Supplementary material for "Supervised Factor Modeling for High-Dimensional Linear Time Series"

Feiqing Huang, Kexin Lu, Yao Zheng and Guodong Li

University of Hong Kong and University of Connecticut

Abstract

This supplementary material provides detailed technical proofs for the theoretical results in the paper in the first four sections. Section 5 includes an additional simulation study, along with complementary details about our simulation setup. Section 6 offers further information on the macroeconomic dataset, while Section 7 presents an empirical study on realized volatility.

S1 Notations and preliminaries

S1.1 Brief introduction to tensor notations and decomposition

This subsection gives a brief introduction to tensor notations and Tucker decomposition, and a detailed review on tensor notations and operations can be referred to in Kolda and Bader (2009). Tensors, also known as multidimensional arrays, are higher-order extensions of matrices, and a multidimensional array $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is called a K-th-order tensor, where the order of a tensor is known as the dimension, way or mode. This paper concentrates on third-order tensors.

For a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, its element is denoted by $\boldsymbol{\mathcal{A}}_{ijk}$ for $1 \leqslant i \leqslant d_1, 1 \leqslant j \leqslant d_2$ and $1 \leqslant k \leqslant d_3$, and the Frobenius norm is defined as $\|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}} = \sqrt{\sum_{i,j,k} \boldsymbol{\mathcal{A}}_{ijk}^2}$. We define its mode-1 multiplication with a matrix $\boldsymbol{B} \in \mathbb{R}^{d_1 \times p_1}$ as $\boldsymbol{\mathcal{A}} \times_1 \boldsymbol{B} \in \mathbb{R}^{p_1 \times d_2 \times d_3}$ with elements of $(\boldsymbol{\mathcal{A}} \times_1 \boldsymbol{B})_{\ell jk} = \sum_{i=1}^{d_1} \boldsymbol{\mathcal{A}}_{ijk} \boldsymbol{B}_{\ell i}$. The mode-2 and -3 multiplications, $\times_2$ and $\times_3$, can be defined similarly.

Matricization or unfolding is an operation to reshape a tensor into matrices of different sizes, and it can help to link the concepts and properties of matrices to those of tensors. The mode-1 matricization of $\boldsymbol{\mathcal{A}}$ is defined as $\boldsymbol{\mathcal{A}}_{(1)} \in \mathbb{R}^{d_1 \times d_2 d_3}$, whose $\{i, (k-1)d_3 + j\}$-th entry is $\boldsymbol{\mathcal{A}}_{ijk}$ for all possible i's, j's and k's, i.e. $\boldsymbol{\mathcal{A}}_{(1)}$ contains all mode-1 fibers $\{(\boldsymbol{\mathcal{A}}_{[:,j,k]}) \in \mathbb{R}^{d_1} : 1 \leqslant j \leqslant d_2, 1 \leqslant k \leqslant d_3\}$. We can similarly define the mode-2 and -3 matricizations of $\boldsymbol{\mathcal{A}}$, denoted by $\boldsymbol{\mathcal{A}}_{(2)} \in \mathbb{R}^{d_2 \times d_1 d_3}$ and $\boldsymbol{\mathcal{A}}_{(3)} \in \mathbb{R}^{d_3 \times d_1 d_2}$, respectively. When the tensor $\boldsymbol{\mathcal{A}}$ has a form of $\boldsymbol{\mathcal{A}}_{(1)} = (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_{d_3})$ with $\boldsymbol{A}_j \in \mathbb{R}^{d_1 \times d_2}$ for all $1 \leqslant j \leqslant d_3$, it holds that $\boldsymbol{\mathcal{A}}_{(2)} = (\boldsymbol{A}_1', \ldots, \boldsymbol{A}_{d_3}')$ and $\boldsymbol{\mathcal{A}}_{(3)} = (\mathrm{vec}(\boldsymbol{A}_1), \ldots, \mathrm{vec}(\boldsymbol{A}_{d_3}))'$.

The multilinear ranks of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is defined as (r_1, r_2, r_3), where

$$r_1 = \mathrm{rank}(\boldsymbol{\mathcal{A}}_{(1)}), \;\; r_2 = \mathrm{rank}(\boldsymbol{\mathcal{A}}_{(2)}) \;\text{ and }\; r_3 = \mathrm{rank}(\boldsymbol{\mathcal{A}}_{(3)}).$$

Accordingly, there exists a Tucker decomposition (Tucker, 1966),

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \times_3 \boldsymbol{U}_3,$$

where $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ is the core tensor, $\boldsymbol{U}_j \in \mathbb{R}^{d_j \times r_j}$ with $1 \leqslant j \leqslant 3$ are factor matrices. Note that the Tucker decomposition is not unique, since $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \times_3 \boldsymbol{U}_3 = (\boldsymbol{\mathcal{G}} \times_1 \boldsymbol{O}_1 \times_2 \boldsymbol{O}_2 \times_3 \boldsymbol{O}_3) \times_1 (\boldsymbol{U}_1 \boldsymbol{O}_1^{-1}) \times_2 (\boldsymbol{U}_2 \boldsymbol{O}_2^{-1}) \times_3 (\boldsymbol{U}_3 \boldsymbol{O}_3^{-1})$ for any invertible matrices $\boldsymbol{O}_i \in \mathbb{R}^{r_i \times r_i}$ with $1 \leqslant i \leqslant 3$. We can consider the higher order singular value decomposition (HOSVD) of $\boldsymbol{\mathcal{A}}$, a special Tucker decomposition uniquely defined by choosing $\boldsymbol{U}_i$ as the tall matrix consisting of the top r_i left singular vectors of $\boldsymbol{\mathcal{A}}_{(i)}$ and then setting $\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{A}} \times_1 \boldsymbol{U}_1^\top \times_2 \cdots \times_d \boldsymbol{U}_d^\top$. Note that $\boldsymbol{U}_i$'s are orthonormal, i.e. $\boldsymbol{U}_i^\top \boldsymbol{U}_i = \boldsymbol{I}_{r_i}$ with $1 \leqslant i \leqslant d$.

The three ranks, r_1, r_2 and r_3, are not equal in general. In particular, when $r_3 = d_3$, the multilinear ranks of $\boldsymbol{\mathcal{A}}$ are denoted by (r_1, r_2) instead, omitting the rank of mode-3 matricization. The multilinear ranks are also known as Tucker ranks, as they are closely related to the Tucker decomposition. There are many other tensor decomposition methods, such as CP decomposition

(Kolda and Bader, 2009), and the ranks of a tensor can be defined in many different ways.

S1.2 More notations and preliminaries used in technical proofs

This subsection gives more notations and preliminaries. Let $\mathcal{S}^{d-1} = \{\mathcal{X} \in \mathbb{R}^d \mid \|\mathcal{X}\|_2 = 1\}$ be the unit sphere of $\mathbb{R}^d$ in Euclidean norm. For $n \geqslant 2$, let $\mathcal{S}^{d_1\times\cdots\times d_n-1} = \{\mathcal{X} \in \mathbb{R}^{d_1\times\cdots\times d_n} \mid \|\mathcal{X}\|_{\mathrm{F}} = 1\}$ be the unit sphere of $\mathbb{R}^{d_1\times\cdots\times d_n}$ in Frobenius norm. Moreover, for any positive integer $r \leqslant N$, denote the set of unit matrices with rank at most r by

$$\boldsymbol{\Theta}_{\mathrm{F}}(r) = \{\boldsymbol{M} \in \mathbb{R}^{N\times N} \mid \|\boldsymbol{M}\|_{\mathrm{F}} = 1, \mathrm{rank}(\boldsymbol{M}) \leqslant r\}.$$

Consider a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{N\times N\times T_0}$ such that $\boldsymbol{\mathcal{A}}_{(1)} = (\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_{T_0})$, i.e. $\boldsymbol{A}_j$ is the j-th frontal slice of $\boldsymbol{\mathcal{A}}$. For any index set $S \subseteq \{1, \ldots, T_0\}$ with cardinality $|S| = s$, denote by $\boldsymbol{\mathcal{A}}_S \in \mathbb{R}^{N\times N\times T_0}$ such that its j-th frontal slice is $\boldsymbol{A}_j$ if $j \in S$, and a zero matrix if $j \notin S$. Let $\|\boldsymbol{\mathcal{A}}\|_{\ddagger} = \sum_{j=1}^{T_0} \|\boldsymbol{A}_j\|_{\mathrm{F}}$, and it holds that

$$\|\boldsymbol{\mathcal{A}}\|_{\ddagger} = \|\boldsymbol{\mathcal{A}}_S\|_{\ddagger} + \|\boldsymbol{\mathcal{A}}_{S^c}\|_{\ddagger}, \tag{S1}$$

where $S^c = \{1, \ldots, T_0\}\backslash S$. Moreover, it can be verified that

$$\|\boldsymbol{\mathcal{A}}_S\|_{\ddagger} = \sum_{i\in S} \|\boldsymbol{A}_i\|_{\mathrm{F}} \leqslant \sqrt{|S| \sum_{i\in S} \|\boldsymbol{A}_i\|_{\mathrm{F}}^2} \leqslant \sqrt{s}\|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}. \tag{S2}$$

Recall that $\boldsymbol{Y} = (\mathbf{y}_T, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_{T_0+1}) \in \mathbb{R}^{N\times T_1}$ and $\boldsymbol{X} = (\mathbf{x}_T, \mathbf{x}_{T-1}, \ldots, \mathbf{x}_{T_0+1}) \in \mathbb{R}^{NT_0\times T_1}$, where $\mathbf{x}_t = (\mathbf{y}_{t-1}', \mathbf{y}_{t-2}', \ldots, \mathbf{y}_{t-T_0}')'$, for $T_0 + 1 \leqslant t \leqslant T$. We further denote

$$\boldsymbol{r}_t = \sum_{j=T_0+1}^{\infty} \boldsymbol{A}_j^* \mathbf{y}_{t-j}, \qquad \boldsymbol{R} = (\boldsymbol{r}_T, \boldsymbol{r}_{T-1}, \ldots, \boldsymbol{r}_{T_0+1}) \in \mathbb{R}^{N\times T_1}$$

and

$$\widetilde{\boldsymbol{\varepsilon}}_t = \boldsymbol{r}_t + \boldsymbol{\varepsilon}_t, \qquad \widetilde{\boldsymbol{\mathcal{E}}} = \boldsymbol{R} + \boldsymbol{\mathcal{E}} = (\widetilde{\boldsymbol{\varepsilon}}_T, \widetilde{\boldsymbol{\varepsilon}}_{T-1}, \ldots, \widetilde{\boldsymbol{\varepsilon}}_{T_0+1}) \in \mathbb{R}^{N\times T_1}$$

where $\boldsymbol{\mathcal{E}} = (\boldsymbol{\varepsilon}_T, \boldsymbol{\varepsilon}_{T-1}, \ldots, \boldsymbol{\varepsilon}_{T_0+1})$. In addition, we partition $\boldsymbol{X}$ into T_0 blocks such that $\boldsymbol{X} = (\boldsymbol{X}_1', \ldots, \boldsymbol{X}_{T_0}')'$, where the j-th block is the $N \times T_1$ matrix

$$\boldsymbol{X}_j = (\mathbf{y}_{T-j}, \mathbf{y}_{T-1-j}, \ldots, \mathbf{y}_{T_0+1-j}) \in \mathbb{R}^{N\times T_1}, \quad j = 1, \ldots, T_0. \tag{S3}$$

Let

$$\boldsymbol{\Sigma}_{T_0} = \mathbb{E}\left(\frac{\boldsymbol{X}\boldsymbol{X}'}{T_1}\right) = \mathbb{E}(\mathbf{x}_t\mathbf{x}_t') = \begin{bmatrix} \boldsymbol{\Gamma}(0) & \boldsymbol{\Gamma}'(1) & \cdots & \boldsymbol{\Gamma}'(T_0-1) \\ \boldsymbol{\Gamma}(1) & \boldsymbol{\Gamma}(0) & \cdots & \boldsymbol{\Gamma}'(T_0-2) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}(T_0-1) & \boldsymbol{\Gamma}(T_0-2) & \cdots & \boldsymbol{\Gamma}(0) \end{bmatrix}, \tag{S4}$$

whose (i,j)-th block is

$$\boldsymbol{\Gamma}(i-j) = \mathbb{E}\left(\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1}\right) = \mathbb{E}(\mathbf{y}_{t-i}\mathbf{y}_{t-j}') \in \mathbb{R}^{N\times N}, \quad 1 \leqslant i \leqslant j \leqslant T_0.$$

S2 Proof of Proposition 1

For a certain general linear process, $\mathbf{y}_t = \boldsymbol{\varepsilon}_t + \sum_{j=1}^{\infty}\boldsymbol{\Psi}_j\boldsymbol{\varepsilon}_{t-j}$, with fixed r_1, r_2 and N, it is implied by Assumption 1 that $\sum_{j=0}^{\infty}\|\boldsymbol{\Psi}_j\|_{\mathrm{F}} < \infty$ and, by directly following (2.2) in Lewis and Reinsel (1985) and its discussion, its VAR(∞) representation can be uniquely identified below,

$$\mathbf{y}_t = \sum_{j=1}^{\infty}\boldsymbol{A}_j\mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t, \quad \text{with} \quad \boldsymbol{A}_j = \boldsymbol{\Psi}_j - \sum_{k=1}^{j-1}\boldsymbol{\Psi}_{j-k}\boldsymbol{A}_k, \ \forall j \geqslant 1. \tag{S1}$$

Moreover, it can be easily verified that $\sum_{j=1}^{\infty}\|\boldsymbol{A}_j\|_{\mathrm{op}} < \infty$.

