $M > 0$ be a constant and

$$-M = a_{-h}^{(h)} < \cdots < a_0^{(h)} < \cdots < a_h^{(h)} = M$$

be a sequence of real numbers satisfying

$$\max_{-h<k\leqslant h}(a_k^{(h)} - a_{k-1}^{(h)}) \leqslant u,$$
$$\lim_{h\to\infty}\max_{-h<k\leqslant h}(a_k^{(h)} - a_{k-1}^{(h)}) = 0. \qquad \text{(A.2)}$$

Given $X_s \in [a_{k-1}^{(h)}, a_k^{(h)}]$, we have that $|X_s - X_t| \leqslant u$ implies $X_t \in [a_{k-1}^h - u, a_k^{(h)} + u]$. Thus, we have

$$\begin{aligned}
&\mathbb{P}(|X_s - X_t| \leqslant u, X_s \in [-M, M]) \\
&= \sum_{-h<k\leqslant h}\mathbb{P}(|X_s - X_t| \leqslant u \mid X_s \in [a_{k-1}, a_k])\mathbb{P}(X_s \in [a_{k-1}, a_k]) \\
&\leqslant \sum_{-h<k\leqslant h}\mathbb{P}(X_t \in [a_{k-1}^{(h)} - u, a_k^{(h)} + u] \mid X_s \in [a_{k-1}, a_k])\cdot \\
&\qquad \mathbb{P}(X_s \in [a_{k-1}, a_k]). \qquad \text{(A.3)}
\end{aligned}$$

On the other hand, given $X_s \in [a_{k-1}^{(h)}, a_k^{(h)}]$, we have $X_t \in [a_k^{(h)} - u, a_{k-1}^{(h)} + u]$ implies $|X_s - X_t| \leqslant u$. Thus, we have

$$\begin{aligned}
&\mathbb{P}(|X_s - X_t| \leqslant u, X_s \in [-M, M]) \\
&= \sum_{-h<k\leqslant h}\mathbb{P}(|X_s - X_t| \leqslant u \mid X_s \in [a_{k-1}, a_k])\mathbb{P}(X_s \in [a_{k-1}, a_k]) \\
&\geqslant \sum_{-h<k\leqslant h}\mathbb{P}(X_t \in [a_k^{(h)} - u, a_{k-1}^{(h)} + u] \mid X_s \in [a_{k-1}, a_k])\cdot \\
&\qquad \mathbb{P}(X_s \in [a_{k-1}, a_k]). \qquad \text{(A.4)}
\end{aligned}$$

Now define $\psi_h^U := \sum_{-h<k\leqslant h}\mathbb{P}(X_t \in [a_{k-1}^{(h)} - u, a_k^{(h)} + u])\mathbb{P}(X_s \in [a_{k-1}, a_k])$, $\psi_h^L := \sum_{-h<k\leqslant h}\mathbb{P}(X_t \in [a_k^{(h)} - u, a_{k-1}^{(h)} + u])\mathbb{P}(X_s \in [a_{k-1}, a_k])$, and

$$\psi_h := \begin{cases} \psi_h^L, & \text{if } \mathbb{P}(|X_s - X_t| \leqslant u, X_s \in [-M, M]) > \psi_h^L; \\ \psi_h^U, & \text{otherwise.} \end{cases}$$

Note that $\psi_h^L \leqslant \psi_h^U$. If $\mathbb{P}(|X_s - X_t| \leqslant u, X_s \in [-M, M]) > \psi_h^L$, by the definition of ψ_h and (A.3), we have

$$\begin{aligned}
&|\mathbb{P}(|X_s - X_t| \leqslant u, X_s \in [-M, M]) - \psi_h| \\
=&\mathbb{P}(|X_s - X_t| \leqslant u, X_s \in [-M, M]) - \psi_h^L \\
\leqslant& \sum_{-h<k\leqslant h} \big|\mathbb{P}(X_t \in [a_{k-1}^{(h)} - u, a_k^{(h)} + u] \mid X_s \in [a_{k-1}, a_k]) - \\
&\quad \mathbb{P}(X_t \in [a_k^{(h)} - u, a_{k-1}^{(h)} + u])\big|\mathbb{P}(X_s \in [a_{k-1}, a_k]) \\
\leqslant& \sum_{-h<k\leqslant h} \big|\mathbb{P}(X_t \in [a_{k-1}^{(h)} - u, a_k^{(h)} + u] \mid X_s \in [a_{k-1}, a_k]) - \\
&\quad \mathbb{P}(X_t \in [a_{k-1}^{(h)} - u, a_k^{(h)} + u])\big|\mathbb{P}(X_s \in [a_{k-1}, a_k]) \\
&+ \sum_{-h<k\leqslant h} \big|\mathbb{P}(X_t \in [a_{k-1}^{(h)} - u, a_k^{(h)} + u]) - \mathbb{P}(X_t \in \\
&\quad [a_k^{(h)} - u, a_{k-1}^{(h)} + u])\big|\mathbb{P}(X_s \in [a_{k-1}, a_k]) \\
\leqslant&\phi(t-s) + \max_{-h<k\leqslant h}\big|\mathbb{P}(X_t \in [a_{k-1}^{(h)} - u, a_k^{(h)} + u]) - \\
&\quad \mathbb{P}(X_t \in [a_k^{(h)} - u, a_{k-1}^{(h)} + u])\big|. \qquad \text{(A.5)}
\end{aligned}$$

On the other hand, if $\mathbb{P}(|X_s - X_t| \leqslant u, X_s \in [-M, M]) \leqslant \psi_h^L$, since $\psi_h^L \leqslant \psi_h^U$, by the definition of ψ_h and (A.4), we have

$$\begin{aligned}
&|\mathbb{P}(|X_s - X_t| \leqslant u, X_s \in [-M, M]) - \psi_h| \\
=&\psi_h^U - \mathbb{P}(|X_s - X_t| \leqslant u, X_s \in [-M, M]) \\
\leqslant& \sum_{-h<k\leqslant h} \big|\mathbb{P}(X_t \in [a_{k-1}^{(h)} - u, a_k^{(h)} + u]) - \mathbb{P}(X_t \in [a_k^{(h)} - u, \\
&\quad a_{k-1}^{(h)} + u] \mid X_s \in [a_{k-1}, a_k])\big|\mathbb{P}(X_s \in [a_{k-1}, a_k]) \\
\leqslant& \sum_{-h<k\leqslant h} \big|\mathbb{P}(X_t \in [a_k^{(h)} - u, a_{k-1}^{(h)} + u] \mid X_s \in [a_{k-1}, a_k]) - \\
&\quad \mathbb{P}(X_t \in [a_k^{(h)} - u, a_{k-1}^{(h)} + u])\big|\mathbb{P}(X_s \in [a_{k-1}, a_k]) \\
&+ \sum_{-h<k\leqslant h} \big|\mathbb{P}(X_t \in [a_k^{(h)} - u, a_{k-1}^{(h)} + u]) - \mathbb{P}(X_t \in \\
&\quad [a_{k-1}^{(h)} - u, a_k^{(h)} + u])\big|\mathbb{P}(X_s \in [a_{k-1}, a_k]) \\
\leqslant&\phi(t-s) + \max_{-h<k\leqslant h}\big|\mathbb{P}(X_t \in [a_k^{(h)} - u, a_{k-1}^{(h)} + u]) - \\
&\quad \mathbb{P}(X_t \in [a_{k-1}^{(h)} - u, a_k^{(h)} + u])\big|. \qquad \text{(A.6)}
\end{aligned}$$

Thus, combining (A.5) and (A.6), we have

$$\begin{aligned}
&|\mathbb{P}(|X_s - X_t| \leqslant u, X_s \in [-M, M]) - \psi_h| \\
\leqslant&\phi(t-s) + \max_{-h<k\leqslant h}\big|\mathbb{P}(X_t \in [a_k^{(h)} - u, a_{k-1}^{(h)} + u]) - \\
&\quad \mathbb{P}(X_t \in [a_{k-1}^{(h)} - u, a_k^{(h)} + u])\big|.
\end{aligned}$$