Let $\mathbb{M}_1 = \mathrm{colspace}\{\boldsymbol{\Psi}_j, j \geqslant 1\}$ and $\mathbb{M}_2 = \mathrm{rowspace}\{\boldsymbol{\Psi}_j, j \geqslant 1\}$ be the column and row spaces of coefficient matrices $\boldsymbol{\Psi}_j$'s, respectively. We first show by induction that for all $j \geqslant 1$,

$$\mathrm{colspace}(\boldsymbol{A}_j) \subseteq \mathbb{M}_1 \quad \text{and} \quad \mathrm{rowspace}(\boldsymbol{A}_j) \subseteq \mathbb{M}_2. \tag{S2}$$

Since $\boldsymbol{A}_1 = \boldsymbol{\Phi}_1$, (S2) holds trivially for $j = 1$. Suppose that (S2) holds for all $j \leqslant k$, then, by the fact that $\boldsymbol{A}_j = \boldsymbol{\Psi}_j - \sum_{k=1}^{j-1}\boldsymbol{\Psi}_{j-k}\boldsymbol{A}_k$, (S2) also holds for $j = k+1$. And hence the induction is completed.

In addition, it can also be verified with $\boldsymbol{\Psi}_j = \boldsymbol{A}_j + \sum_{k=1}^{j-1}\boldsymbol{A}_k\boldsymbol{\Psi}_{j-k}$ for all $j \geqslant 1$. By a method similar to (S2), we can show that

$$\mathbb{M}_1 \subseteq \mathrm{colspace}\{\boldsymbol{A}_j, j \geqslant 1\} \quad \text{and} \quad \mathbb{M}_2 \subseteq \mathrm{rowspace}\{\boldsymbol{A}_j, j \geqslant 1\}.$$

This hence accomplishes the proof.

S3 Proofs of theoretical results in Section 4.1

We first give the proofs of Theorems 1 and 2, and Corollary 1 in Sections B.1-B.4, respectively. Section B.5 gives four lemmas used in the proof of Theorem 1, and two auxiliary lemmas are provided in Section B.6.

S3.1 Proof of Theorem 1

Let $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\mathcal{A}}} - \boldsymbol{\mathcal{A}}^*$. Since $\widehat{\boldsymbol{\mathcal{A}}}, \boldsymbol{\mathcal{A}}^* \in \boldsymbol{\Theta}(r_1, r_2)$, we can show that $\widehat{\boldsymbol{\Delta}} \in \boldsymbol{\Theta}(2r_1, 2r_2) = \{\boldsymbol{\mathcal{A}} \in \mathbb{R}^{N\times N\times T_0} \mid \text{rank}(\boldsymbol{\mathcal{A}}_{(i)}) \leqslant 2r_i, i = 1, 2\}$. By the optimality of $\widehat{\boldsymbol{\mathcal{A}}}$, we have

$$\frac{1}{2T_1}\|\boldsymbol{Y} - \widehat{\boldsymbol{\mathcal{A}}}_{(1)}\boldsymbol{X}\|_{\text{F}}^2 + \lambda\|\widehat{\boldsymbol{\mathcal{A}}}\|_{\ddagger} \leqslant \frac{1}{2T_1}\|\boldsymbol{Y} - \boldsymbol{\mathcal{A}}^*_{(1)}\boldsymbol{X}\|_{\text{F}}^2 + \lambda\|\boldsymbol{\mathcal{A}}^*\|_{\ddagger}. \tag{S1}$$

Define the event

$$\mathcal{E}_1 = \left\{\sup_{\boldsymbol{\Delta}\in\boldsymbol{\Theta}_{\ddagger}(2r_1,2r_2)} \langle \boldsymbol{\Delta}_{(1)}, \frac{1}{T_1}\widetilde{\boldsymbol{\mathcal{E}}}\boldsymbol{X}'\rangle \leqslant \lambda/2\right\},$$

where

$$\boldsymbol{\Theta}_{\ddagger}(r_1, r_2) = \{\boldsymbol{\mathcal{A}} \in \mathbb{R}^{N\times N\times T_0} \mid \text{rank}(\boldsymbol{\mathcal{A}}_{(i)}) \leqslant r_i, i = 1, 2, \|\boldsymbol{\mathcal{A}}\|_{\ddagger} = 1\}. \tag{S2}$$

It is then implied by (S1) that, on the event $\mathcal{E}_1$,

$$\begin{aligned}
\frac{1}{T_1}\|\widehat{\boldsymbol{\Delta}}_{(1)}\boldsymbol{X}\|_{\text{F}}^2 &\leqslant 2\langle \widehat{\boldsymbol{\Delta}}_{(1)}, \frac{1}{T_1}\widetilde{\boldsymbol{\mathcal{E}}}\boldsymbol{X}'\rangle + 2\lambda\left(\|\boldsymbol{\mathcal{A}}^*\|_{\ddagger} - \|\widehat{\boldsymbol{\mathcal{A}}}\|_{\ddagger}\right)\\
&\leqslant 2\|\widehat{\boldsymbol{\Delta}}\|_{\ddagger}\sup_{\boldsymbol{\Delta}\in\boldsymbol{\Theta}_{\ddagger}(2r_1,2r_2)}\langle \boldsymbol{\Delta}_{(1)}, \frac{1}{T_1}\widetilde{\boldsymbol{\mathcal{E}}}\boldsymbol{X}'\rangle + 2\lambda\left(\|\boldsymbol{\mathcal{A}}^*\|_{\ddagger} - \|\widehat{\boldsymbol{\mathcal{A}}}\|_{\ddagger}\right)\\
&\leqslant \lambda\left\{\|\widehat{\boldsymbol{\Delta}}\|_{\ddagger} + 2\left(\|\boldsymbol{\mathcal{A}}^*\|_{\ddagger} - \|\widehat{\boldsymbol{\mathcal{A}}}\|_{\ddagger}\right)\right\}\\
&\leqslant \lambda\left(4\|\boldsymbol{\mathcal{A}}^*_{S^c}\|_{\ddagger} + 3\|\widehat{\boldsymbol{\Delta}}_S\|_{\ddagger} - \|\widehat{\boldsymbol{\Delta}}_{S^c}\|_{\ddagger}\right),
\end{aligned} \tag{S3}$$

where the last inequality follows from (S1) and the triangle inequality. Moreover, since the left hand side of (S3) is non-negative, we have $\widehat{\boldsymbol{\Delta}} \in \mathbb{C}(S) \cap \boldsymbol{\Theta}(2r_1, 2r_2)$, where the restricted set $\mathbb{C}(S)$ is defined as

$$\mathbb{C}(S) = \left\{\boldsymbol{\Delta} \in \mathbb{R}^{N\times N\times T_0} \mid \|\boldsymbol{\Delta}_{S^c}\|_{\ddagger} \leqslant 3\|\boldsymbol{\Delta}_S\|_{\ddagger} + 4\|\boldsymbol{\mathcal{A}}^*_{S^c}\|_{\ddagger}\right\}.$$

For any $\boldsymbol{\Delta} \in \mathbb{C}(S)$, by the triangle inequality and (S2), we have

$$\|\boldsymbol{\Delta}\|_{\ddagger}^2 \leqslant (4\|\boldsymbol{\Delta}_S\|_{\ddagger} + 4\|\boldsymbol{\mathcal{A}}^*_{S^c}\|_{\ddagger})^2 \leqslant 32s\|\boldsymbol{\Delta}\|_{\text{F}}^2 + 32\|\boldsymbol{\mathcal{A}}^*_{S^c}\|_{\ddagger}^2. \tag{S4}$$

On the other hand, let $\mathcal{E}_2$ be the event that the following restricted eigenvalue (RE) condition holds:

$$\frac{1}{T_1}\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2 \geqslant \kappa_{\mathrm{RSC}}\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2 - \tau^2\|\boldsymbol{\Delta}\|_{\ddagger}^2 \quad \text{for all} \quad \boldsymbol{\Delta} \in \mathbb{R}^{N\times N\times T_0},$$

where $\tau^2 = C\sqrt{(N+\log T_0)/T_1}$. Note that, from (S2) and (S3), $T_1{}^{-1}\|\widehat{\boldsymbol{\Delta}}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2 \leqslant 4\lambda\|\boldsymbol{\mathcal{A}}_{S^c}^*\|_{\ddagger} + 3\lambda\sqrt{s}\|\widehat{\boldsymbol{\Delta}}\|_{\mathrm{F}}$. As a result, from (S4) and on the event $\mathcal{E}_1 \cap \mathcal{E}_2$,

$$4\lambda\|\boldsymbol{\mathcal{A}}_{S^c}^*\|_{\ddagger} + 3\lambda\sqrt{s}\|\widehat{\boldsymbol{\Delta}}\|_{\mathrm{F}} \geqslant \left(\kappa_{\mathrm{RSC}} - 32\tau^2 s\right)\|\widehat{\boldsymbol{\Delta}}\|_{\mathrm{F}}^2 - 32\tau^2\|\boldsymbol{\mathcal{A}}_{S^c}^*\|_{\ddagger}^2 \geqslant \frac{\kappa_{\mathrm{RSC}}}{2}\|\widehat{\boldsymbol{\Delta}}\|_{\mathrm{F}}^2 - 32\tau^2\|\boldsymbol{\mathcal{A}}_{S^c}^*\|_{\ddagger}^2, \quad \text{(S5)}$$

if $s \leqslant \kappa_{\mathrm{RSC}}/(64\tau^2)$, i.e., as long as

$$T_1 \gtrsim s^2(N + \log T_0). \quad \text{(S6)}$$

In view of (S5), by solving the quadratic function in $\|\widehat{\boldsymbol{\Delta}}\|_{\mathrm{F}}$ we can show that, on the event $\mathcal{E}_1 \cap \mathcal{E}_2$, if (S6) holds, then

$$\|\widehat{\boldsymbol{\Delta}}\|_{\mathrm{F}}^2 \lesssim \kappa_{\mathrm{RSC}}^{-1}(\lambda^2 s + \lambda\|\boldsymbol{\mathcal{A}}_{S^c}^*\|_{\ddagger} + \tau^2\|\boldsymbol{\mathcal{A}}_{S^c}^*\|_{\ddagger}^2). \quad \text{(S7)}$$

Since

$$\sup_{\boldsymbol{\Delta}\in\boldsymbol{\Theta}_{\ddagger}(2r_1,2r_2)}\langle \boldsymbol{\Delta}_{(1)}, \frac{1}{T_1}\widetilde{\boldsymbol{\mathcal{E}}}\boldsymbol{X}'\rangle \leqslant \sup_{\boldsymbol{\Delta}\in\boldsymbol{\Theta}_{\ddagger}(2r_1,2r_2)}\langle \boldsymbol{\Delta}_{(1)}, \frac{1}{T_1}\boldsymbol{\mathcal{E}}\boldsymbol{X}'\rangle + \sup_{\boldsymbol{\Delta}\in\boldsymbol{\Theta}_{\ddagger}(2r_1,2r_2)}\langle \boldsymbol{\Delta}_{(1)}, \frac{1}{T_1}\boldsymbol{R}\boldsymbol{X}'\rangle,$$

by Lemmas B.1 and B.2, if $\lambda \gtrsim \sqrt{\{(r_1 \wedge r_2)N + \log T_0\}/T_1}$, we then have

$$\mathbb{P}(\mathcal{E}_1^c) \leqslant \mathbb{P}\left\{\sup_{\boldsymbol{\Delta}\in\boldsymbol{\Theta}_{\ddagger}(2r_1,2r_2)}\langle \boldsymbol{\Delta}_{(1)}, \frac{1}{T_1}\widetilde{\boldsymbol{\mathcal{E}}}\boldsymbol{X}'\rangle \geqslant C\sqrt{\frac{(r_1\wedge r_2)N + \log T_0}{T_1}}\right\} \leqslant Ce^{-(r_1\wedge r_2)N - \log T_0}. \quad \text{(S8)}$$

In addition, by Lemma B.3,

$$\mathbb{P}(\mathcal{E}_2^c) \leqslant Ce^{-N-\log T_0}. \quad \text{(S9)}$$

Combining (S7)–(S9), we prove the upper bound for $\|\widehat{\boldsymbol{\Delta}}\|_{\mathrm{F}}$.

Lastly, by (S2) and (S3), we have

$$\frac{1}{T_1}\|\widehat{\boldsymbol{\Delta}}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2 \leqslant \lambda\left(4\|\boldsymbol{\mathcal{A}}_{S^c}^*\|_{\ddagger} + 3\sqrt{s}\|\widehat{\boldsymbol{\Delta}}\|_{\mathrm{F}}\right) \lesssim \lambda\|\boldsymbol{\mathcal{A}}_{S^c}^*\|_{\ddagger} + \lambda^2 s + \|\widehat{\boldsymbol{\Delta}}\|_{\mathrm{F}}^2,$$

since $2\lambda\sqrt{s}\|\widehat{\boldsymbol{\Delta}}\|_{\mathrm{F}} \leqslant \lambda^2 s + \|\widehat{\boldsymbol{\Delta}}\|_{\mathrm{F}}^2$. Combining this with (S7)–(S9) and the condition in (S6), we can similarly prove the upper bound for $T_1{}^{-1}\|\widehat{\boldsymbol{\Delta}}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2$.

S3.2 Proof of Theorem 2

By Assumptions 2 & 4 and the low-rank conditions at (2.1),

$$e_{\text{trunc}} = \sum_{j=T_0+1}^{\infty} \|\boldsymbol{A}_j^*\|_{\text{F}}^2 \leqslant \frac{(r_1 \wedge r_2)\rho^{2T_0}}{1-\rho} \lesssim \frac{r_1 \wedge r_2}{T_1^4}, \tag{S10}$$

which is clearly dominated by $\|\widehat{\boldsymbol{\Delta}}\|_{\text{F}}^2$ in (S7) under the condition on λ. Therefore, combining (S7) and (S10), we have that, with probability at least $1 - Ce^{-(r_1 \wedge r_2)N - \log T_0}$,

$$e_{\text{est}}(\widehat{\boldsymbol{\mathcal{A}}}_\infty) = \|\widehat{\boldsymbol{\mathcal{A}}}_\infty - \boldsymbol{\mathcal{A}}_\infty^*\|_{\text{F}}^2 = \|\widehat{\boldsymbol{\Delta}}\|_{\text{F}}^2 + e_{\text{trunc}} \lesssim \kappa_{\text{RSC}}^{-1}(\lambda^2 s + \lambda \|\boldsymbol{\mathcal{A}}_{S^c}^*\|_{\ddagger} + \tau^2 \|\boldsymbol{\mathcal{A}}_{S^c}^*\|_{\ddagger}^2). \tag{S11}$$

Moreover, by (S2) and (S3), we have with probability at least $1 - Ce^{-(r_1 \wedge r_2)N - \log T_0}$,

$$\begin{aligned} e_{\text{pred}}(\widehat{\boldsymbol{\mathcal{A}}}_\infty) &= \frac{1}{T_1} \sum_{t=T_0+1}^{T} \|\widehat{\boldsymbol{\Delta}}_{(1)} \mathbf{x}_t + \boldsymbol{r}_t\|_2^2 \\ &= \frac{1}{T_1} \|\widehat{\boldsymbol{\Delta}}_{(1)} \boldsymbol{X}\|_{\text{F}}^2 + \frac{2}{T_1} \langle \widehat{\boldsymbol{\Delta}}_{(1)} \boldsymbol{X}, \boldsymbol{R} \rangle + \frac{1}{T_1} \|\boldsymbol{R}\|_{\text{F}}^2 \\ &\leqslant \frac{1}{T_1} \|\widehat{\boldsymbol{\Delta}}_{(1)} \boldsymbol{X}\|_{\text{F}}^2 + 2\|\widehat{\boldsymbol{\Delta}}\|_{\text{F}} \sup_{\boldsymbol{\Delta} \in \boldsymbol{\Theta}_{\ddagger}(2r_1, 2r_2)} \langle \boldsymbol{\Delta}_{(1)}, \frac{1}{T_1} \boldsymbol{R} \boldsymbol{X}' \rangle + \frac{1}{T_1} \|\boldsymbol{R}\|_{\text{F}}^2 \\ &\lesssim \kappa_{\text{RSC}}^{-1}(\lambda^2 s + \lambda \|\boldsymbol{\mathcal{A}}_{S^c}^*\|_{\ddagger} + \tau^2 \|\boldsymbol{\mathcal{A}}_{S^c}^*\|_{\ddagger}^2), \end{aligned} \tag{S12}$$

where the last inequality uses Theorem 1, Lemmas B.2 and B.4.

Next, we characterize the optimal bounds for $e_{\text{est}}(\widehat{\boldsymbol{\mathcal{A}}}_\infty)$ in (S11) and $e_{\text{pred}}(\widehat{\boldsymbol{\mathcal{A}}}_\infty)$ in (S12). Consider a family of subsets indexed by a threshold $\gamma > 0$: $S_\gamma = \{j \in \{1, \ldots, T_0\} \mid \|\boldsymbol{A}_j^*\|_{\text{F}} > \gamma\}$, and let $S_\gamma^c = \{1, \ldots, T_0\} \backslash S_\gamma$. Note that under Assumption 2 and the low-rank conditions at (2.4), $\|\boldsymbol{A}_j^*\|_{\text{F}} \leqslant C\sqrt{r_1 \wedge r_2}\rho^j$ for $j \geqslant 1$. Let Q_γ be the smallest integer such that $C\sqrt{r_1 \wedge r_2}\rho^j \leqslant \gamma$ for all $j \geqslant Q_\gamma$. Then,

$$Q_\gamma = \left\lceil \frac{\log(C\sqrt{r_1 \wedge r_2}/\gamma)}{\log(1/\rho)} \right\rceil, \tag{S13}$$

where $\lceil \cdot \rceil$ is the ceiling function. Moreover, since $C\sqrt{r_1 \wedge r_2}\rho^{Q_\gamma} \leqslant \gamma$, we have

$$\|\boldsymbol{\mathcal{A}}_{S_\gamma^c}^*\|_{\ddagger} = \sum_{j \in S_\gamma^c \cap \{1, \ldots, Q_\gamma\}} \|\boldsymbol{A}_j^*\|_{\text{F}} + \sum_{j=Q_\gamma+1}^{T_0} \|\boldsymbol{A}_j^*\|_{\text{F}} \leqslant \gamma Q_\gamma + \sum_{j=Q_\gamma+1}^{\infty} C\sqrt{r_1 \wedge r_2}\rho^j \lesssim \gamma Q_\gamma, \tag{S14}$$

as long as $Q_\gamma \geqslant c$ for some absolute constant $c > 0$. Since $\|\boldsymbol{A}_j^*\|_{\text{F}} > \gamma \geqslant C\sqrt{N}\rho^{Q_\gamma}$ for any $j \in S_\gamma$, we have $S_\gamma \subseteq \{1, \ldots, Q_\gamma\}$, and hence $|S_\gamma| \leqslant Q_\gamma$. Then, the upper bounds of $\lambda^2|S_\gamma|$ and $\lambda\|\boldsymbol{\mathcal{A}}_{S_\gamma^c}^*\|_{\ddagger}$, i.e., $\lambda^2 Q_\gamma$ and $\lambda\gamma Q_\gamma$, are balanced when $\gamma \asymp \lambda$. With this choice of γ and the rate of λ specified in the theorem, by (S13), we can show that

$$Q_\gamma \lesssim \frac{\log T_1}{\log(1/\rho)}$$

and $\tau^2 Q_\gamma \leqslant C$. Thus, $\tau^2 \|\boldsymbol{\mathcal{A}}^*_{S^c_\gamma}\|^2_\ddagger$ is dominated by $\lambda \|\boldsymbol{\mathcal{A}}^*_{S^c_\gamma}\|_\ddagger$. From (S11), $e_{\text{est}}(\widehat{\boldsymbol{\mathcal{A}}}_\infty) \lesssim \lambda^2 Q_\gamma$. In addition, substituting the upper bound of $|S_\gamma|$, i.e., $\log T_1/\log(1/\rho)$, into the sample size condition in Theorem 1, we have (4.2). As a result, the upper bound for $e_{\text{est}}(\widehat{\boldsymbol{\mathcal{A}}}_\infty)$ in this theorem holds. Finally, the result for $e_{\text{pred}}(\widehat{\boldsymbol{\mathcal{A}}}_\infty)$ can be proved by a similar method.

S3.3 Proof of Corollary 1

This proof largely follows from the proof of Theorem 1 in Section S3.1. For the finite-order AR model with a fixed order T_0, $\boldsymbol{A}^*_j = \boldsymbol{0}$ for all $j \geqslant T_0 + 1$ leading to $\boldsymbol{R} = \boldsymbol{0}$ and $\widetilde{\boldsymbol{\mathcal{E}}} = \boldsymbol{\mathcal{E}}$. Subsequently, the event $\mathcal{E}_1$ becomes

$$\mathcal{E}_1 = \left\{ \sup_{\boldsymbol{\Delta} \in \boldsymbol{\Theta}_\ddagger(2r_1, 2r_2)} \langle \boldsymbol{\Delta}_{(1)}, \frac{1}{T_1} \boldsymbol{\mathcal{E}} \boldsymbol{X}' \rangle \leqslant \lambda/2 \right\},$$

and its probability can be established directly from Lemma B.1. Meanwhile, the other intermediate steps are the same as in the proof of Theorem 1 in Section S3.1.

S3.4 Four lemmas used in the proof of Theorem 1

We first state the four lemmas and then give their technical proofs.

Lemma B.1. *Suppose that Assumptions 1 – 3 hold. If $T_1 \gtrsim (r_1 \wedge r_2)N + \log T_0$, then*

$$\mathbb{P}\left\{ \sup_{\boldsymbol{\Delta} \in \boldsymbol{\Theta}_\ddagger(2r_1, 2r_2)} \langle \boldsymbol{\Delta}_{(1)}, \frac{1}{T_1} \boldsymbol{\mathcal{E}} \boldsymbol{X}' \rangle \geqslant C \sqrt{\frac{(r_1 \wedge r_2)N + \log T_0}{T_1}} \right\} \leqslant C e^{-(r_1 \wedge r_2)N - \log T_0}.$$

Lemma B.2. *Suppose that Assumptions 1 – 4 hold. If $T_1 \gtrsim (r_1 \wedge r_2)N + \log T_0$, then*

$$\mathbb{P}\left\{ \sup_{\boldsymbol{\Delta} \in \boldsymbol{\Theta}_\ddagger(2r_1, 2r_2)} \langle \boldsymbol{\Delta}_{(1)}, \frac{1}{T_1} \boldsymbol{R} \boldsymbol{X}' \rangle \geqslant C \frac{\sqrt{N}}{T_1} \right\} \leqslant C e^{-(r_1 \wedge r_2)N - \log T_0}.$$

Lemma B.3 (RE condition)**.** *Suppose that Assumptions 1 – 3 hold. If* $T_1 \gtrsim N + \log T_0$*, then with probability at least* $1 - Ce^{-N-\log T_0}$*, the following RE condition is satisfied:*

$$\frac{1}{T_1}\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2 \geqslant \kappa_{\mathrm{RSC}}\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2 - \tau^2\|\boldsymbol{\Delta}\|_{\ddagger}^2 \quad \text{for all} \quad \boldsymbol{\Delta} \in \mathbb{R}^{N\times N\times T_0},$$

where $\kappa_{\mathrm{RSC}} = \lambda_{\min}(\boldsymbol{\Sigma}_\varepsilon)\mu_{\min}(\boldsymbol{\Psi}_*)$ *and* $\tau^2 = C\sqrt{(N + \log T_0)/T_1}$*.*

Lemma B.4. *If Assumptions 1 – 4 hold, then*

$$\mathbb{P}\left\{\frac{1}{T_1}\|\boldsymbol{R}\|_{\mathrm{F}}^2 \geqslant C\frac{N^{3/2}\rho^{T_0}}{T_1}\right\} \leqslant Ce^{-N}.$$

S3.4.1 Proof of Lemma B.1

The proof of Lemma B.1 relies on the following discretization result.

Lemma B.5 (Discretization)**.** *Let* $\bar{\boldsymbol{\Theta}}_{\mathrm{F}}(2r_{\min})$ *be a minimal* $1/2$*-net for* $\boldsymbol{\Theta}_{\mathrm{F}}(2r_{\min})$ *in the Frobenius norm.*

$$\sup_{\boldsymbol{\Delta}\in\boldsymbol{\Theta}_{\ddagger}(2r_1,2r_2)}\langle\boldsymbol{\Delta}_{(1)}, \boldsymbol{\mathcal{E}}\boldsymbol{X}'\rangle \leqslant 4\max_{1\leqslant j\leqslant T_0}\max_{\boldsymbol{M}\in\bar{\boldsymbol{\Theta}}_{\mathrm{F}}(2r_{\min})}\langle\boldsymbol{M}, \boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\rangle,$$

where $r_{\min} = r_1 \wedge r_2$*, and* $\boldsymbol{\Theta}_{\ddagger}(2r_1, 2r_2)$ *is defined as in* (S2)*.*

Proof of Lemma B.5. For any $\boldsymbol{\Delta} \in \boldsymbol{\Theta}_{\ddagger}(2r_1, 2r_2)$, there exist $\boldsymbol{\mathcal{H}} \in \mathbb{R}^{2r_1\times 2r_2\times T_0}$, $\boldsymbol{U}_1 \in \mathcal{O}^{N\times 2r_1}$ and $\boldsymbol{U}_2 \in \mathcal{O}^{N\times 2r_2}$ such that

$$\boldsymbol{\Delta} = \boldsymbol{\mathcal{H}} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2.$$

By the orthonormality of $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$, we have $\|\boldsymbol{\mathcal{H}}\|_{\ddagger} = \|\boldsymbol{\Delta}\|_{\ddagger}$. Let $\boldsymbol{\mathcal{H}}_{(1)} = (\boldsymbol{H}_1, \ldots, \boldsymbol{H}_{T_0})$, where $\boldsymbol{H}_j \in \mathbb{R}^{2r_1\times 2r_2}$ are the frontal slices. Note that

$$\langle\boldsymbol{\Delta}_{(1)}, \boldsymbol{\mathcal{E}}\boldsymbol{X}'\rangle = \langle\boldsymbol{U}_1\boldsymbol{\mathcal{H}}_{(1)}(\boldsymbol{I}_{T_0}\otimes\boldsymbol{U}_2)', \boldsymbol{\mathcal{E}}\boldsymbol{X}'\rangle = \langle\boldsymbol{\mathcal{H}}_{(1)}, \boldsymbol{U}_1'\boldsymbol{\mathcal{E}}\boldsymbol{X}'(\boldsymbol{I}_{T_0}\otimes\boldsymbol{U}_2)\rangle = \sum_{j=1}^{T_0}\langle\boldsymbol{H}_j, \boldsymbol{U}_1'\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\boldsymbol{U}_2\rangle.$$

Then we can show that

$$\begin{aligned}
\sup_{\boldsymbol{\Delta}\in\boldsymbol{\Theta}_{\ddagger}(2r_1,2r_2)}\langle\boldsymbol{\Delta}_{(1)},\boldsymbol{\mathcal{E}}\boldsymbol{X}'\rangle &\leqslant \sup_{\boldsymbol{U}_1\in\mathcal{O}^{N\times 2r_1},\boldsymbol{U}_2\in\mathcal{O}^{N\times 2r_2}}\ \sup_{\sum_{j=1}^{T_0}\|\boldsymbol{H}_j\|_{\mathrm{F}}=1}\sum_{j=1}^{T_0}\langle\boldsymbol{H}_j,\boldsymbol{U}_1'\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\boldsymbol{U}_2\rangle \\
&= \sup_{\boldsymbol{U}_1\in\mathcal{O}^{N\times 2r_1},\boldsymbol{U}_2\in\mathcal{O}^{N\times 2r_2}}\ \max_{1\leqslant j\leqslant T_0}\|\boldsymbol{U}_1'\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\boldsymbol{U}_2\|_{\mathrm{F}} \\
&= \max_{1\leqslant j\leqslant T_0}\ \sup_{\boldsymbol{U}_1\in\mathcal{O}^{N\times 2r_1},\boldsymbol{U}_2\in\mathcal{O}^{N\times 2r_2}}\ \sup_{\boldsymbol{M}\in\mathcal{S}^{2r_1\times 2r_2}}\langle\boldsymbol{M},\boldsymbol{U}_1'\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\boldsymbol{U}_2\rangle \qquad \text{(S15)} \\
&= \max_{1\leqslant j\leqslant T_0}\ \sup_{\boldsymbol{U}_1\in\mathcal{O}^{N\times 2r_1},\boldsymbol{U}_2\in\mathcal{O}^{N\times 2r_2}}\ \sup_{\boldsymbol{M}\in\mathcal{S}^{2r_1\times 2r_2}}\langle\boldsymbol{U}_1\boldsymbol{M}\boldsymbol{U}_2',\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\rangle \\
&= \max_{1\leqslant j\leqslant T_0}\ \sup_{\boldsymbol{M}\in\boldsymbol{\Theta}_{\mathrm{F}}(2r_{\min})}\langle\boldsymbol{M},\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\rangle.
\end{aligned}$$