Let $h \to \infty$. Using (A.2) and the assumption that X_t is absolutely continuous, we have

$$\begin{aligned}
&\Big|\mathbb{P}(|X_s - X_t| \leqslant u, X_s \in [-M, M]) - \\
&\quad \int_{-M}^{M} \mathbb{P}(X_s \in [a-u, a+u]) d\mathbb{P}(X_s = a)\Big| \leqslant \phi(t-s).
\end{aligned}$$

Now, let $M \to \infty$, we further obtain

$$\begin{aligned}
&\Big|\mathbb{P}(|X_s - X_t| \leqslant u) - \int \mathbb{P}(X_s \in [a-t, a+t]) d\mathbb{P}(X_s = a)\Big| \\
&\leqslant \phi(t-s).
\end{aligned}$$

Noting that

$$\begin{aligned}
&\int \mathbb{P}(X_s \in [a-u, a+u]) d\mathbb{P}(X_s = a) \\
=&\int \mathbb{P}(X_s \in [a-u, a+u]) d\mathbb{P}(\widetilde{X} = a) \\
=&\mathbb{P}(|X_1 - \widetilde{X}| \leqslant u) = G(u),
\end{aligned}$$

we have $\Big|\mathbb{P}(|X_s - X_t| \leqslant u) - G(u)\Big| \leqslant \phi(t-s)$. Hence, we have

$$\begin{aligned}
&\big|\mathbb{E}U_T(\phi_u) - G(u)\big| \\
\leqslant&\frac{2}{T(T-1)} \sum_{1\leqslant s<t\leqslant T} \big|\mathbb{P}(|X_s - X_t| \leqslant u) - G(u)\big| \\
\leqslant&\frac{2}{T(T-1)} \sum_{1\leqslant s<t\leqslant T} \phi(t-s) \\
=&\frac{2}{T(T-1)} \sum_{k=1}^{T-1} (T-k)\phi(k) = \frac{2\phi_T}{T}.
\end{aligned}$$

This completes the proof. □

Lemma A.2. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary process with ϕ-mixing coefficient $\phi(n)$, and $U_T(K_u)$ be defined in (A.1). Then, for any $u > 0$, we have*

$$\mathbb{P}\{|U_T(K_u) - \mathbb{E}U_T(K_u)| \geqslant \tau\} \leqslant 2\exp\Big\{-\frac{T\tau^2}{2(1+2\phi_T)}\Big\}$$

for any $\tau > 0$, where $\phi_T = \sum_{n=1}^{T} \phi(n)$.

The following lemma is needed for proving Lemma A.2

Lemma A.3. (Kontorovich et al., 2008; Mohri & Rostamizadeh, 2010) *Let $f : \Omega^T \to \mathbb{R}$ be a measurable function that is M-Lipschitz with respect to the Hamming metric for some $M > 0$:*

$$\sup_{x_1,\ldots,x_T,x_t'} \big|f(x_1,\ldots,x_t,\ldots,x_T) - f(x_1,\ldots,x_t',\ldots,x_T)\big| \leqslant M.$$

Then, for a stationary process $\{X_t\}_{t\in\mathbb{Z}}$ with ϕ-mixing coefficient $\phi(n)$, we have

$$\begin{aligned}
&\mathbb{P}\{|f(X_1,\ldots,X_T) - \mathbb{E}f(X_1,\ldots,X_T)| \geqslant \tau\} \\
&\leqslant 2\exp\Big[-\frac{2\tau^2}{M^2 T\{1 + 2\sum_{k=1}^{T} \phi(k)\}}\Big].
\end{aligned}$$

for any $\tau > 0$.

Proof of Lemma A.2. Let

$$f(x_1,\ldots,x_T):=TU_T(K_u)=\frac{2}{T-1}\sum_{s<t}I(|x_s-x_t|\leqslant u).$$

since replacing an element in $(x_1,\ldots,x_T)$, say, x_t, by x_t' only affects $T-1$ terms in the summation above, we have

$$\left|f(x_1,\ldots,x_t,\ldots,x_T)-f(x_1,\ldots,x_t',\ldots,x_T)\right|\leqslant 2.$$

Thus, by Lemma A.3, we have

$$\mathbb{P}\{T|U_T(K_u)-\mathbb{E}U_T(K_u)|\geqslant\tau\} \leqslant 2\exp\Big[-\frac{\tau^2}{2T\{1+2\sum_{k=1}^T\phi(k)\}}\Big]$$

for any $\tau>0$. Replacing τ with $T\tau$ in the above equation, we obtain

$$\mathbb{P}\{|U_T(K_u)-\mathbb{E}U_T(K_u)|\geqslant\tau\} \leqslant 2\exp\Big[-\frac{T\tau^2}{2\{1+2\phi_T\}}\Big]$$

This completes the proof. □

Lemma A.4. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be an absolutely continuous stationary process with ϕ-mixing coefficient $\phi(n)$. Let $U_T(K_u)$ and $G(u)$ be defined as in Lemma A.1. Then, for any $u>0$, we have*

$$\mathbb{P}\{|U_T(K_u)-G(u)|\geqslant\tau\}\leqslant 2\exp\Big\{-\frac{T}{2(1+2\phi_T)}\Big(\tau-\frac{2\phi_T}{T}\Big)^2\Big\}$$

for $\tau>2\phi_T/T$ and $\phi_T=\sum_{n=1}^T\phi(n)$.

Proof. Using Lemma A.1, we have

$$\begin{aligned}&\mathbb{P}\{|U_T(K_u)-G(u)|\geqslant\tau\}\\ \leqslant&\mathbb{P}\{|U_T(K_u)-\mathbb{E}U_T(K_u)|+|\mathbb{E}U_T(K_u)-G(u)|\geqslant\tau\}\\ \leqslant&\mathbb{P}\Big\{|U_T(K_u)-\mathbb{E}U_T(K_u)|\geqslant\tau-\frac{2\phi_T}{T}\Big\}.\end{aligned}$$

Applying Lemma A.2 completes the proof. □

Lemma A.5. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary process with ϕ-mixing coefficient $\phi(n)$. Let $\widetilde{X}_1$ be an independent copy of X_1, and $q\in[0,1]$ be an absolute constant. Suppose the following assumptions hold:*

1. *$Q(|X_1-\widetilde{X}_1|;q)$ and $\widehat{Q}(\{|X_s-X_t|\}_{1\leqslant s<t\leqslant T};q)$ are unique with probability 1.*
2. *There exist constants $\kappa>0$ and $\eta>0$ such that*

$$\inf_{|y-Q(|X_1-\widetilde{X}_1|;q)|\leqslant\kappa}\frac{d}{dy}G(y)\geqslant\eta,$$

where G is the distribution function of $|X_1-\widetilde{X}_1|$.

Then, we have

$$\begin{aligned}&\mathbb{P}\big[|\widehat{Q}(\{|X_s-X_t|\}_{1\leqslant s<t\leqslant T};q)-Q(|X_1-\widetilde{X}_1|;q)|\geqslant u\big]\\ &\leqslant 2\exp\Big\{-\frac{T}{2(1+2\phi_T)}\Big(\eta u-\frac{4\phi_T}{T}\Big)^2\Big\},\end{aligned}\tag{A.7}$$

when $4\phi_T/(\eta T)\leqslant u\leqslant\kappa$. Here $\phi_T=\sum_{n=1}^T\phi(n)$.