Since $\bar{\boldsymbol{\Theta}}_{\mathrm{F}}(2r_{\min})$ is a minimal 1/2-net for $\boldsymbol{\Theta}_{\mathrm{F}}(2r_{\min})$ in the Frobenius norm, for any $\boldsymbol{M}\in\boldsymbol{\Theta}_{\mathrm{F}}(2r_{\min})$, there exists $\overline{\boldsymbol{M}}\in\bar{\boldsymbol{\Theta}}_{\mathrm{F}}(2r_{\min})$ such that $\|\boldsymbol{M}-\overline{\boldsymbol{M}}\|_{\mathrm{F}}\leqslant 1/2$. Notice that $\boldsymbol{M}-\overline{\boldsymbol{M}}$ is of rank at most $4r_{\min}$. Then, we can find $\boldsymbol{T}_1,\boldsymbol{T}_2$, each with rank at most $2r_{\min}$, such that $\boldsymbol{M}-\overline{\boldsymbol{M}}=\boldsymbol{T}_1+\boldsymbol{T}_2$ and $\langle\boldsymbol{T}_1,\boldsymbol{T}_2\rangle=0$. Moreover, it holds $\|\boldsymbol{T}_1\|_{\mathrm{F}}+\|\boldsymbol{T}_2\|_{\mathrm{F}}\leqslant\sqrt{2}\|\boldsymbol{T}_1+\boldsymbol{T}_2\|_{\mathrm{F}}=\sqrt{2}\|\boldsymbol{M}-\overline{\boldsymbol{M}}\|_{\mathrm{F}}\leqslant\sqrt{2}/2$, and $\boldsymbol{T}_i/\|\boldsymbol{T}_i\|_{\mathrm{F}}\in\boldsymbol{\Theta}_{\mathrm{F}}(2r_{\min})$. As a result, we can show that

$$\begin{aligned}
\sup_{\boldsymbol{M}\in\boldsymbol{\Theta}_{\mathrm{F}}(2r_{\min})}\langle\boldsymbol{M},\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\rangle &= \sup_{\boldsymbol{M}\in\boldsymbol{\Theta}_{\mathrm{F}}(2r_{\min})}\left\{\langle\overline{\boldsymbol{M}},\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\rangle+\sum_{i=1}^{2}\langle\frac{\boldsymbol{T}_i}{\|\boldsymbol{T}_i\|_{\mathrm{F}}},\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\rangle\|\boldsymbol{T}_i\|_{\mathrm{F}}\right\} \\
&\leqslant \max_{\overline{\boldsymbol{M}}\in\bar{\boldsymbol{\Theta}}_{\mathrm{F}}(2r_{\min})}\langle\overline{\boldsymbol{M}},\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\rangle+(\|\boldsymbol{T}_1\|_{\mathrm{F}}+\|\boldsymbol{T}_2\|_{\mathrm{F}})\sup_{\boldsymbol{M}\in\boldsymbol{\Theta}_{\mathrm{F}}(2r_{\min})}\langle\boldsymbol{M},\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\rangle \\
&\leqslant \max_{\overline{\boldsymbol{M}}\in\bar{\boldsymbol{\Theta}}_{\mathrm{F}}(2r_{\min})}\langle\overline{\boldsymbol{M}},\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\rangle+\frac{\sqrt{2}}{2}\sup_{\boldsymbol{M}\in\boldsymbol{\Theta}_{\mathrm{F}}(2r_{\min})}\langle\boldsymbol{M},\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\rangle,
\end{aligned}$$

which implies

$$\sup_{\boldsymbol{M}\in\boldsymbol{\Theta}_{\mathrm{F}}(2r_{\min})}\langle\boldsymbol{M},\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\rangle\leqslant 4\max_{\boldsymbol{M}\in\bar{\boldsymbol{\Theta}}_{\mathrm{F}}(2r_{\min})}\langle\boldsymbol{M},\boldsymbol{\mathcal{E}}\boldsymbol{X}_j'\rangle. \qquad \text{(S16)}$$

By combining (S15) and (S16), we accomplish the proof of this lemma. □

Now we are ready to prove Lemma B.1.

Proof of Lemma B.1. Let $\bar{\boldsymbol{\Theta}}_{\mathrm{F}}(2r_{\min})$ be a minimal 1/2-net of $\boldsymbol{\Theta}_{\mathrm{F}}(2r_{\min})$ in the Frobenius norm, where $r_{\min}=r_1\wedge r_2$. Then its cardinality satisfies

$$\log|\bar{\boldsymbol{\Theta}}_{\mathrm{F}}(2r_{\min})|\leqslant(4N+2)r_{\min}\log 18\leqslant 18r_{\min}N; \qquad \text{(S17)}$$

see Lemma 3.1 of Candes and Plan (2011). Denote $\check{D}_{\mathcal{M}} = 18r_{\min}N + \log T_0$. By (S17) and Lemma B.5, for any $K > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\Delta}\in\boldsymbol{\Theta}_{\ddagger}(2r_1,2r_2)}\langle\boldsymbol{\Delta}_{(1)},\frac{1}{T_1}\boldsymbol{\mathcal{E}}\boldsymbol{X}'\rangle \geqslant K\right)\\
&\leqslant \mathbb{P}\left(\max_{1\leqslant k\leqslant T_0}\max_{\boldsymbol{M}\in\bar{\boldsymbol{\Theta}}_{\mathrm{F}}(2r_{\min})}\frac{1}{T_1}\langle\boldsymbol{M},\boldsymbol{\mathcal{E}}\boldsymbol{X}_k'\rangle \geqslant K/4\right)\\
&\leqslant T_0|\bar{\boldsymbol{\Theta}}_{\mathrm{F}}(2r_{\min})|\max_{1\leqslant k\leqslant T_0}\max_{\|\boldsymbol{M}\|_{\mathrm{F}}=1}\mathbb{P}\left(\frac{1}{T_1}\langle\boldsymbol{M},\boldsymbol{\mathcal{E}}\boldsymbol{X}_k'\rangle \geqslant K/4\right)\\
&\leqslant \exp(\check{D}_{\mathcal{M}})\max_{1\leqslant k\leqslant T_0}\max_{\|\boldsymbol{M}\|_{\mathrm{F}}=1}\mathbb{P}\left(\frac{1}{T_1}\langle\boldsymbol{M},\boldsymbol{\mathcal{E}}\boldsymbol{X}_k'\rangle \geqslant K/4\right).
\end{aligned} \tag{S18}$$

Since $\boldsymbol{\varepsilon}_t = \boldsymbol{\Sigma}_\varepsilon^{1/2}\boldsymbol{\xi}_t$ and $\mathbf{y}_t = \sum_{j=0}^{\infty}\boldsymbol{\Psi}_j^*\boldsymbol{\Sigma}_\varepsilon^{1/2}\boldsymbol{\xi}_{t-j}$, for any $1\leqslant k\leqslant T_0$, we have

$$\begin{aligned}
\langle\boldsymbol{M},\boldsymbol{\mathcal{E}}\boldsymbol{X}_k'\rangle = \sum_{t=T_0+1}^{T}\langle\boldsymbol{M},\boldsymbol{\varepsilon}_t\mathbf{y}_{t-k}'\rangle &= \sum_{t=T_0+1}^{T}\left\langle\boldsymbol{M},\boldsymbol{\Sigma}_\varepsilon^{1/2}\boldsymbol{\xi}_t\sum_{j=0}^{\infty}\boldsymbol{\xi}_{t-k-j}'\boldsymbol{\Sigma}_\varepsilon^{1/2}\boldsymbol{\Psi}_j^{*\prime}\right\rangle\\
&= \sum_{j=0}^{\infty}\sum_{t=T_0+1}^{T}\langle\boldsymbol{M},\boldsymbol{\Sigma}_\varepsilon^{1/2}\boldsymbol{\xi}_t\boldsymbol{\xi}_{t-k-j}'\boldsymbol{\Sigma}_\varepsilon^{1/2}\boldsymbol{\Psi}_j^{*\prime}\rangle\\
&= \sum_{j=0}^{\infty}\sum_{t=T_0+1}^{T}\langle\widetilde{\boldsymbol{M}}_j\boldsymbol{\xi}_{t-k-j},\boldsymbol{\xi}_t\rangle,
\end{aligned}$$

where

$$\widetilde{\boldsymbol{M}}_j = \boldsymbol{\Sigma}_\varepsilon^{1/2}\boldsymbol{M}\boldsymbol{\Psi}_j^*\boldsymbol{\Sigma}_\varepsilon^{1/2}.$$

Note that for any fixed k and j and any $\delta_j\in(0,1]$, by Lemma B.7(i), we have

$$\mathbb{P}\left\{\sum_{t=T_0+1}^{T}\langle\widetilde{\boldsymbol{M}}_j\boldsymbol{\xi}_{t-k-j},\boldsymbol{\xi}_t\rangle \geqslant C\sigma^2\|\widetilde{\boldsymbol{M}}_j\|_{\mathrm{F}}\left\{\log(1/\delta_j)+\sqrt{T_1\log(1/\delta_j)}\right\}\right\}\leqslant 2\delta_j.$$

For simplicity, denote

$$a_j = \sigma^2\|\widetilde{\boldsymbol{M}}_j\|_{\mathrm{F}}\left\{\log(1/\delta_j)+\sqrt{T_1\log(1/\delta_j)}\right\}.$$

Then it follows that

$$\mathbb{P}\left(\frac{1}{T_1}\langle\boldsymbol{M},\boldsymbol{\mathcal{E}}\boldsymbol{X}_k'\rangle \geqslant \frac{C}{T_1}\sum_{j=0}^{\infty}a_j\right)\leqslant 2\sum_{j=0}^{\infty}\delta_j. \tag{S19}$$

Moreover, by Assumption 2, if $\|\boldsymbol{M}\|_{\mathrm{F}} = 1$, then $\|\widetilde{\boldsymbol{M}}_j\|_{\mathrm{F}} \leqslant \lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon)\|\boldsymbol{M}\|_{\mathrm{F}}\|\boldsymbol{\Psi}_j^*\|_{\mathrm{op}} \leqslant C\rho^j$. By choosing

$$\delta_j = \exp\left\{-4\rho^{-(j+1)/2}\check{D}_{\mathcal{M}}/\log(1/\rho)\right\},$$

i.e., $\log(1/\delta_j) = 4\rho^{-(j+1)/2}\check{D}_{\mathcal{M}}/\log(1/\rho)$, we can show that

$$\begin{aligned}
\frac{1}{T_1}\sum_{j=0}^{\infty} a_j &\leqslant \frac{4\sigma^2}{T_1}\sum_{j=0}^{\infty}\left\{\frac{\rho^{(j-1)/2}\check{D}_{\mathcal{M}}}{\log(1/\rho)} + \rho^{(3j-1)/4}\sqrt{\frac{T_1\check{D}_{\mathcal{M}}}{\log(1/\rho)}}\right\} \\
&= 4\sigma^2\left\{\frac{1}{\sqrt{\rho}-\rho}\cdot\frac{\check{D}_{\mathcal{M}}}{T_1\log(1/\rho)} + \frac{1}{\rho^{1/4}-\rho}\cdot\sqrt{\frac{\check{D}_{\mathcal{M}}}{T_1\log(1/\rho)}}\right\} \\
&\leqslant C\sqrt{\frac{\check{D}_{\mathcal{M}}}{T_1}}. \qquad \text{(S20)}
\end{aligned}$$

In addition, using the inequality $x^k \geqslant k\log x$ for any $k \geqslant 0$ and $x > 1$, we can show that

$$\sum_{j=0}^{\infty}\delta_j \leqslant \sum_{j=0}^{\infty} e^{-2(j+1)\check{D}_{\mathcal{M}}} = \frac{e^{-2\check{D}_{\mathcal{M}}}}{1-e^{-2\check{D}_{\mathcal{M}}}}. \qquad \text{(S21)}$$

Combining (S19)–(S21), if $\|\boldsymbol{M}\|_{\mathrm{F}} = 1$, for any $1 \leqslant k \leqslant P$, we have

$$\mathbb{P}\left\{\frac{1}{T_1}\langle \boldsymbol{M}, \boldsymbol{\mathcal{E}}\boldsymbol{X}'_k\rangle \geqslant C\sqrt{\frac{\check{D}_{\mathcal{M}}}{T_1}}\right\} \leqslant \frac{2e^{-2\check{D}_{\mathcal{M}}}}{1-e^{-2\check{D}_{\mathcal{M}}}} \leqslant Ce^{-\check{D}_{\mathcal{M}}},$$

which together with (S18) and $T_1 \gtrsim \check{D}_{\mathcal{M}}$ implies the result of Lemma B.1. □

S3.4.2 Proof of Lemma B.2

Note that $\boldsymbol{R}\boldsymbol{X}' = (\boldsymbol{R}\boldsymbol{X}'_1, \ldots, \boldsymbol{R}\boldsymbol{X}'_{T_0})$. Along the lines of Lemma B.5 in Section S3.4.1, we can show the following discretization result:

$$\sup_{\boldsymbol{\Delta}\in\boldsymbol{\Theta}_{\ddagger}(2r_1,2r_2)}\langle\boldsymbol{\Delta}_{(1)}, \boldsymbol{R}\boldsymbol{X}'\rangle \leqslant 4\max_{1\leqslant k\leqslant T_0}\max_{\boldsymbol{M}\in\bar{\boldsymbol{\Theta}}_2(2r_{\min})}\langle\boldsymbol{M}, \boldsymbol{R}\boldsymbol{X}'_k\rangle.$$

Then, similarly to (S18), for any $K > 0$, by (S17), we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\Delta}\in\boldsymbol{\Theta}_{\ddagger}(2r_1,2r_2)}\langle\boldsymbol{\Delta}_{(1)}, \frac{1}{T_1}\boldsymbol{R}\boldsymbol{X}'\rangle \geqslant K\right) \\
&\qquad \leqslant \exp(\check{D}_{\mathcal{M}})\max_{1\leqslant k\leqslant T_0}\max_{\|\boldsymbol{M}\|_{\mathrm{F}}=1}\mathbb{P}\left(\frac{1}{T_1}\langle\boldsymbol{M}, \boldsymbol{R}\boldsymbol{X}'_k\rangle \geqslant K/4\right), \qquad \text{(S22)}
\end{aligned}$$

where $\check{D}_{\mathcal{M}} = 18(r_1 \wedge r_2)N + \log T_0$.

Since $\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{\Psi}_j^* \mathbf{\Sigma}_\varepsilon^{1/2} \boldsymbol{\xi}_{t-j}$, for any $1 \leqslant k \leqslant T_0$, we have

$$
\begin{aligned}
\langle \boldsymbol{M}, \boldsymbol{R}\boldsymbol{X}_k' \rangle &= \sum_{t=T_0+1}^{T} \langle \boldsymbol{M}, \boldsymbol{r}_t \mathbf{y}_{t-k}' \rangle \\
&= \sum_{t=T_0+1}^{T} \left\langle \boldsymbol{M}, \sum_{j=T_0+1}^{\infty} \boldsymbol{A}_j^* \sum_{i=0}^{\infty} \mathbf{\Psi}_i^* \mathbf{\Sigma}_\varepsilon^{1/2} \boldsymbol{\xi}_{t-j-i} \sum_{\ell=0}^{\infty} \boldsymbol{\xi}_{t-k-\ell}' \mathbf{\Sigma}_\varepsilon^{1/2} \mathbf{\Psi}_\ell^{*\prime} \right\rangle \\
&= \sum_{j=T_0+1}^{\infty} \sum_{i=0}^{\infty} \sum_{\ell=0}^{\infty} \sum_{t=T_0+1}^{T} \langle \boldsymbol{M}, \boldsymbol{A}_j^* \mathbf{\Psi}_i^* \mathbf{\Sigma}_\varepsilon^{1/2} \boldsymbol{\xi}_{t-j-i} \boldsymbol{\xi}_{t-k-\ell}' \mathbf{\Sigma}_\varepsilon^{1/2} \mathbf{\Psi}_\ell^{*\prime} \rangle \\
&= \sum_{j=T_0+1}^{\infty} \sum_{i=0}^{\infty} \sum_{\ell=0}^{\infty} \sum_{t=T_0+1}^{T} \langle \widetilde{\boldsymbol{M}}_{i,j,\ell} \boldsymbol{\xi}_{t-k-\ell}, \boldsymbol{\xi}_{t-j-i} \rangle := B_1 + B_2,
\end{aligned} \tag{S23}
$$

where

$$
\begin{aligned}
B_1 &= \sum_{j=T_0+1}^{\infty} \sum_{i=0}^{\infty} \sum_{\ell=0}^{\infty} \sum_{t=T_0+1}^{T} \langle \widetilde{\boldsymbol{M}}_{i,j,\ell} \boldsymbol{\xi}_{t-k-\ell}, \boldsymbol{\xi}_{t-j-i} \rangle I\{\ell \neq i+j-k\}, \\
B_2 &= \sum_{j=T_0+1}^{\infty} \sum_{i=0}^{\infty} \sum_{\ell=0}^{\infty} \sum_{t=T_0+1}^{T} \langle \widetilde{\boldsymbol{M}}_{i,j,\ell} \boldsymbol{\xi}_{t-k-\ell}, \boldsymbol{\xi}_{t-j-i} \rangle I\{\ell = i+j-k\},
\end{aligned}
$$

with $I(\cdot)$ being the indicator function, and

$$
\widetilde{\boldsymbol{M}}_{i,j,\ell} = \mathbf{\Sigma}_\varepsilon^{1/2} \mathbf{\Psi}_i^{*\prime} \boldsymbol{A}_j^{*\prime} \boldsymbol{M} \mathbf{\Psi}_\ell^* \mathbf{\Sigma}_\varepsilon^{1/2}.
$$

Note that for any fixed i, j, ℓ, we can write $\sum_{t=T_0+1}^{T} \langle \widetilde{\boldsymbol{M}}_{i,j,\ell} \boldsymbol{\xi}_{t-k-\ell}, \boldsymbol{\xi}_{t-j-i} \rangle = \sum_{t=T_0'}^{T_1'} \langle \widetilde{\boldsymbol{M}}_{i,j,\ell} \boldsymbol{\xi}_{t-s}, \boldsymbol{\xi}_t \rangle$, with $T_0' = T_0 + 1 - j - i$, $T_1' = T - j - i$ and $s = k + \ell - j - i$ (cf. Lemma B.7). In addition, note that $T_1' - T_0' + 1 = T_1$.

Then, by Lemma B.7 and a method similar to that for (S19), for any fixed k, we can show that

$$
\mathbb{P}\left(B_1 \geqslant \sum_{j=T_0+1}^{\infty} \sum_{i=0}^{\infty} \sum_{\ell=0}^{\infty} C b_{i,j,\ell} I\{\ell \neq i+j-k\} \right) \leqslant \sum_{j=T_0+1}^{\infty} \sum_{i=0}^{\infty} \sum_{\ell=0}^{\infty} 2\delta_{i,j,\ell} I\{\ell \neq i+j-k\}
$$

and

$$
\mathbb{P}\left\{ B_2 \geqslant \sum_{j=T_0+1}^{\infty} \sum_{i=0}^{\infty} \sum_{\ell=0}^{\infty} (C b_{i,j,\ell} + c_{i,j,\ell}) I\{\ell = i+j-k\} \right\} \leqslant \sum_{j=T_0+1}^{\infty} \sum_{i=0}^{\infty} \sum_{\ell=0}^{\infty} \delta_{i,j,\ell} I\{\ell = i+j-k\},
$$

where $\delta_{i,j,\ell} \in (0,1)$ will be specified shortly (see (S26) below), and

$$
b_{i,j,\ell} = \sigma^2 \|\widetilde{\boldsymbol{M}}_{i,j,\ell}\|_{\mathrm{F}} \left\{ \log(1/\delta_{i,j,\ell}) + \sqrt{T_1 \log(1/\delta_{i,j,\ell})} \right\} \quad \text{and} \quad c_{i,j,\ell} = T_1 \sqrt{N} \|\widetilde{\boldsymbol{M}}_{i,j,\ell}\|_{\mathrm{F}}.
$$

Combining the above inequalities with (S23), we have

$$\mathbb{P}\left\{\frac{1}{T_1}\langle \boldsymbol{M}, \boldsymbol{R}\boldsymbol{X}_k'\rangle \geqslant \frac{C}{T_1}\sum_{j=T_0+1}^{\infty}\sum_{i=0}^{\infty}\sum_{\ell=0}^{\infty} b_{i,j,\ell} + \frac{1}{T_1}\sum_{j=T_0+1}^{\infty}\sum_{i=0}^{\infty}\sum_{\ell=0}^{\infty} c_{i,j,\ell} I\{\ell = i+j-k\}\right\}$$
$$\leqslant 2\sum_{j=T_0+1}^{\infty}\sum_{i=0}^{\infty}\sum_{\ell=0}^{\infty}\delta_{i,j,\ell}. \tag{S24}$$

Note that by Assumption 2, if $\|\boldsymbol{M}\|_{\mathrm{F}} = 1$, we can upper bound each $\|\widetilde{\boldsymbol{M}}_{i,j,\ell}\|_{\mathrm{F}}$ as follows:

$$\|\widetilde{\boldsymbol{M}}_{i,j,\ell}\|_{\mathrm{F}} \leqslant \lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon)\|\boldsymbol{A}_j^*\|_{\mathrm{op}}\|\boldsymbol{\Psi}_i^*\|_{\mathrm{op}}\|\boldsymbol{\Psi}_\ell^*\|_{\mathrm{op}}\|\boldsymbol{M}\|_{\mathrm{F}} \leqslant C\rho^{i+j+\ell}.$$

Then, for any $1 \leqslant k \leqslant T_0$, we have

$$\begin{aligned}
\sum_{j=T_0+1}^{\infty}\sum_{i=0}^{\infty}\sum_{\ell=0}^{\infty} c_{i,j,\ell} I\{\ell = i+j-k\} &\leqslant CT_1\sqrt{N}\sum_{j=T_0+1}^{\infty}\sum_{i=0}^{\infty}\rho^{2i+2j-T_0} \\
&= \frac{C\rho^2}{(1-\rho^2)^2}\sqrt{N}\rho^{T_0}T_1 \\
&\leqslant C\sqrt{N},
\end{aligned} \tag{S25}$$

where the last inequality follows from Assumption 4.

Moreover, by choosing

$$\delta_{i,j,\ell} = \exp\left\{-2\rho^{-(i+j+\ell)/2}\check{D}_{\mathcal{M}}/\log(1/\rho)\right\}, \tag{S26}$$

i.e., $\log(1/\delta_{i,j,\ell}) = 2\rho^{-(i+j+\ell)/2}\check{D}_{\mathcal{M}}/\log(1/\rho)$, we can similarly show that

$$\begin{aligned}
\sum_{j=T_0+1}^{\infty}\sum_{i=0}^{\infty}\sum_{\ell=0}^{\infty} b_{i,j,\ell} &\leqslant C\sigma^2\sum_{j=T_0+1}^{\infty}\sum_{i=0}^{\infty}\sum_{\ell=0}^{\infty}\left\{\frac{\rho^{(i+j+\ell)/2}\check{D}_{\mathcal{M}}}{\log(1/\rho)} + \rho^{3(i+j+\ell)/4}\sqrt{\frac{T_1\check{D}_{\mathcal{M}}}{\log(1/\rho)}}\right\} \\
&= C\sigma^2\left\{\frac{\sqrt{\rho}}{(1-\sqrt{\rho})^3}\cdot\frac{\rho^{T_0/2}\check{D}_{\mathcal{M}}}{\log(1/\rho)} + \frac{\rho^{3/4}}{(1-\rho^{3/4})^3}\cdot\sqrt{\frac{\rho^{3T_0/2}T_1\check{D}_{\mathcal{M}}}{\log(1/\rho)}}\right\} \\
&\leqslant C\left(\frac{\check{D}_{\mathcal{M}}}{T_1} + \frac{\sqrt{\check{D}_{\mathcal{M}}}}{T_1}\right) \\
&\leqslant C\frac{\check{D}_{\mathcal{M}}}{T_1},
\end{aligned} \tag{S27}$$

where Assumption 4 is used in the second to last inequality.

In addition, similarly to (S21), we can show that

$$\sum_{j=T_0+1}^{\infty}\sum_{i=0}^{\infty}\sum_{\ell=0}^{\infty}\delta_{i,j,\ell} \leqslant \sum_{j=T_0+1}^{\infty}\sum_{i=0}^{\infty}\sum_{\ell=0}^{\infty}\exp\left\{-(i+j+\ell)\check{D}_{\mathcal{M}}\right\} = \frac{e^{-(T_0+1)\check{D}_{\mathcal{M}}}}{\left(1-e^{-\check{D}_{\mathcal{M}}}\right)^3}. \tag{S28}$$

Combining (S24), (S25), (S27) and (S28), if $\|\boldsymbol{M}\|_{\mathrm{F}} = 1$, for any $1 \leqslant k \leqslant T_0$, we have

$$\mathbb{P}\left\{\frac{1}{T_1}\langle \boldsymbol{M}, \boldsymbol{R}\boldsymbol{X}_k'\rangle \geqslant C\frac{\sqrt{N}}{T_1}\right\} \leqslant \frac{2e^{-(T_0+1)\check{D}_{\mathcal{M}}}}{\left(1-e^{-\check{D}_{\mathcal{M}}}\right)^3} \leqslant Ce^{-\check{D}_{\mathcal{M}}}.$$

Combining this with (S22) and $T_1 \gtrsim \check{D}_{\mathcal{M}}$, we accomplish the proof of this lemma.

S3.4.3 Proof of Lemma B.3

The proof of Lemma B.3 relies on the following result.

Lemma B.6. *Suppose that Assumptions 1 – 3 hold. If $T_1 \gtrsim N + \log T_0$, then*

$$\mathbb{P}\left\{\max_{1\leqslant i\leqslant T_0}\max_{1\leqslant j\leqslant T_0}\left\|\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1} - \boldsymbol{\Gamma}(i-j)\right\|_{\mathrm{op}} \geqslant \tau^2\right\} \leqslant Ce^{-N-\log T_0},$$

where $\tau^2 = C\sqrt{(N+\log T_0)/T_1}$.