Proof. We denote by G_T the empirical distribution function of $\{|X_s-X_t|\}_{1\leqslant s<t\leqslant T}$. G_T is non-decreasing and since $\widehat{Q}(\{|X_s-X_t|\}_{1\leqslant s<t\leqslant T};q)$ is unique, we have

$$q\leqslant G_T\big\{\widehat{Q}(\{|X_s-X_t|\}_{1\leqslant s<t\leqslant T};q)\big\}\leqslant q+\frac{2}{T(T-1)}.$$

Denote $G^{-1}(q)=Q(|X_1-\widetilde{X}_1|;q)$. Since $Q(|X_1-\widetilde{X}_1|;q)$ is unique, we have $G\{G^{-1}(q)\}=q$. Thus, we have

$$\begin{aligned}&\mathbb{P}\big[\widehat{Q}(\{|X_s-X_t|\}_{1\leqslant s<t\leqslant T};q)-Q(|X_1-\widetilde{X}_1|;q)\geqslant u\big]\\ &\leqslant\mathbb{P}\big[G_T\{\widehat{Q}(\{|X_s-X_t|\}_{1\leqslant s<t\leqslant T};q)\}\geqslant G_T\{G^{-1}(q)+u\}\big]\\ &\leqslant\mathbb{P}\big[q+\frac{2}{T(T-1)}\geqslant U_T\{\psi_{G^{-1}(q)+u}\}\big]\\ &=\mathbb{P}\big[-U_T\{\psi_{G^{-1}(q)+u}\}+G\{G^{-1}(q)+u\}\geqslant\\ &\qquad G\{G^{-1}(q)+u\}-q-\frac{2}{T(T-1)}\big],\end{aligned}$$

where $U_T\{\psi_{G^{-1}(q)+u}\}$ is defined in Lemma A.1. By Assumption 2, we have $G\{G^{-1}(q)+u\}-q\leqslant\eta$ when $u\leqslant\kappa$. Now, using Lemma A.1, we have

$$\begin{aligned}&\mathbb{P}\big[\widehat{Q}(\{|X_s-X_t|\}_{1\leqslant s<t\leqslant T};q)-Q(|X_1-\widetilde{X}_1|;q)\geqslant u\big]\\ &\leqslant\mathbb{P}\big[|U_T\{\psi_{G^{-1}(q)+u}\}-G\{G^{-1}(q)+u\}|\geqslant\eta u-\frac{2}{T(T-1)}\big]\\ &\leqslant 2\exp\Big[-\frac{T}{2(1+2\phi_T)}\Big\{\eta u-\frac{2}{T(T-1)}-\frac{2\phi_T}{T}\Big\}^2\Big]\\ &\leqslant 2\exp\Big[-\frac{T}{2(1+2\phi_T)}\Big\{\eta u-\frac{4\phi_T}{T}\Big\}^2\Big],\end{aligned}\tag{A.8}$$

provided that $4\phi_T/(\eta T)\leqslant u\leqslant\kappa$. On the other hand, using the same technique, we have

$$\begin{aligned}&\mathbb{P}\big[\widehat{Q}(\{|X_s-X_t|\}_{1\leqslant s<t\leqslant T};q)-Q(|X_1-\widetilde{X}|;q)\leqslant -u\big]\\ &\leqslant\mathbb{P}\big[G_T\{\widehat{Q}(\{|X_s-X_t|\}_{1\leqslant s<t\leqslant T};q)\}\leqslant G_T\{G^{-1}(q)-u\}\big]\\ &\leqslant\mathbb{P}\big[U_T\{\psi_{G^{-1}(q)+u}\}-G\{G^{-1}(q)-u\}\geqslant q-G\{G^{-1}(q)-u\}\big]\\ &\leqslant\mathbb{P}\big[|U_T\{\psi_{G^{-1}(q)+u}\}-G\{G^{-1}(q)-u\}|\geqslant\eta u\big]\\ &\leqslant 2\exp\Big\{-\frac{T}{2(1+2\phi_T)}\Big(\eta u-\frac{2\phi_T}{T}\Big)^2\Big\},\end{aligned}\tag{A.9}$$

provided that $2\phi_T/(\eta T)\leqslant u\leqslant\kappa$. Combining (A.8) and (A.9) completes the proof. □

A.2. Proof of Main Results

A.2.1. Proof of Lemma 3.4

Proof. Part 1 of Lemma 3.4 is immediate by the definition of $\boldsymbol{L}$ and Proposition 2.3.

To prove Part 2, we start with sufficiency. Suppose Equations (3.6) - (3.9) hold. Since $(\boldsymbol{X}_t^\mathsf{T}, \boldsymbol{E}_{t+1}^\mathsf{T})^\mathsf{T}$ can be obtained by a linear transformation of $\boldsymbol{L}$, by Proposition 2.3, we have

$$\begin{pmatrix} \boldsymbol{X}_t \\ \boldsymbol{E}_{t+1} \end{pmatrix} \sim \mathrm{EC}_{2d}\Big(\mathbf{0}, \begin{pmatrix} [\boldsymbol{\Omega}_X]_{I_t I_t} & [\boldsymbol{\Omega}_{XE}]_{I_t I_t} \\ [\boldsymbol{\Omega}_{XE}]_{I_t I_t}^\mathsf{T} & [\boldsymbol{\Omega}_E]_{I_t I_t} \end{pmatrix}, \xi\Big)$$

for some random variable ξ. By (3.6) and (3.5), we have $[\boldsymbol{\Omega}_X]_{I_t I_t} = \boldsymbol{\Sigma}$ and $[\boldsymbol{\Omega}_E]_{I_t I_t} = \boldsymbol{\Psi}$; by (3.9), we have $[\boldsymbol{\Omega}_{XE}]_{I_t I_t} = \mathbf{0}$. Thus, Condition 2 in Definition 3.1 hold.

To prove necessity, suppose that $\boldsymbol{L}$ satisfies Condition 2 in Definition 3.1. By stationarity of $\{\boldsymbol{X}_t\}_{t=1}^T$, the diagonal blocks of $\boldsymbol{\Omega}_X$ equal $\boldsymbol{\Sigma}$. Thus, we have (3.6). On the other hand, since $\boldsymbol{L} = \mathbf{B}\boldsymbol{L}_0$ and $\boldsymbol{L}_0 \sim \mathrm{EC}_{Td}(\mathbf{0}, \mathrm{diag}(\boldsymbol{\Sigma}, \boldsymbol{\Psi}, \ldots, \boldsymbol{\Psi}), \zeta)$, by Proposition 2.3, we have $\boldsymbol{\Omega} = \mathbf{B}\mathrm{diag}(\boldsymbol{\Sigma}, \boldsymbol{\Psi}, \ldots, \boldsymbol{\Psi})\mathbf{B}^\mathsf{T}$. Plugging in the definition of $\mathbf{B}$, we have (3.9). Comparing the leading T diagonal blocks of $\mathbf{B}\mathrm{diag}(\boldsymbol{\Sigma}, \boldsymbol{\Psi}, \ldots, \boldsymbol{\Psi})\mathbf{B}^\mathsf{T}$ with those of $\boldsymbol{\Omega}_X$, we have (3.7). Plugging (3.7) into the off-diagonal blocks of $\mathbf{B}\mathrm{diag}(\boldsymbol{\Sigma}, \boldsymbol{\Psi}, \ldots, \boldsymbol{\Psi})\mathbf{B}^\mathsf{T}$, we obtain (3.8). This completes the proof. □

A.2.2. Proof of Theorem 3.6

Proof. To prove Theorem 3.6, we first introduce an equivalent definition of elliptical random vectors. Specifically, $\boldsymbol{X}$ is an elliptical random vector with location $\boldsymbol{\mu}$ and scatter $\mathbf{S}$ if and only if the characteristic function of $\boldsymbol{X}$ is $\psi_{\boldsymbol{X}}(\boldsymbol{t}) = \exp(i\boldsymbol{t}^\mathsf{T}\boldsymbol{\mu})\varphi(\boldsymbol{t}^\mathsf{T}\mathbf{S}\boldsymbol{t})$ for some function φ (Fang et al., 1990).

Let $\boldsymbol{R} := (\boldsymbol{X}_1^\mathsf{T}, \boldsymbol{X}_2^\mathsf{T})^\mathsf{T}$ and the characteristic function of $\boldsymbol{R}$ be $\psi_{\boldsymbol{R}}(\boldsymbol{t}) = \varphi(\boldsymbol{t}^\mathsf{T}\boldsymbol{\Theta}\boldsymbol{t})$, where

$$\boldsymbol{\Theta} = \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}_1 \\ \boldsymbol{\Sigma}_1^\mathsf{T} & \boldsymbol{\Sigma} \end{pmatrix}.$$

Suppose $\widetilde{\boldsymbol{R}} = (\widetilde{\boldsymbol{X}}_1^\mathsf{T}, \widetilde{\boldsymbol{X}}_2^\mathsf{T})^\mathsf{T}$ is an independent copy of $\boldsymbol{R}$. The characteristic function of $\boldsymbol{R} - \widetilde{\boldsymbol{R}}$ is

$$\begin{aligned}\psi_{\boldsymbol{R}-\widetilde{\boldsymbol{R}}}(\boldsymbol{t}) &= \mathbb{E}\exp\{i\boldsymbol{t}^\mathsf{T}(\boldsymbol{R} - \widetilde{\boldsymbol{R}})\} \\ &= \mathbb{E}\exp(i\boldsymbol{t}^\mathsf{T}\boldsymbol{R})\mathbb{E}\exp(i\boldsymbol{t}^\mathsf{T}\widetilde{\boldsymbol{R}}) = \varphi(\boldsymbol{t}^\mathsf{T}\boldsymbol{\Theta}\boldsymbol{t})^2.\end{aligned}$$

Thus, $\boldsymbol{R} - \widetilde{\boldsymbol{R}}$ is also an elliptical random vector with scatter $\boldsymbol{\Theta}$. Suppose $\boldsymbol{R} - \widetilde{\boldsymbol{R}} \sim \mathrm{EC}_{2d}(\mathbf{0}, \boldsymbol{\Theta}, \nu)$. Let $r_{\boldsymbol{\Theta}} = \mathrm{rank}(\boldsymbol{\Theta})$ be the rank of $\boldsymbol{\Theta}$. For any $j \in \{1, \ldots, d\}$, since $X_{1j} - \widetilde{X}_{1j}$ can be obtained by a linear transformation of $\boldsymbol{R} - \widetilde{\boldsymbol{R}}$, by Proposition 2.3, we have

$$X_{1j} - \widetilde{X}_{1j} \sim \mathrm{EC}_1(0, \boldsymbol{\Sigma}_{jj}, \nu\sqrt{D}),$$

where $D \sim \mathrm{Beta}(1/2, (r_{\boldsymbol{\Theta}} - 1)/2)$ is a Beta random variable. Thus, we have

$$(X_{1j} - \widetilde{X}_{1j}) \overset{\mathsf{d}}{=} \sqrt{\boldsymbol{\Sigma}_{jj}}\nu\sqrt{D}.$$

By the definition of $\mathbf{R}_{jj}^{\mathrm{Q}}$, we have

$$\begin{aligned}\mathbf{R}_{jj}^{\mathrm{Q}} &= Q(|X_{1j} - \widetilde{X}_{1j}|; 1/4)^2 = Q\{(X_{1j} - \widetilde{X}_{1j})^2; 1/4\} \\ &= \boldsymbol{\Sigma}_{jj}Q(\nu^2 D; 1/4).\end{aligned} \tag{A.10}$$

Now, for $j, k, j', k' \in \{1, \ldots, d\}$ and $j \neq k$, let

$$\begin{aligned}&X_{jk}^+ := X_{1j} + X_{1k},\ X_{jk}^- := X_{1j} - X_{1k}, \\ &\widetilde{X}_{jk}^+ := \widetilde{X}_{1j} + \widetilde{X}_{1k},\ \widetilde{X}_{jk}^- := \widetilde{X}_{1j} - \widetilde{X}_{1k}, \\ &Y_{j'k'}^+ := X_{1j'} + X_{2k'},\ Y_{j'k'}^- := X_{1j'} - X_{2k'}, \\ &\widetilde{Y}_{j'k'}^+ := \widetilde{X}_{1j'} + \widetilde{X}_{2k'},\ \widetilde{Y}_{j'k'}^- := \widetilde{X}_{1j'} - \widetilde{X}_{2k'}.\end{aligned}$$

Apply Proposition 2.3, we have

$$\begin{aligned}&X_{jk}^+ - \widetilde{X}_{jk}^+ \sim \mathrm{EC}_1(0, \boldsymbol{\Sigma}_{jj} + \boldsymbol{\Sigma}_{kk} + 2\boldsymbol{\Sigma}_{jk}, \nu\sqrt{D}), \\ &X_{jk}^- - \widetilde{X}_{jk}^- \sim \mathrm{EC}_1(0, \boldsymbol{\Sigma}_{jj} + \boldsymbol{\Sigma}_{kk} - 2\boldsymbol{\Sigma}_{jk}, \nu\sqrt{D}), \\ &Y_{j'k'}^+ - \widetilde{Y}_{j'k'}^+ \sim \mathrm{EC}_1(0, \boldsymbol{\Sigma}_{j'j'} + \boldsymbol{\Sigma}_{k'k'} + 2(\boldsymbol{\Sigma}_1)_{j'k'}, \nu\sqrt{D}), \\ &Y_{j'k'}^- - \widetilde{Y}_{j'k'}^- \sim \mathrm{EC}_1(0, \boldsymbol{\Sigma}_{j'j'} + \boldsymbol{\Sigma}_{k'k'} - 2(\boldsymbol{\Sigma}_1)_{j'k'}, \nu\sqrt{D}).\end{aligned}$$

Using the same technique as in (A.10), we can obtain

$$\begin{aligned}&Q(|X_{jk}^+ - \widetilde{X}_{jk}^+|; 1/4)^2 = (\boldsymbol{\Sigma}_{jj} + \boldsymbol{\Sigma}_{kk} + 2\boldsymbol{\Sigma}_{jk})Q(\nu^2 D; 1/4), \\ &Q(|X_{jk}^- - \widetilde{X}_{jk}^-|; 1/4)^2 = (\boldsymbol{\Sigma}_{jj} + \boldsymbol{\Sigma}_{kk} - 2\boldsymbol{\Sigma}_{jk})Q(\nu^2 D; 1/4), \\ &Q(|Y_{j'k'}^+ - \widetilde{Y}_{j'k'}^+|; 1/4)^2 = \\ &\quad (\boldsymbol{\Sigma}_{j'j'} + \boldsymbol{\Sigma}_{k'k'} + 2(\boldsymbol{\Sigma}_1)_{j'k'})Q(\nu^2 D; 1/4), \\ &Q(|Y_{j'k'}^- - \widetilde{Y}_{j'k'}^-|; 1/4)^2 = \\ &\quad (\boldsymbol{\Sigma}_{j'j'} + \boldsymbol{\Sigma}_{k'k'} - 2(\boldsymbol{\Sigma}_1)_{j'k'})Q(\nu^2 D; 1/4).\end{aligned}$$

Thus, by the definitions of $\mathbf{R}_{jk}^{\mathrm{Q}}$ and $(\mathbf{R}_1^{\mathrm{Q}})_{jk}$, we have

$$\mathbf{R}_{jk}^{\mathrm{Q}} = \boldsymbol{\Sigma}_{jk}Q(\nu^2 D; 1/4) \tag{A.11}$$

for $j \neq k \in \{1, \ldots, d\}$ and

$$(\mathbf{R}_1^{\mathrm{Q}})_{j'k'} = (\boldsymbol{\Sigma}_1)_{j'k'}Q(\nu^2 D; 1/4) \tag{A.12}$$

for $j', k' \in \{1, \ldots, d\}$. Combining (A.10), (A.11), and (A.12) leads to (3.11) and with $m^{\mathrm{Q}} = Q(\nu^2 D; 1/4)$. □