Proof. Let $\bar{\mathcal{S}}^{N-1}$ be a minimal 1/4-net of $\mathcal{S}^{N-1}$ in the Euclidean norm. By Vershynin (2010), its cardinality satisfies $|\bar{\mathcal{S}}^{N-1}| \leqslant 9^N$, and for any fixed $1 \leqslant i, j \leqslant T_0$,

$$\left\|\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1} - \boldsymbol{\Gamma}(i-j)\right\|_{\mathrm{op}} \leqslant 2\max_{\boldsymbol{u}\in\bar{\mathcal{S}}^{N-1}}\left|\boldsymbol{u}'\left\{\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1} - \boldsymbol{\Gamma}(i-j)\right\}\boldsymbol{u}\right|.$$

Denote $\widetilde{D}_{\mathcal{M}} = N\log 9 + 2\log T_0$. Hence, for any $K > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left\{\max_{1\leqslant i\leqslant T_0}\max_{1\leqslant j\leqslant T_0}\left\|\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1} - \boldsymbol{\Gamma}(i-j)\right\|_{\mathrm{op}} \geqslant K\right\} \\
&\leqslant T_0^2\max_{1\leqslant i\leqslant T_0}\max_{1\leqslant j\leqslant T_0}\mathbb{P}\left\{\left\|\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1} - \boldsymbol{\Gamma}(i-j)\right\|_{\mathrm{op}} \geqslant K\right\} \\
&\leqslant T_0^2|\bar{\mathcal{S}}^{N-1}|\max_{1\leqslant i\leqslant T_0}\max_{1\leqslant j\leqslant T_0}\max_{\|\boldsymbol{u}\|_2=1}\mathbb{P}\left[\left|\boldsymbol{u}'\left\{\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1} - \boldsymbol{\Gamma}(i-j)\right\}\boldsymbol{u}\right| \geqslant K/2\right] \\
&\leqslant \exp(\widetilde{D}_{\mathcal{M}})\max_{1\leqslant i\leqslant T_0}\max_{1\leqslant j\leqslant T_0}\max_{\|\boldsymbol{u}\|_2=1}\mathbb{P}\left[\left|\boldsymbol{u}'\left\{\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1} - \boldsymbol{\Gamma}(i-j)\right\}\boldsymbol{u}\right| \geqslant K/2\right].
\end{aligned} \tag{S29}$$

Since $\mathbf{y}_t = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j^* \boldsymbol{\Sigma}_{\varepsilon}^{1/2} \boldsymbol{\xi}_{t-j}$, for any $1 \leqslant i, j \leqslant T_0$, we have

$$\frac{\boldsymbol{u}'\boldsymbol{X}_i\boldsymbol{X}_j'\boldsymbol{u}}{T_1} = \sum_{t=T_0+1}^{T} \frac{\boldsymbol{u}'\mathbf{y}_{t-i}\mathbf{y}_{t-j}'\boldsymbol{u}}{T_1} = \frac{1}{T_1}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\sum_{t=T_0+1}^{T} \boldsymbol{u}'\boldsymbol{\Psi}_k^*\boldsymbol{\Sigma}_{\varepsilon}^{1/2}\boldsymbol{\xi}_{t-i-k}\boldsymbol{\xi}_{t-j-\ell}'\boldsymbol{\Sigma}_{\varepsilon}^{1/2}\boldsymbol{\Psi}_\ell^{*\prime}\boldsymbol{u}$$

$$= \frac{1}{T_1}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\sum_{t=T_0+1}^{T} \langle \widetilde{\boldsymbol{M}}_{k,\ell}\boldsymbol{\xi}_{t-j-\ell}, \boldsymbol{\xi}_{t-i-k}\rangle := G_1 + G_2, \qquad \text{(S30)}$$

where

$$G_1 = \frac{1}{T_1}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\sum_{t=T_0+1}^{T} \langle \widetilde{\boldsymbol{M}}_{k,\ell}\boldsymbol{\xi}_{t-j-\ell}, \boldsymbol{\xi}_{t-i-k}\rangle I\{\ell \neq i+k-j\},$$

$$G_2 = \frac{1}{T_1}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\sum_{t=T_0+1}^{T} \langle \widetilde{\boldsymbol{M}}_{k,\ell}\boldsymbol{\xi}_{t-j-\ell}, \boldsymbol{\xi}_{t-i-k}\rangle I\{\ell = i+k-j\},$$

with $I(\cdot)$ being the indicator function, and

$$\widetilde{\boldsymbol{M}}_{k,\ell} = \boldsymbol{\Sigma}_{\varepsilon}^{1/2}\boldsymbol{\Psi}_k^{*\prime}\boldsymbol{u}\boldsymbol{u}'\boldsymbol{\Psi}_\ell^*\boldsymbol{\Sigma}_{\varepsilon}^{1/2}.$$

By Lemma B.7 and a method similar to the proof of Lemma B.2, we can show that

$$\mathbb{P}\left(|G_1| \geqslant \frac{C}{T_1}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} b_{k,\ell} I\{\ell \neq i+k-j\}\right) \leqslant \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} 4\delta_{k,\ell} I\{\ell \neq i+k-j\}$$

and

$$\mathbb{P}\left\{|G_2 - \mathbb{E}(G_2)| \geqslant \frac{C}{T_1}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} b_{k,\ell} I\{\ell = i+k-j\}\right\} \leqslant \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} 2\delta_{k,\ell} I\{\ell = i+k-j\},$$

where $\delta_{k,\ell} \in (0,1)$ will be specified below, and

$$b_{k,\ell} = \sigma^2 \|\widetilde{\boldsymbol{M}}_{k,\ell}\|_{\mathrm{F}} \left\{\log(1/\delta_{k,\ell}) + \sqrt{T_1 \log(1/\delta_{k,\ell})}\right\}.$$

Since $\mathbb{E}(G_1) = 0$, combining the above results with (S30), we have

$$\mathbb{P}\left\{\left|\boldsymbol{u}'\left\{\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1} - \boldsymbol{\Gamma}(i-j)\right\}\boldsymbol{u}\right| \geqslant \frac{C}{T_1}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} b_{k,\ell}\right\}$$

$$\leqslant \mathbb{P}\left\{|G_1| + |G_2 - \mathbb{E}(G_2)| \geqslant \frac{C}{T_1}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} b_{k,\ell}\right\}$$

$$\leqslant 4\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} \delta_{k,\ell}. \qquad \text{(S31)}$$

Note that by Assumption 2, if $\|\boldsymbol{u}\|_2 = 1$, we have

$$\|\widetilde{\boldsymbol{M}}_{k,\ell}\|_{\mathrm{F}} \leqslant \lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon)\|\boldsymbol{\Psi}_k^*\|_{\mathrm{op}}\|\boldsymbol{\Psi}_\ell^*\|_{\mathrm{op}} \leqslant C\rho^{k+\ell}.$$

Then, by choosing

$$\delta_{k,\ell} = \exp\left\{-4\rho^{-(k+\ell+1)/2}\widetilde{D}_{\mathcal{M}}/\log(1/\rho)\right\}.$$

i.e., $\log(1/\delta_{k,\ell}) = 4\rho^{-(k+\ell+1)/2}\widetilde{D}_{\mathcal{M}}/\log(1/\rho)$, we have

$$\begin{aligned}
\frac{1}{T_1}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} b_{k,\ell} &\leqslant \frac{C\sigma^2}{T_1}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\left\{\frac{\rho^{(k+\ell-1)/2}\widetilde{D}_{\mathcal{M}}}{\log(1/\rho)} + \rho^{(3k+3\ell-1)/4}\sqrt{\frac{T_1\widetilde{D}_{\mathcal{M}}}{\log(1/\rho)}}\right\} \\
&= C\sigma^2\left\{\frac{1}{\sqrt{\rho}(1-\sqrt{\rho})^2}\cdot\frac{\widetilde{D}_{\mathcal{M}}}{T_1\log(1/\rho)} + \frac{1}{\rho^{1/4}(1-\rho^{3/4})^2}\cdot\sqrt{\frac{\widetilde{D}_{\mathcal{M}}}{T_1\log(1/\rho)}}\right\} \\
&\leqslant C\sqrt{\frac{\widetilde{D}_{\mathcal{M}}}{T_1}}. \qquad \text{(S32)}
\end{aligned}$$

In addition, similarly to the proof of Lemma B.1, we can show that

$$\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\delta_{k,\ell} \leqslant \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} e^{-2(k+\ell+1)\widetilde{D}_{\widetilde{\mathcal{M}}}} = \frac{e^{-2\widetilde{D}_{\mathcal{M}}}}{(1-e^{-2\widetilde{D}_{\mathcal{M}}})^2}. \qquad \text{(S33)}$$

By (S31)–(S33), if $\|\boldsymbol{u}\|_2 = 1$, for any $1 \leqslant i, j \leqslant T_0$, we have

$$\mathbb{P}\left\{\left|\boldsymbol{u}'\left\{\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1} - \boldsymbol{\Gamma}(i-j)\right\}\boldsymbol{u}\right| \geqslant C\sqrt{\frac{\widetilde{D}_{\mathcal{M}}}{T_1}}\right\} \leqslant \frac{4e^{-2\widetilde{D}_{\mathcal{M}}}}{(1-e^{-2\widetilde{D}_{\mathcal{M}}})^2} \leqslant Ce^{-\widetilde{D}_{\mathcal{M}}}.$$

Combining this with (S29) and $T_1 \gtrsim \widetilde{D}_{\mathcal{M}}$, we accomplish the proof of this lemma. □

Now we are ready to prove Lemma B.3.

Proof of Lemma B.3. By Basu and Michailidis (2015), we have $\sigma_{\min}(\boldsymbol{\Sigma}_{T_0}) \geqslant \lambda_{\min}(\boldsymbol{\Sigma}_\varepsilon)\mu_{\min}(\boldsymbol{\Psi}_*) = \kappa_{\mathrm{RSC}}$, and hence

$$\frac{\mathbb{E}\left(\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2\right)}{T_1} = \mathrm{tr}\left(\boldsymbol{\Delta}_{(1)}\boldsymbol{\Sigma}_{T_0}\boldsymbol{\Delta}_{(1)}'\right) \geqslant \sigma_{\min}(\boldsymbol{\Sigma}_{T_0})\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2 \geqslant \kappa_{\mathrm{RSC}}\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2. \qquad \text{(S34)}$$

Moreover, observe that

$$
\begin{aligned}
\frac{\left|\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2-\mathbb{E}\left(\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2\right)\right|}{T_1} &= \operatorname{tr}\left\{\boldsymbol{\Delta}_{(1)}\left(\frac{\boldsymbol{X}\boldsymbol{X}'}{T_1}-\boldsymbol{\Sigma}_{T_0}\right)\boldsymbol{\Delta}_{(1)}'\right\} \\
&= \sum_{i=1}^{T_0}\sum_{j=1}^{T_0}\operatorname{tr}\left[\boldsymbol{\Delta}_i\left\{\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1}-\boldsymbol{\Gamma}(i-j)\right\}\boldsymbol{\Delta}_j'\right] \\
&\leqslant \sum_{i=1}^{T_0}\sum_{j=1}^{T_0}\|\boldsymbol{\Delta}_i\|_{\mathrm{F}}\|\boldsymbol{\Delta}_j\|_{\mathrm{F}}\left\|\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1}-\boldsymbol{\Gamma}(i-j)\right\|_{\mathrm{op}} \\
&\leqslant \|\boldsymbol{\Delta}\|_{\ddagger}^2\max_{1\leqslant i\leqslant T_0}\max_{1\leqslant j\leqslant T_0}\left\|\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1}-\boldsymbol{\Gamma}(i-j)\right\|_{\mathrm{op}}.
\end{aligned}
$$

As a result, for any $\boldsymbol{\Delta}\in\mathbb{R}^{N\times N\times T_0}$, we have

$$
\begin{aligned}
\frac{1}{T_1}\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2 &\geqslant \frac{\mathbb{E}\left(\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2\right)}{T_1}-\frac{\left|\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2-\mathbb{E}\left(\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2\right)\right|}{T_1} \\
&\geqslant \kappa_{\mathrm{RSC}}\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2-\|\boldsymbol{\Delta}\|_{\ddagger}^2\max_{1\leqslant i\leqslant T_0}\max_{1\leqslant j\leqslant T_0}\left\|\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1}-\boldsymbol{\Gamma}(i-j)\right\|_{\mathrm{op}}.
\end{aligned}
$$

Combining this with Lemma B.6, we accomplish the proof of Lemma B.3. □

S3.4.4 Proof of Lemma B.4

Since $\boldsymbol{r}_t=\sum_{j=T_0+1}^{\infty}\boldsymbol{A}_j^*\mathbf{y}_{t-j}=\sum_{j=T_0+1}^{\infty}\sum_{\ell=0}^{\infty}\boldsymbol{A}_j^*\boldsymbol{\Psi}_\ell^*\boldsymbol{\Sigma}_\varepsilon^{1/2}\boldsymbol{\xi}_{t-j-\ell}$, we have

$$
\begin{aligned}
\|\boldsymbol{R}\|_{\mathrm{F}}^2=\sum_{t=T_0+1}^{T}\langle\boldsymbol{r}_t,\boldsymbol{r}_t\rangle &= \sum_{t=T_0+1}^{T}\sum_{i=T_0+1}^{\infty}\sum_{j=T_0+1}^{\infty}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\left\langle\boldsymbol{A}_j^*\boldsymbol{\Psi}_\ell^*\boldsymbol{\Sigma}_\varepsilon^{1/2}\boldsymbol{\xi}_{t-j-\ell},\boldsymbol{A}_i^*\boldsymbol{\Psi}_k^*\boldsymbol{\Sigma}_\varepsilon^{1/2}\boldsymbol{\xi}_{t-i-k}\right\rangle \\
&= \sum_{i=T_0+1}^{\infty}\sum_{j=T_0+1}^{\infty}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\sum_{t=T_0+1}^{T}\langle\widetilde{\boldsymbol{M}}_{i,j,k,\ell}\boldsymbol{\xi}_{t-i-k},\boldsymbol{\xi}_{t-j-\ell}\rangle,
\end{aligned}
$$

and

$$
\widetilde{\boldsymbol{M}}_{i,j,k,\ell}=\boldsymbol{\Sigma}_\varepsilon^{1/2}\boldsymbol{\Psi}_\ell^{*\prime}\boldsymbol{A}_j^{*\prime}\boldsymbol{A}_i^*\boldsymbol{\Psi}_k^*\boldsymbol{\Sigma}_\varepsilon^{1/2}.
$$