A.2.3. Proof of Theorem 5.2

Proof. We first prove (5.1). For brevity, we denote

$$\begin{aligned}&\widehat{\sigma}_j^{\mathrm{Q}} := \widehat{\sigma}^{\mathrm{Q}}(\{X_{tj}\}_{t=1}^T),\ \sigma_j^{\mathrm{Q}} := \sigma^{\mathrm{Q}}(X_j), \\ &\widehat{\sigma}_{jk+}^{\mathrm{Q}} := \widehat{\sigma}^{\mathrm{Q}}(\{X_{tj} + X_{tk}\}_{t=1}^T), \sigma_{jk+}^{\mathrm{Q}} := \sigma^{\mathrm{Q}}(X_{1j} + X_{1k}),\end{aligned}$$

$\widehat{\sigma}^{\mathrm{Q}}_{jk-} := \widehat{\sigma}^{\mathrm{Q}}(\{X_{tj}-X_{tk}\}_{t=1}^{T}), \sigma^{\mathrm{Q}}_{jk-} := \sigma^{\mathrm{Q}}(X_{1j}-X_{1k}),$

for $j \neq k \in \{1,\ldots,d\}$. By definition, for any $u>0$, we have

$$\begin{aligned}
&\mathbb{P}(|\widehat{\mathbf{R}}^{\mathrm{Q}}_{jj} - \mathbf{R}^{\mathrm{Q}}_{jj}| \geqslant u) = \mathbb{P}(|\widehat{\sigma}^{\mathrm{Q}^2}_{j} - \sigma^{\mathrm{Q}^2}_{j}| \geqslant u)\\
\leqslant&\mathbb{P}(\{\widehat{\sigma}^{\mathrm{Q}}_{j} - \sigma^{\mathrm{Q}}_{j}\}^2 + 2\sigma^{\mathrm{Q}}_{j}|\widehat{\sigma}^{\mathrm{Q}}_{j} - \sigma^{\mathrm{Q}}_{j}| \geqslant u)\\
\leqslant&\mathbb{P}\Big(|\widehat{\sigma}^{\mathrm{Q}}_{j} - \sigma^{\mathrm{Q}}_{j}| \geqslant \sqrt{\frac{u}{2}}\Big) + \mathbb{P}\Big(|\widehat{\sigma}^{\mathrm{Q}}_{j} - \sigma^{\mathrm{Q}}_{j}| \geqslant \frac{u}{4\sigma^{\mathrm{Q}}_{j}}\Big). \quad \text{(A.13)}
\end{aligned}$$

The quantiles in the definitions of $\mathbf{R}^{\mathrm{Q}}$ and $\widehat{\mathbf{R}}^{\mathrm{Q}}$ are unique due to Condition 1 and absolute continuity of $\boldsymbol{X}_1$. Hence, applying Lemma A.5 and noting that $\sigma^{\mathrm{Q}}_{j} \leqslant \sigma^{\mathrm{Q}}_{\max}$, we have

$$\begin{aligned}
&\mathbb{P}(|\widehat{\mathbf{R}}^{\mathrm{Q}}_{jj} - \mathbf{R}^{\mathrm{Q}}_{jj}| \geqslant u) \leqslant\\
&2\exp\Big\{-\frac{T}{2(1+2\Theta(T))}\Big(\eta\sqrt{\frac{u}{2}} - \frac{4\Theta(T)}{T}\Big)^2\Big\} +\\
&2\exp\Big\{-\frac{T}{2(1+2\Theta(T))}\Big(\frac{\eta u}{4\sigma^{\mathrm{Q}}_{\max}} - \frac{4\Theta(T)}{T}\Big)^2\Big\}, \quad \text{(A.14)}
\end{aligned}$$

when $4\Theta(T)/(\eta T) \leqslant \sqrt{u/2}, u/(4\sigma^{\mathrm{Q}}_{\max}) \leqslant \kappa$. Now, for the off-diagonal entries, we have

$$\begin{aligned}
&\mathbb{P}(|\widehat{\mathbf{R}}^{\mathrm{Q}}_{jk} - \mathbf{R}^{\mathrm{Q}}_{jk}| \geqslant u)\\
\leqslant&\mathbb{P}\Big(|\widehat{\sigma}^{\mathrm{Q}^2}_{jk+} - \sigma^{\mathrm{Q}^2}_{jk+}| + |\widehat{\sigma}^{\mathrm{Q}^2}_{jk-} - \sigma^{\mathrm{Q}^2}_{jk-}| \geqslant 4u\Big)\\
\leqslant&\mathbb{P}\Big(|\widehat{\sigma}^{\mathrm{Q}^2}_{jk+} - \sigma^{\mathrm{Q}^2}_{jk+}| \geqslant 2u\Big) + \mathbb{P}\Big(|\widehat{\sigma}^{\mathrm{Q}^2}_{jk-} - \sigma^{\mathrm{Q}^2}_{jk-}| \geqslant 2u\Big).
\end{aligned}$$

Using the same technique as in (A.13), we further have

$$\begin{aligned}
&\mathbb{P}(|\widehat{\mathbf{R}}^{\mathrm{Q}}_{jk} - \mathbf{R}^{\mathrm{Q}}_{jk}| \geqslant u)\\
\leqslant&\mathbb{P}\Big(|\widehat{\sigma}^{\mathrm{Q}}_{jk+} - \sigma^{\mathrm{Q}}_{jk+}| \geqslant \sqrt{u}\Big) + \mathbb{P}\Big(|\widehat{\sigma}^{\mathrm{Q}}_{jk+} - \sigma^{\mathrm{Q}}_{jk+}| \geqslant \frac{u}{2\sigma^{\mathrm{Q}}_{jk+}}\Big) +\\
&\mathbb{P}\Big(|\widehat{\sigma}^{\mathrm{Q}}_{jk-} - \sigma^{\mathrm{Q}}_{jk-}| \geqslant \sqrt{u}\Big) + \mathbb{P}\Big(|\widehat{\sigma}^{\mathrm{Q}}_{jk-} - \sigma^{\mathrm{Q}}_{jk-}| \geqslant \frac{u}{2\sigma^{\mathrm{Q}}_{jk-}}\Big).
\end{aligned}$$

Applying Lemma A.5 and noting that $\sigma^{\mathrm{Q}}_{jk+} \leqslant \sigma^{\mathrm{Q}}_{\max}$ and $\sigma^{\mathrm{Q}}_{jk-} \leqslant \sigma^{\mathrm{Q}}_{\max}$, we have

$$\begin{aligned}
&\mathbb{P}(|\widehat{\mathbf{R}}^{\mathrm{Q}}_{jk} - \mathbf{R}^{\mathrm{Q}}_{jk}| \geqslant u) \leqslant\\
&4\exp\Big\{-\frac{T}{2(1+2\Theta(T))}\Big(\eta\sqrt{u} - \frac{4\Theta(T)}{T}\Big)^2\Big\} +\\
&4\exp\Big\{-\frac{T}{2(1+2\Theta(T))}\Big(\frac{\eta u}{2\sigma^{\mathrm{Q}}_{\max}} - \frac{4\Theta(T)}{T}\Big)^2\Big\}, \quad \text{(A.15)}
\end{aligned}$$

when $4\Theta(T)/(\eta T) \leqslant \sqrt{u}, u/(2\sigma^{\mathrm{Q}}_{\max}) \leqslant \kappa$. Combining (A.14) and (A.15), we obtain

$$\begin{aligned}
&\mathbb{P}(\|\widehat{\mathbf{R}}^{\mathrm{Q}} - \mathbf{R}^{\mathrm{Q}}\|_{\max} \geqslant u) \leqslant \sum_{j,k=1}^{d} \mathbb{P}(|\widehat{\mathbf{R}}^{\mathrm{Q}}_{jk} - \mathbf{R}^{\mathrm{Q}}_{jk}| \geqslant u)\\
\leqslant& 4d^2\Big[\exp\Big\{-\frac{T}{2(1+2\Theta(T))}\Big(\eta\sqrt{\frac{u}{2}} - \frac{4\Theta(T)}{T}\Big)^2\Big\} +\\
&\exp\Big\{-\frac{T}{2(1+2\Theta(T))}\Big(\frac{\eta u}{4\sigma^{\mathrm{Q}}_{\max}} - \frac{4\Theta(T)}{T}\Big)^2\Big\}\Big]\\
\leqslant& 8\max\Big\{\underbrace{d^2\exp\Big[-\frac{T}{2(1+2\Theta(T))}\Big(\eta\sqrt{\frac{u}{2}} - \frac{4\Theta(T)}{T}\Big)^2\Big]}_{A_1(u)},\\
&\underbrace{d^2\exp\Big[-\frac{T}{2(1+2\Theta(T))}\Big(\frac{\eta u}{4\sigma^{\mathrm{Q}}_{\max}} - \frac{4\Theta(T)}{T}\Big)^2\Big]}_{A_2(u)}\Big\},
\end{aligned}$$

when we have

$$\frac{4\Theta(T)}{\eta T} \leqslant \sqrt{\frac{u}{2}}, \sqrt{u}, \frac{u}{4\sigma^{\mathrm{Q}}_{\max}}, \frac{u}{2\sigma^{\mathrm{Q}}_{\max}} \leqslant \kappa. \qquad \text{(A.16)}$$