Then, by Lemma B.7 and a method similar to the proof of Lemma B.2, we can show that

$$
\begin{aligned}
&\mathbb{P}\left\{\|\boldsymbol{R}\|_{\mathrm{F}}^2\geqslant C\sum_{i=T_0+1}^{\infty}\sum_{j=T_0+1}^{\infty}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}b_{i,j,k,\ell}+\sum_{i=T_0+1}^{\infty}\sum_{j=T_0+1}^{\infty}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}c_{i,j,k,\ell}I\{\ell=i+k-j\}\right\} \\
&\quad\leqslant 2\sum_{i=T_0+1}^{\infty}\sum_{j=T_0+1}^{\infty}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\delta_{i,j,k,\ell}, \qquad \text{(S35)}
\end{aligned}
$$

where $I(\cdot)$ is the indicator function, $\delta_{i,j,k,\ell} \in (0,1)$ will be specified shortly,

$$b_{i,j,k,\ell} = \sigma^2 \|\widetilde{\boldsymbol{M}}_{i,j,k,\ell}\|_{\mathrm{F}} \left\{ \log(1/\delta_{i,j,k,\ell}) + \sqrt{T_1 \log(1/\delta_{i,j,k,\ell})} \right\}$$

and

$$c_{i,j,k,\ell} = T_1 \sqrt{N} \|\widetilde{\boldsymbol{M}}_{i,j,k,\ell}\|_{\mathrm{F}}.$$

By Assumption 2 and 3,

$$\|\widetilde{\boldsymbol{M}}_{i,j,k,\ell}\|_{\mathrm{F}} \leqslant \lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon) \|\boldsymbol{\Psi}_\ell^*\|_{\mathrm{op}} \|\boldsymbol{A}_j^*\|_{\mathrm{F}} \|\boldsymbol{A}_i^*\|_{\mathrm{op}} \|\boldsymbol{\Psi}_k^*\|_{\mathrm{op}} \leqslant C\rho^{i+k+\ell} \|\boldsymbol{A}_j^*\|_{\mathrm{F}}.$$

Then, since $c_{i,j,k,\ell} \geqslant 0$, we have

$$\begin{aligned}
\sum_{i=T_0+1}^{\infty} \sum_{j=T_0+1}^{\infty} \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} c_{i,j,k,\ell} I\{\ell = i+k-j\} &\leqslant \sum_{i=T_0+1}^{\infty} \sum_{j=T_0+1}^{\infty} \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} c_{i,j,k,\ell} \\
&\leqslant C T_1 \sqrt{N} \sum_{i=T_0+1}^{\infty} \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \rho^{i+k+\ell} \sum_{j=T_0+1}^{\infty} \|\boldsymbol{A}_j^*\|_{\mathrm{F}} \\
&= \frac{C\rho}{(1-\rho)^3} \sqrt{N} \rho^{T_0} T_1 \sum_{j=T_0+1}^{\infty} \left\|\boldsymbol{A}_j^*\right\|_{\mathrm{F}} \\
&\leqslant \frac{C\rho^2}{(1-\rho)^4} \lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon) N \rho^{2T_0} T_1 \\
&\leqslant C N \rho^{3T_0/2}, \qquad \text{(S36)}
\end{aligned}$$

where the last but two inequality follows from Assumption 2 and the last inequality follows from Assumption 4. Moreover, by choosing

$$\delta_{i,j,k,\ell} = \exp\left\{-2\rho^{-(i+k+\ell+j)/2} N / \log(1/\rho)\right\},$$

i.e., $\log(1/\delta_{i,j,k,\ell}) = 2\sqrt{j}\rho^{-(i+k+\ell)/2}N/\log(1/\rho)$, we can show that

$$\begin{aligned}
&\sum_{i=T_0+1}^{\infty}\sum_{j=T_0+1}^{\infty}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} b_{i,j,k,\ell} \\
&\leqslant C\sigma^2\sqrt{N}\sum_{i=T_0+1}^{\infty}\sum_{j=T_0+1}^{\infty}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\left\{\frac{\rho^{(i+k+\ell+j)/2}N}{\log(1/\rho)} + \rho^{3(i+k+\ell+j)/4}\sqrt{\frac{T_1 N}{\log(1/\rho)}}\right\} \\
&= C\sigma^2\sqrt{N}\left\{\frac{\rho}{(1-\sqrt{\rho})^4}\cdot\frac{\rho^{T_0}N}{\log(1/\rho)} + \frac{\rho^{3/2}}{(1-\rho^{3/4})^2}\cdot\sqrt{\frac{\rho^{3T_0}T_1 N}{\log(1/\rho)}}\right\} \\
&\leqslant C\sigma^2\sqrt{N}\left(\rho^{T_0}N + \sqrt{\rho^{5T_0/2}N}\right) \\
&\leqslant CN^{3/2}\rho^{T_0},
\end{aligned} \tag{S37}$$

where we used Assumption 2 in the first inequality and Assumption 4 in the second to last inequality.

In addition, similarly to (S21), we can show that

$$\begin{aligned}
\sum_{i=T_0+1}^{\infty}\sum_{j=T_0+1}^{\infty}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\delta_{i,j,k,\ell} &\leqslant \sum_{i=T_0+1}^{\infty}\sum_{j=T_0+1}^{\infty}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\exp\{-(i+k+\ell+j)N\} \\
&\leqslant \frac{e^{-(T_0+1)^2N}}{(1-e^{-N})^4} \leqslant \frac{e^{-T_0^2N}}{(1-e^{-N})^4} \leqslant Ce^{-N}.
\end{aligned} \tag{S38}$$

Combining (S35)–(S38), the proof of this lemma is complete.

S3.5 Two auxiliary lemmas

Below we present two auxiliary lemmas. Lemma B.7 is used in the proofs of Lemmas B.1, B.2, B.4, and B.6, while Lemma B.8 establishes the basic martingale concentration bound which is used to prove Lemmas B.7 and B.3.

Lemma B.7. *Let $T_0 < T$ be arbitrary fixed time points, and $T_1 = T - T_0$. Suppose that $\boldsymbol{\xi}_t$ is a random vector with independent, zero-mean, unit-variance, σ^2-sub-Gaussian coordinates for any $T_0 + 1 \leqslant t \leqslant T$. Then*

(i) for any $\delta \in (0,1]$, $\boldsymbol{M} \in \mathbb{R}^{N\times N}$, and fixed nonzero integer $1 \leqslant s \leqslant T_0$,

$$\mathbb{P}\left\{\sum_{t=T_0+1}^{T}\langle \boldsymbol{M}\boldsymbol{\xi}_{t-s}, \boldsymbol{\xi}_t\rangle \geqslant C\sigma^2\|\boldsymbol{M}\|_{\mathrm{F}}\left\{\log(1/\delta) + \sqrt{T_1\log(1/\delta)}\right\}\right\} \leqslant 2\delta \tag{S39}$$

and

$$\mathbb{P}\left\{\left|\sum_{t=T_0+1}^{T}\langle \boldsymbol{M}\boldsymbol{\xi}_{t-s}, \boldsymbol{\xi}_t\rangle\right| \geqslant C\sigma^2\|\boldsymbol{M}\|_{\mathrm{F}}\left\{\log(1/\delta) + \sqrt{T_1\log(1/\delta)}\right\}\right\} \leqslant 4\delta; \tag{S40}$$

(ii) for any $\delta \in (0,1]$ *and* $\boldsymbol{M} \in \mathbb{R}^{N\times N}$,

$$\mathbb{P}\left\{\sum_{t=T_0+1}^{T}\langle \boldsymbol{M}\boldsymbol{\xi}_{t}, \boldsymbol{\xi}_t\rangle \geqslant C\sigma^2\|\boldsymbol{M}\|_{\mathrm{F}}\left\{\log(1/\delta) + \sqrt{T_1\log(1/\delta)}\right\} + T_1\sqrt{N}\|\boldsymbol{M}\|_{\mathrm{F}}\right\} \leqslant \delta \tag{S41}$$

and

$$\mathbb{P}\left\{\left|\sum_{t=T_0+1}^{T}\langle \boldsymbol{M}\boldsymbol{\xi}_{t}, \boldsymbol{\xi}_t\rangle - E\left(\sum_{t=T_0+1}^{T}\langle \boldsymbol{M}\boldsymbol{\xi}_{t}, \boldsymbol{\xi}_t\rangle\right)\right| \geqslant C\sigma^2\|\boldsymbol{M}\|_{\mathrm{F}}\left\{\log(1/\delta) + \sqrt{T_1\log(1/\delta)}\right\}\right\}$$
$$\leqslant 2\delta. \tag{S42}$$

Proof. For any integer s, denote $\boldsymbol{\xi}_{[s]} = (\boldsymbol{\xi}'_{T-s}, \boldsymbol{\xi}'_{T-1-s}, \ldots, \boldsymbol{\xi}'_{T_0+1-s})' \in \mathbb{R}^{T_1 N}$, $S_{\xi,[s]} = \sum_{t=T_0+1}^{T}\langle \boldsymbol{M}\boldsymbol{\xi}_{t-s}, \boldsymbol{\xi}_t\rangle$, and $V_{\xi,[s]} = \sum_{t=T_0+1}^{T}\|\boldsymbol{M}\boldsymbol{\xi}_{t-s}\|^2$. Note that $\boldsymbol{\xi}_{[s]}$ is a random vector with independent, zero-mean, unit-variance, σ^2-sub-Gaussian coordinates for any fixed integer s. For simplicity, in the following we omit the subscript $[s]$ in $\boldsymbol{\xi}_{[s]}$, $S_{\xi,[s]}$, and $V_{\xi,[s]}$ whenever $s \neq 0$.

We first prove claim (i) of this lemma. Without loss of generality, assume that s is a positive integer. Note that $V_\xi = \|(\boldsymbol{I}_{T_1}\otimes \boldsymbol{M})\boldsymbol{\xi}\|^2 = \boldsymbol{\xi}'\boldsymbol{Q}\boldsymbol{\xi}$, where $\boldsymbol{Q} = \boldsymbol{I}_{T_1}\otimes \boldsymbol{M}'\boldsymbol{M}$. Then $\mathbb{E}(V_\xi) = \|\boldsymbol{I}_{T_1}\otimes \boldsymbol{M}\|_{\mathrm{F}}^2 = T_1\|\boldsymbol{M}\|_{\mathrm{F}}^2$. By the (one-sided) Hanson-Wright inequality (Vershynin, 2010), for any $\eta \geqslant 0$,

$$\mathbb{P}\left(V_\xi - T_1\|\boldsymbol{M}\|_{\mathrm{F}}^2 \geqslant \eta\right) \leqslant \exp\left\{-c\min\left(\frac{\eta}{\sigma^2\|\boldsymbol{Q}\|_{\mathrm{op}}}, \frac{\eta^2}{\sigma^4\|\boldsymbol{Q}\|_{\mathrm{F}}^2}\right)\right\},$$

where $c > 0$ is an absolute constant. Since $\|\boldsymbol{Q}\|_{\mathrm{op}} = \|\boldsymbol{M}\|_{\mathrm{op}}^2 \leqslant \|\boldsymbol{M}\|_{\mathrm{F}}^2$, and $\|\boldsymbol{Q}\|_{\mathrm{F}} \leqslant \|\boldsymbol{I}_{T_1}\otimes \boldsymbol{M}\|_{\mathrm{F}}\|\boldsymbol{I}_{T_1}\otimes \boldsymbol{M}\|_{\mathrm{op}} \leqslant \sqrt{T_1}\|\boldsymbol{M}\|_{\mathrm{F}}^2$, we can show that

$$\mathbb{P}\left(V_\xi \geqslant T_1\|\boldsymbol{M}\|_{\mathrm{F}}^2 + \eta\right) \leqslant \exp\left\{-c\min\left(\frac{\eta}{\sigma^2\|\boldsymbol{M}\|_{\mathrm{F}}^2}, \frac{\eta^2}{\sigma^4 T_1\|\boldsymbol{M}\|_{\mathrm{F}}^4}\right)\right\} \leqslant \delta.$$

by choosing

$$\eta = C\sigma^2\|\boldsymbol{M}\|_{\mathrm{F}}^2\left\{\log(1/\delta) + \sqrt{T_1\log(1/\delta)}\right\}, \tag{S43}$$

where C is dependent on c. Moreover, by Lemma B.8, for any $\alpha, \beta > 0$, we have

$$\mathbb{P}(S_\xi \geqslant \alpha) \leqslant \mathbb{P}(S_\xi \geqslant \alpha,\ V_\xi \leqslant \beta) + \mathbb{P}(V_\xi \geqslant \beta) \leqslant \exp\left(-\frac{\alpha^2}{2\sigma^2\beta}\right) + \mathbb{P}(V_\xi \geqslant \beta).$$

This implies that, if $\beta = T_1\|\boldsymbol{M}\|_{\mathrm{F}}^2 + \eta$ and $\alpha \geqslant \sqrt{2\sigma^2\beta\log(1/\delta)}$, then $\mathbb{P}(S_\xi \geqslant \alpha) \leqslant \delta$. Hence, we can establish (S39) by choosing

$$\alpha = C\sigma^2\|\boldsymbol{M}\|_{\mathrm{F}}\left\{\log(1/\delta) + \sqrt{T_1\log(1/\delta)}\right\}.$$

Furthermore, applying (S39) to $-\boldsymbol{M}$, we directly have

$$\mathbb{P}\left\{\sum_{t=T_0+1}^{T}\langle \boldsymbol{M}\boldsymbol{\xi}_{t-s}, \boldsymbol{\xi}_t\rangle \leqslant -C\sigma^2\|\boldsymbol{M}\|_{\mathrm{F}}\left\{\log(1/\delta) + \sqrt{T_1\log(1/\delta)}\right\}\right\} \leqslant 2\delta,$$

which, combined with (S39), yields the two-sided bound in (S40).