Setting $A_1(u_1) = 1/d^2$, we obtain

$$u_1 = \frac{2}{\eta^2}\Big[\sqrt{\frac{8(1+2\Theta(T))\log d}{T}} + \frac{4\Theta(T)}{T}\Big]^2.$$

Setting $A_2(u_2) = 1/d^2$, we obtain

$$u_2 = \frac{4\sigma^{\mathrm{Q}}_{\max}}{\eta}\Big[\sqrt{\frac{8(1+2\Theta(T))\log d}{T}} + \frac{4\Theta(T)}{T}\Big].$$

Now set $u = r(T) = \max(u_1, u_2)$. (A.16) is satisfied when T is large enough. If $u_1 \geqslant u_2$, since $A_2(u)$ is a non-increasing function of u, we have $A_2(u_1) \leqslant A_2(u_2) = 1/d^2$. Thus, we have

$$\mathbb{P}(\|\widehat{\mathbf{R}}^{\mathrm{Q}} - \mathbf{R}^{\mathrm{Q}}\|_{\max} \geqslant r(T)) \leqslant 8\max\{A_1(u), A_2(u)\} \leqslant 8/d^2.$$

On the other hand, if $u_1 < u_2$, we have $r(T) = u_2$. Since $A_1(u)$ is a non-increasing function of u, we have $A_1(u_2) \leqslant A_1(u_1) = 1/d^2$. Thus, we still have

$$\mathbb{P}(\|\widehat{\mathbf{R}}^{\mathrm{Q}} - \mathbf{R}^{\mathrm{Q}}\|_{\max} \geqslant r(T)) \leqslant 8\max\{A_1(u), A_2(u)\} \leqslant 8/d^2.$$

This proves (5.1).

To prove (5.2), we employ the same technique as above. Specifically, denote

$$\begin{aligned}
&\widehat{\tau}^{\mathrm{Q}}_{j'k'+} := \widehat{\sigma}^{\mathrm{Q}}(\{X_{tj'} + X_{t+1,k'}\}_{t=1}^{T-1}), \tau^{\mathrm{Q}}_{j'k'+} := \sigma^{\mathrm{Q}}(X_{1j'} + X_{2k'}),\\
&\widehat{\tau}^{\mathrm{Q}}_{j'k'-} := \widehat{\tau}^{\mathrm{Q}}(\{X_{tj'} - X_{t+1,k'}\}_{t=1}^{T-1}), \tau^{\mathrm{Q}}_{j'k'-} := \sigma^{\mathrm{Q}}(X_{1j'} - X_{2k'}),
\end{aligned}$$

for $j', k' \in \{1,\ldots,d\}$. Using the same technique in deriving (A.15), we can obtain

$$\begin{aligned}
&\mathbb{P}(|(\widehat{\mathbf{R}}^{\mathrm{Q}}_1)_{j'k'} - (\mathbf{R}^{\mathrm{Q}}_1)_{j'k'}| \geqslant u) \leqslant\\
&4\exp\Big\{-\frac{T-1}{2(1+2\Theta(T-1))}\Big(\eta\sqrt{u} - \frac{4\Theta(T-1)}{T-1}\Big)^2\Big\} +\\
&4\exp\Big\{-\frac{T-1}{2(1+2\Theta(T-1))}\Big(\frac{\eta u}{2\sigma^{\mathrm{Q}}_{\max}} - \frac{4\Theta(T-1)}{T-1}\Big)^2\Big\},
\end{aligned} \qquad \text{(A.17)}$$

when $4\Theta(T-1)/(\eta(T-1)) \leqslant \sqrt{u}, u/(2\sigma_{\max}^{Q}) \leqslant \kappa$. Hence, we obtain

$$\mathbb{P}(\|\widehat{\mathbf{R}}_1^{Q} - \mathbf{R}_1^{Q}\|_{\max} \geqslant u) \leqslant \sum_{j',k'=1}^{d} \mathbb{P}(|\widehat{\mathbf{R}}_{j'k'}^{Q} - \mathbf{R}_{j'k'}^{Q}| \geqslant u)$$
$$\leqslant 8\max\Big\{\underbrace{d^2 \exp\Big[-\frac{T}{4(1+2\Theta(T))}\Big(\eta\sqrt{u} - \frac{8\Theta(T)}{T}\Big)^2\Big]}_{B_1(u)},$$
$$\underbrace{d^2 \exp\Big[-\frac{T}{4(1+2\Theta(T))}\Big(\frac{\eta u}{2\tau_{\max}^{Q}} - \frac{8\Theta(T)}{T}\Big)^2\Big]}_{B_2(u)}\Big\},$$

when we have

$$\frac{8\Theta(T)}{\eta T} \leqslant \sqrt{u}, \frac{u}{2\tau_{\max}^{Q}} \leqslant \kappa. \tag{A.18}$$

Here we used the fact that $\Theta(T-1) \leqslant \Theta(T)$ and $T-1 > T/2$ when $T > 3$. Again, (A.18) can be fulfilled when T is large enough. Setting $B_1(u_3) = 1/d^2$, we obtain

$$u_3 = \frac{1}{\eta^2}\Big[\sqrt{\frac{16(1+2\Theta(T))\log d}{T}} + \frac{8\Theta(T)}{T}\Big]^2.$$

Setting $B_2(u_4) = 1/d^2$, we obtain

$$u_4 = \frac{2\tau_{\max}^{Q}}{\eta}\Big[\sqrt{\frac{16(1+2\Theta(T))\log d}{T}} + \frac{8\Theta(T)}{T}\Big].$$

Let $r_1(T) = \max(u_3, u_4)$. Using the same argument as in deriving (5.1), we may conclude that

$$\mathbb{P}(\|\widehat{\mathbf{R}}_1^{Q} - \mathbf{R}_1^{Q}\|_{\max} \geqslant r_1(T)) \leqslant 8/d^2.$$

This completes the proof. □

A.2.4. Proof of Theorem 5.3

Proof. We first show that with large probability, $\mathbf{A}$ is feasible to the optimization problem (4.1). By Theorem 3.7, we have $\mathbf{A}^{\mathsf{T}} = (\mathbf{R}^{Q})^{-1}\mathbf{R}_1^{Q}$. Thus, we have

$$\begin{aligned}
\|\widehat{\mathbf{R}}^{Q}\mathbf{A}^{\mathsf{T}} - \widehat{\mathbf{R}}_1^{Q}\|_{\max} &= \|\widehat{\mathbf{R}}^{Q}(\mathbf{R}^{Q})^{-1}\mathbf{R}_1^{Q} - \widehat{\mathbf{R}}_1^{Q}\|_{\max}\\
=&\|\widehat{\mathbf{R}}^{Q}(\mathbf{R}^{Q})^{-1}\mathbf{R}_1^{Q} - \mathbf{R}_1^{Q} + \mathbf{R}_1^{Q} - \widehat{\mathbf{R}}_1^{Q}\|_{\max}\\
\leqslant&\|[\widehat{\mathbf{R}}^{Q}(\mathbf{R}^{Q})^{-1} - \mathbf{I}]\mathbf{R}_1^{Q}\|_{\max} + \|\mathbf{R}_1^{Q} - \widehat{\mathbf{R}}_1^{Q}\|_{\max}\\
=&\|(\widehat{\mathbf{R}}^{Q} - \mathbf{R}^{Q})\mathbf{A}\|_{\max} + \|\mathbf{R}_1^{Q} - \widehat{\mathbf{R}}_1^{Q}\|_{\max}\\
\leqslant&\|\widehat{\mathbf{R}}^{Q} - \mathbf{R}^{Q}\|_{\max}\|\mathbf{A}\|_1 + \|\mathbf{R}_1^{Q} - \widehat{\mathbf{R}}_1^{Q}\|_{\max}\\
\leqslant&\|\widehat{\mathbf{R}}^{Q} - \mathbf{R}^{Q}\|_{\max}M_T + \|\mathbf{R}_1^{Q} - \widehat{\mathbf{R}}_1^{Q}\|_{\max}.
\end{aligned}$$

The last inequality is due to $\mathbf{A} \in \mathcal{M}(\alpha, s, M_T)$. By Lemma 5.2, we have, with probability no smaller than $1 - 8/d^2$,

$$\begin{aligned}
\|\widehat{\mathbf{R}}^{Q}\mathbf{A}^{\mathsf{T}} - \widehat{\mathbf{R}}_1^{Q}\|_{\max} \leqslant& r(T)M_T + r_1(T)\\
\leqslant&(1+M_T)r_{\max}(T) = \lambda.
\end{aligned}$$

Next, we prove (5.5). Using $\mathbf{A}^{\mathsf{T}} = (\mathbf{R}^{Q})^{-1}\mathbf{R}_1^{Q}$, we have

$$\begin{aligned}
\|\widehat{\mathbf{A}} - \mathbf{A}\|_{\max} &= \|\widehat{\mathbf{A}} - (\mathbf{R}^{Q})^{-1}\mathbf{R}_1^{Q}\|_{\max}\\
=&\|(\mathbf{R}^{Q})^{-1}(\mathbf{R}^{Q}\widehat{\mathbf{A}} - \mathbf{R}_1^{Q})\|_{\max}\\
=&\|(\mathbf{R}^{Q})^{-1}(\mathbf{R}^{Q}\widehat{\mathbf{A}} - \widehat{\mathbf{R}}^{Q}\widehat{\mathbf{A}} + \widehat{\mathbf{R}}_1^{Q} - \mathbf{R}_1^{Q} + \widehat{\mathbf{R}}^{Q}\widehat{\mathbf{A}} - \widehat{\mathbf{R}}_1^{Q})\|_{\max}\\
\leqslant&\|(\mathbf{R}^{Q})^{-1}\|_1(\|\mathbf{R}^{Q} - \widehat{\mathbf{R}}^{Q}\|_{\max}\|\widehat{\mathbf{A}}\|_1 +\\
&\|\widehat{\mathbf{R}}_1^{Q} - \mathbf{R}_1^{Q}\|_{\max} + \|\widehat{\mathbf{R}}^{Q}\widehat{\mathbf{A}} - \widehat{\mathbf{R}}_1^{Q}\|_{\max}).
\end{aligned}$$

Since $\mathbf{A}$ is feasible to optimization problem (4.1) with probability no smaller than $1 - 1/d^2$, and $\widehat{\mathbf{A}}$ is the solution to (4.1), we have $\|\widehat{\mathbf{A}}\|_1 \leqslant \|\mathbf{A}\|_1$ with probability no smaller than $1 - 1/d^2$. Using Lemma 5.2 and $\|\widehat{\mathbf{R}}^{Q}\widehat{\mathbf{A}} - \widehat{\mathbf{R}}_1^{Q}\|_{\max} \leqslant \lambda$, we further have

$$\begin{aligned}
&\|\widehat{\mathbf{A}} - \mathbf{A}\|_{\max}\\
\leqslant&\|(\mathbf{R}_1^{Q})^{-1}\|_1[r(T)\|\mathbf{A}\|_1 + r_1(T) + (1+M_T)r_{\max}(T)]\\
\leqslant&2\|(\mathbf{R}_1^{Q})^{-1}\|_1(1+M_T)r_{\max}(T),
\end{aligned}$$

with probability no smaller than $1-1/d^2$. This proves (5.5).

To prove (5.6), let λ_1 be a parameter to be defined later, and denote

$$s_1 := \max_{1\leqslant j\leqslant d}\sum_{k=1}^{d}\min(|\mathbf{A}_{jk}/\lambda_1|, 1) \text{ and } S_j := \{k : |\mathbf{A}_{jk}| \leqslant \lambda_1\}.$$

It follows that $|S_j| \leqslant s_1$, where $|S_j|$ denote the cardinality of S_j. We have

$$\begin{aligned}
\|\widehat{\mathbf{A}}_{j,*} - \mathbf{A}_{j,*}\|_1 \leqslant& \|\widehat{\mathbf{A}}_{j,S_j^c}\|_1 + \|\mathbf{A}_{j,S_j^c}\|_1\\
&+ \|\widehat{\mathbf{A}}_{j,S_j} - \mathbf{A}_{j,S_j}\|_1.
\end{aligned} \tag{A.19}$$

By the equivalence of (4.1) and (4.2), we have $\|\widehat{\mathbf{A}}_{j,*}\|_1 \leqslant \|\mathbf{A}_{j,*}\|_1$ for any $j \in \{1, \ldots, d\}$, with probability no smaller than $1 - 1/d^2$. Thus, we have

$$\begin{aligned}
\|\widehat{\mathbf{A}}_{j,S_j^c}\|_1 =& \|\widehat{\mathbf{A}}_{j,*}\|_1 - \|\widehat{\mathbf{A}}_{j,S_j}\|_1 \leqslant \|\mathbf{A}_{j,*}\|_1 - \|\widehat{\mathbf{A}}_{j,S_j}\|_1\\
=&\|\mathbf{A}_{j,S_j}\|_1 + \|\mathbf{A}_{j,S_j^c}\|_1 - \|\widehat{\mathbf{A}}_{j,S_j}\|_1\\
\leqslant&\|\mathbf{A}_{j,S_j} - \widehat{\mathbf{A}}_{j,S_j}\|_1 + \|\mathbf{A}_{j,S_j^c}\|_1
\end{aligned}$$

Plugging the above equation into (A.19), we obtain

$$\|\widehat{\mathbf{A}}_{j,*} - \mathbf{A}_{j,*}\|_1 \leqslant 2\|\mathbf{A}_{j,S_j^c}\|_1 + 2\|\widehat{\mathbf{A}}_{j,S_j} - \mathbf{A}_{j,S_j}\|_1. \tag{A.20}$$

When $k \notin S_j$, we have $\mathbf{A}_{jk} < \lambda_1$. Thus, we have

$$\begin{aligned}
\|\mathbf{A}_{j,S_j^c}\|_1 =& \lambda_1 \sum_{k\in S_j^c} |\mathbf{A}_{jk}|/\lambda_1\\
=&\lambda_1 \sum_{k\in S_j^c} \min(|\mathbf{A}_{jk}|/\lambda_1, 1) \leqslant \lambda_1 s_1.
\end{aligned} \tag{A.21}$$

Regarding the second term on the right hand side of (A.20), we have

$$\begin{aligned}\|\widehat{\mathbf{A}}_{j,S_j} - \mathbf{A}_{j,S_j}\|_1 \leqslant& \|\widehat{\mathbf{A}} - \mathbf{A}\|_{\max}|S_j| \\ \leqslant& 2\|(\mathbf{R}_1^{\mathrm{Q}})^{-1}\|_1(1 + M_T)r_{\max}(T)s_1. \end{aligned} \tag{A.22}$$

Combining (A.19), (A.21), and (A.22), we have

$$\|\widehat{\mathbf{A}}_{j,*}-\mathbf{A}_{j,*}\|_1\leqslant[2\lambda_1+4\|(\mathbf{R}_1^{\mathrm{Q}})^{-1}\|_1(1+M_T)r_{\max}(T)]s_1.$$

Let $\lambda_1 = 2\|(\mathbf{R}_1^{\mathrm{Q}})^{-1}\|_1(1+M_T)r_{\max}(T)$, we have $\|\widehat{\mathbf{A}}_{j,*} - \mathbf{A}_{j,*}\|_1 \leqslant 4\lambda_1 s_1$. By the definition of s_1, since $\alpha \in [0, 1)$, we have

$$\begin{aligned} s_1 \leqslant& \max_{1\leqslant j\leqslant d}\sum_{k=1}^{d}\min(|\mathbf{A}_{jk}|^{\alpha}/\lambda_1^{\alpha}, 1) \\ \leqslant& \max_{1\leqslant j\leqslant d}\sum_{k=1}^{d}|\mathbf{A}_{jk}|^{\alpha}/\lambda_1^{\alpha} \leqslant s/\lambda_1^{\alpha}. \end{aligned}$$

Hence, we have

$$\begin{aligned}\|\widehat{\mathbf{A}}_{j,*} - \mathbf{A}_{j,*}\|_1 \leqslant& 4\lambda_1^{1-\alpha}s \\ =& 4s\Big[2\|(\mathbf{R}^{\mathrm{Q}})^{-1}\|_1(1 + M_T)r_{\max}(T)\Big]^{1-\alpha}. \end{aligned}$$

Since the above equation holds for any $j \in \{1, \ldots, d\}$, we have (5.6). □

A.2.5. Proof of Theorem 6.2

Proof. In order to prove Claim 1, we only need to prove that $\{X_{tk}\}_{t\in\mathbb{Z}}$ doesn't Granger cause $\{X_{tj}\}_{t\in\mathbb{Z}}$ implies $\mathbf{A}_{jk} = 0$. Suppose for some $t \in \mathbb{Z}$, we have

$$\mathbb{P}(X_{t+1,j} \in A \mid \{\boldsymbol{X}_s\}_{s\leqslant t}) = \mathbb{P}(X_{t+1,j} \in A \mid \{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t}),$$

for any measurable set A. The above equation implies that conditioning on $\{X_{s,\backslash k}\}_{s\leqslant t}$, $X_{t+1,j}$ is independent of $\{X_{sk}\}_{s\leqslant t}$. Hence, we have

$$\mathrm{Cov}(X_{t+1,j}, X_{tk} \mid \{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t}) = 0.$$

Plugging $X_{t+1,j} = \sum_{l=1}^{d}\mathbf{A}_{jl}X_{tl} + E_{t+1,j}$ into the above equation, we have

$$\begin{aligned} 0 =& \mathrm{Cov}(\mathbf{A}_{jk}X_{tk}, X_{tk} \mid \{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t})+ \\ & \mathrm{Cov}(\sum_{l\neq k}\mathbf{A}_{jl}X_{tl}, X_{tk} \mid \{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t})+ \\ & \mathrm{Cov}(E_{t+1,j}, X_{tk} \mid \{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t}). \end{aligned}$$

The second term on the right hand side is 0, since given $\{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t}$, $\sum_{l\neq k}\mathbf{A}_{jl}X_{tl}$ is constant. Since $(\boldsymbol{\Omega}_{XE})_{jk} = 0$ for $j \leqslant k$, we have $\mathrm{Cov}(E_{t+1,j}, \boldsymbol{X}_{sk}) = 0$ for any $s \leqslant t$. Using Theorem 2.18 in Fang et al. (1990), we have the third term is also 0. Thus, we have $\mathbf{A}_{jk}\,\mathrm{Var}(X_{tk} \mid \boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t}) = 0$, and hence $\mathbf{A}_{jk} = 0$. This proves Claim 1.

Given Claim 1, to prove Claim 2, it remains to prove that $\mathbf{A}_{jk} = 0$ implies that $\{X_{tk}\}_{t\in\mathbb{Z}}$ doesn't Granger cause $\{X_{tj}\}_{t\in\mathbb{Z}}$. Since $\mathbf{A}_{jk} = 0$, we have

$$\begin{aligned} & p(X_{t+1,j}, \{X_{sk}\}_{s\leqslant t} \mid \{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t}) \\ =& p(\sum_{l\neq k}\mathbf{A}_{jl}X_{tl} + E_{t+1,j}, \{X_{sk}\}_{s\leqslant t} \mid \{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t}) \\ =& p(\sum_{l\neq k}\mathbf{A}_{jl}X_{tl} + E_{t+1,j} \mid \{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t})\cdot \\ & \quad p(\{X_{sk}\}_{s\leqslant t} \mid \{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t}). \end{aligned}$$

Here p is the conditional probability density function. The last equation is because $\boldsymbol{E}_{t+1}$ is independent of $\{\boldsymbol{X}_s\}_{s\leqslant t}$, and the fact that $\sum_{l\neq k}\mathbf{A}_{jl}X_{tl}$ is constant given $\{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t}$. Hence, we have

$$\begin{aligned} & p(X_{t+1,j}, \{X_{sk}\}_{s\leqslant t} \mid \{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t}) \\ =& p(X_{t+1,j} \mid \{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t})p(\{X_{sk}\}_{s\leqslant t} \mid \{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t}), \end{aligned}$$

and thus

$$p(X_{t+1,j} \mid \{\boldsymbol{X}_s\}_{s\leqslant t}) = p(X_{t+1,j} \mid \{\boldsymbol{X}_{s,\backslash k}\}_{s\leqslant t}).$$

This completes the proof. □

A.2.6. Proof of Theorem 6.4

Proof. Theorem 6.4 is a consequence of Theorem 5.3. Indetail, if $\mathbf{A}_{jk} > 0$, by (6.1), we have $\mathbf{A}_{jk} \geqslant 2\gamma$. By Theorem 5.3, with probability no smaller than $1-8/d^2$, we have $|\widehat{\mathbf{A}}_{jk}-\mathbf{A}_{jk}| \leqslant \gamma$. Thus, we have $\widehat{\mathbf{A}}_{jk} \geqslant \gamma$ with probability no smaller than $1 - 8/d^2$. By the definition of $\widetilde{\mathbf{A}}$, we have $\widetilde{\mathbf{A}}_{jk} = \widehat{\mathbf{A}}_{jk} \geqslant \gamma > 0$.

If $\mathbf{A}_{jk} < 0$, by (6.1), we have $\mathbf{A}_{jk} \leqslant -2\gamma$. Using Theorem 5.3, we have $\widehat{\mathbf{A}}_{jk} \leqslant -\gamma$ with probability no smaller than $1 - 8/d^2$. By the definition of $\widetilde{\mathbf{A}}$, we have $\widetilde{\mathbf{A}}_{jk} = \widehat{\mathbf{A}}_{jk} \leqslant -\gamma < 0$.

If $\mathbf{A}_{jk} = 0$, using Theorem 5.3, we have $\widehat{\mathbf{A}}_{jk} < \gamma$ with probability no smaller than $1 - 8/d^2$, since $\mathbb{P}(\widehat{\mathbf{A}}_{jk} = \gamma) = 0$. By the definition of $\widetilde{\mathbf{A}}$, we have $\widetilde{\mathbf{A}}_{jk} = 0$. □

References

Fang, Kai-Tai, Kotz, Samuel, and Ng, Kai Wang. *Symmetric Multivariate and Related Distributions*. Chapman and Hall, 1990.

Kontorovich, Leonid Aryeh, Ramanan, Kavita, et al. Concentration inequalities for dependent random variables via the martingale method. *The Annals of Probability*, 36(6):2126–2158, 2008.

Mohri, Mehryar and Rostamizadeh, Afshin. Stability bounds for stationary ϕ-mixing and β-mixing processes. *The Journal of Machine Learning Research*, 11:789–814, 2010.