The proof of claim (ii) is similar to the analysis of V_ξ above. Note that $S_{\xi,[0]} = \boldsymbol{\xi}_{[0]}'(\boldsymbol{I}_{T_1}\otimes\boldsymbol{M})'\boldsymbol{\xi}_{[0]}$. Applying the (two-sided) Hanson-Wright inequality, we have

$$\mathbb{P}\left\{\left|S_{\xi,[0]} - \mathbb{E}(S_{\xi,[0]})\right| \geqslant \eta_0\right\} \leqslant 2\exp\left\{-c\min\left(\frac{\eta_0}{\sigma^2\|\boldsymbol{M}\|_{\mathrm{op}}}, \frac{\eta_0^2}{\sigma^4 T_1\|\boldsymbol{M}\|_{\mathrm{F}}^2}\right)\right\} \leqslant 2\delta$$

if we choose

$$\eta_0 = C\sigma^2\|\boldsymbol{M}\|_{\mathrm{F}}\left\{\log(1/\delta) + \sqrt{T_1\log(1/\delta)}\right\}, \tag{S44}$$

where C is dependent on c. This leads to (S42) in the lemma. Moreover, since $\mathbb{E}(S_{\xi,[0]}) = T_1\operatorname{tr}(\boldsymbol{M}) \leqslant T_1\sqrt{N}\|\boldsymbol{M}\|_{\mathrm{F}}$, similarly we can also obtain the one-sided result:

$$\mathbb{P}\left\{S_{\xi,[0]} \geqslant T_1\operatorname{tr}(\boldsymbol{M}) + \eta_0\right\} \leqslant \exp\left\{-c\min\left(\frac{\eta_0}{\sigma^2\|\boldsymbol{M}\|_{\mathrm{op}}}, \frac{\eta_0^2}{\sigma^4 T_1\|\boldsymbol{M}\|_{\mathrm{F}}^2}\right)\right\} \leqslant \delta$$

with the same choice of η_0 as in (S44). Therefore, (S42) is proved as well. □

Lemma B.8 (Martingale concentration)**.** *Let $\{\mathcal{F}_t, t \in \mathbb{Z}\}$ be a filtration. Suppose that $\{\boldsymbol{w}_t\}$ and $\{\boldsymbol{e}_t\}$ are processes taking values in $\mathbb{R}^d$, and for each integer t, $\boldsymbol{w}_t$ is $\mathcal{F}_{t-1}$-measurable, $\boldsymbol{e}_t$ is $\mathcal{F}_t$-measurable, and $\boldsymbol{e}_t \mid \mathcal{F}_{t-1}$ is mean-zero and σ^2-sub-Gaussian. Let $T_0 < T$ be arbitrary fixed time points. Then, for any $\alpha, \beta > 0$, we have*

$$\mathbb{P}\left\{\sum_{t=T_0+1}^{T}\langle \boldsymbol{w}_t, \boldsymbol{e}_t\rangle \geqslant \alpha,\ \sum_{t=T_0+1}^{T}\|\boldsymbol{w}_t\|^2 \leqslant \beta\right\} \leqslant \exp\left(-\frac{\alpha^2}{2\sigma^2\beta}\right).$$

Proof. See Lemma 4.2 in Simchowitz et al. (2018). □

S4 Proofs of theoretical results in Section 4.2

This section gives the proofs of Theorems 3, 4 and Corollary 2 in Sections C.1-C.3, respectively. Section C.4 provides five auxiliary lemmas, which are used in the proof of Theorem 4. We first introduce several parameter spaces of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{N\times N\times T_0}$ below,

$$\begin{aligned}
\boldsymbol{\Theta}(r_1,r_2) &= \{\boldsymbol{\mathcal{A}} \in \mathbb{R}^{N\times N\times T_0} \mid \mathrm{rank}(\boldsymbol{\mathcal{A}}_{(1)}) \leqslant r_1,\ \mathrm{rank}(\boldsymbol{\mathcal{A}}_{(2)}) \leqslant r_2\},\\
\boldsymbol{\Theta}^{\mathrm{SP}}(r_1,r_2,s) &= \{\boldsymbol{\mathcal{A}} \in \mathbb{R}^{N\times N\times T_0} \mid \mathrm{rank}(\boldsymbol{\mathcal{A}}_{(1)}) \leqslant r_1,\ \mathrm{rank}(\boldsymbol{\mathcal{A}}_{(2)}) \leqslant r_2,\ \|\boldsymbol{\mathcal{A}}\|_0 \leqslant s,\},\\
\boldsymbol{\Theta}_{\ddagger}(r_1,r_2) &= \{\boldsymbol{\mathcal{A}} \in \boldsymbol{\Theta}(r_1,r_2),\ \|\boldsymbol{\mathcal{A}}\|_{\ddagger}=1\} \quad \text{and} \quad \boldsymbol{\Theta}_1^{\mathrm{SP}}(r_1,r_2,s) = \{\boldsymbol{\mathcal{A}} \in \boldsymbol{\Theta}^{\mathrm{SP}}(r_1,r_2,s),\ \|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}=1\}.
\end{aligned}$$

S4.1 Proof of Theorem 3

It can be verified that $[\nabla\mathcal{L}(\boldsymbol{\mathcal{A}})]_{(1)} = -T_1^{-1}(\boldsymbol{Y}-\boldsymbol{\mathcal{A}}_{(1)}\boldsymbol{X})\boldsymbol{X}'$ and $\boldsymbol{Y} = (\boldsymbol{\mathcal{A}}^*_{S_\gamma})_{(1)}\boldsymbol{X} + (\boldsymbol{\mathcal{A}}^*_{S_\gamma^c})_{(1)}\boldsymbol{X} + \widetilde{\boldsymbol{\mathcal{E}}}$. As a result, for any $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{N\times N\times T_0}$,

$$\langle \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*_{S_\gamma}), \boldsymbol{\mathcal{M}}\rangle = -\langle {T_1}^{-1}\widetilde{\boldsymbol{\mathcal{E}}}\boldsymbol{X}', \boldsymbol{\mathcal{M}}_{(1)}\rangle - \langle {T_1}^{-1}(\boldsymbol{\mathcal{A}}^*_{S_\gamma^c})_{(1)}\boldsymbol{X}\boldsymbol{X}', \boldsymbol{\mathcal{M}}_{(1)}\rangle. \tag{S1}$$

From Lemmas B.1 and B.2 and by a method similar to (S8), we can show that, if $T_1 \gtrsim (r_1 \wedge r_2)N + \log T_0$ and $\gamma \gtrsim \sqrt{\{(r_1\wedge r_2)N + \log T_0\}/T_1}$, then

$$\mathbb{P}\left\{\sup_{\boldsymbol{\Delta}\in\boldsymbol{\Theta}_{\ddagger}(r_1,r_2)} \langle \frac{1}{T_1}\widetilde{\boldsymbol{\mathcal{E}}}\boldsymbol{X}', \boldsymbol{\Delta}_{(1)}\rangle \geqslant C\gamma\right\} \leqslant Ce^{-(r_1\wedge r_2)N - \log T_0},$$

which, together with the fact that $\|\boldsymbol{\mathcal{M}}\|_{\ddagger} \leqslant \sqrt{s}\|\boldsymbol{\mathcal{M}}\|_{\mathrm{F}} = \sqrt{s}$ for any $\boldsymbol{\mathcal{M}} \in \boldsymbol{\Theta}_1^{\mathrm{SP}}(r_1,r_2,s)$, implies that

$$\langle \frac{1}{T_1}\widetilde{\boldsymbol{\mathcal{E}}}\boldsymbol{X}', \boldsymbol{\mathcal{M}}_{(1)}\rangle \leqslant \|\boldsymbol{\mathcal{M}}\|_{\ddagger} \sup_{\boldsymbol{\Delta}\in\boldsymbol{\Theta}_{\ddagger}(r_1,r_2)} \langle \frac{1}{T_1}\widetilde{\boldsymbol{\mathcal{E}}}\boldsymbol{X}', \boldsymbol{\Delta}_{(1)}\rangle \leqslant C\gamma\sqrt{s} \tag{S2}$$

holds with probability at least $1 - Ce^{-(r_1\wedge r_2)N - \log T_0}$.

We next handle the second term at the right hand side of (S1). It holds that, from Assumptions 2 and 3,

$$\begin{aligned}
\frac{\mathbb{E}\langle (\boldsymbol{\mathcal{A}}^*_{S_\gamma^c})_{(1)}\boldsymbol{X}\boldsymbol{X}', \boldsymbol{\mathcal{M}}_{(1)}\rangle}{T_1} &= \mathrm{tr}\left((\boldsymbol{\mathcal{A}}^*_{S_\gamma^c})_{(1)}\boldsymbol{\Sigma}_{T_0}\boldsymbol{\mathcal{M}}'_{(1)}\right)\\
&\leqslant \lambda_{\max}(\boldsymbol{\Sigma}_{T_0})\|\boldsymbol{\mathcal{A}}^*_{S_\gamma^c}\|_{\mathrm{F}}\|\boldsymbol{\mathcal{M}}\|_{\mathrm{F}} \leqslant C^2\kappa_{\mathrm{RSS}}\|\boldsymbol{\mathcal{A}}^*_{S_\gamma^c}\|_{\mathrm{F}}\|\boldsymbol{\mathcal{M}}\|_{\mathrm{F}},
\end{aligned}$$

and

$$\begin{aligned}&\frac{|\langle(\boldsymbol{\mathcal{A}}^*_{S^c_\gamma})_{(1)}\boldsymbol{X}\boldsymbol{X}',\boldsymbol{\mathcal{M}}_{(1)}\rangle-\mathbb{E}\langle(\boldsymbol{\mathcal{A}}^*_{S^c_\gamma})_{(1)}\boldsymbol{X}\boldsymbol{X}',\boldsymbol{\mathcal{M}}_{(1)}\rangle|}{T_1}\\&\qquad\leqslant\sum_{i=1}^{T_0}\sum_{j=1}^{T_0}\left|\operatorname{tr}\left[\boldsymbol{A}_i^*\left\{\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1}-\boldsymbol{\Gamma}(i-j)\right\}\boldsymbol{M}_j\right]I\{i\in S^c_\gamma\}\right|\\&\qquad\leqslant\|\boldsymbol{\mathcal{A}}^*_{S^c_\gamma}\|_\ddagger\|\boldsymbol{\mathcal{M}}\|_\ddagger\max_{1\leqslant i\leqslant T_0}\max_{1\leqslant j\leqslant T_0}\left\|\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1}-\boldsymbol{\Gamma}(i-j)\right\|_{\text{op}}.\end{aligned}$$

As a result, by Lemma B.6 and the fact that $\|\boldsymbol{\mathcal{M}}\|_\ddagger\leqslant\sqrt{s}\|\boldsymbol{\mathcal{M}}\|_{\text{F}}$,

$$\begin{aligned}&{T_1}^{-1}\langle(\boldsymbol{\mathcal{A}}^*_{S^c_\gamma})_{(1)}\boldsymbol{X}\boldsymbol{X}',\boldsymbol{\mathcal{M}}_{(1)}\rangle\\&\qquad\leqslant\frac{\mathbb{E}\langle(\boldsymbol{\mathcal{A}}^*_{S^c_\gamma})_{(1)}\boldsymbol{X}\boldsymbol{X}',\boldsymbol{\mathcal{M}}_{(1)}\rangle}{T_1}+\frac{|\langle(\boldsymbol{\mathcal{A}}^*_{S^c_\gamma})_{(1)}\boldsymbol{X}\boldsymbol{X}',\boldsymbol{\mathcal{M}}_{(1)}\rangle-\mathbb{E}\langle(\boldsymbol{\mathcal{A}}^*_{S^c_\gamma})_{(1)}\boldsymbol{X}\boldsymbol{X}',\boldsymbol{\mathcal{M}}_{(1)}\rangle|}{T_1}\\&\qquad\leqslant\left(C^2\kappa_{\text{RSS}}\|\boldsymbol{\mathcal{A}}^*_{S^c_\gamma}\|_{\text{F}}+\tau^2\sqrt{s}\|\boldsymbol{\mathcal{A}}^*_{S^c_\gamma}\|_\ddagger\right)\|\boldsymbol{\mathcal{M}}\|_{\text{F}}.\end{aligned}\tag{S3}$$

holds with probability at least $1-Ce^{-N-\log T_0}$ when $T_1\gtrsim s^2(N+\log T_0)$. We accomplish the proof by combining (S1) – (S3) and letting $\|\boldsymbol{\mathcal{M}}\|_{\text{F}}=1$.

S4.2 Proof of Theorem 4

This proof is divided into six steps. Some notations and conditions are given in the first step, and verified in the last step. Without loss of generality, we assume that $0<\sigma_L<1<\sigma_U$ and $0<\kappa_{\text{RSC}}<1<\kappa_{\text{RSS}}$ throughout this proof.

Step 1 (Notations and conditions) Without confusion, we use $\boldsymbol{\mathcal{A}}^*$ to denote $\boldsymbol{\mathcal{A}}^*_{S_\gamma}$ for simplicity in this proof. Since $\boldsymbol{\mathcal{A}}^*$ has Tucker ranks r_1 and r_2 along the first two modes, its Tucker decomposition can be assumed to have the form of $\boldsymbol{\mathcal{A}}^*=\boldsymbol{\mathcal{G}}^*\times_1\boldsymbol{U}_1^*\times_2\boldsymbol{U}_2^*$, where $\boldsymbol{\mathcal{G}}^*\in\mathbb{R}^{r_1\times r_2\times T_0}$, $\boldsymbol{U}_i^*\in\mathbb{R}^{N\times r_i}$ and $\boldsymbol{U}_i^{*\prime}\boldsymbol{U}_i^*=b^2\boldsymbol{I}_{r_i}$ for $1\leqslant i\leqslant 2$. Moreover, at the k-th iteration of Algorithm 1, the pre-thresholding estimator $\widetilde{\boldsymbol{\mathcal{A}}}^{k+1}$ has the Tucker form of $\widetilde{\boldsymbol{\mathcal{G}}}^{k+1}\times_1\boldsymbol{U}_1^{k+1}\times_2\boldsymbol{U}_2^{k+1}$, where $\widetilde{\boldsymbol{\mathcal{G}}}^{k+1}$,$\boldsymbol{U}_1^{k+1}$ and $\boldsymbol{U}_2^{k+1}$ are obtained by one-step gradient descent, and the hard-thresholding operation gives

$$\boldsymbol{\mathcal{A}}^{k+1}=\text{HT}(\widetilde{\boldsymbol{\mathcal{A}}}^{k+1},\text{s})=\boldsymbol{\mathcal{G}}^{k+1}\times_1\boldsymbol{U}_1^{k+1}\times_2\boldsymbol{U}_2^{k+1}.$$

Note that the zero frontal slices in $\boldsymbol{\mathcal{A}}^*$, $\widetilde{\boldsymbol{\mathcal{A}}}^{k+1}$ and $\boldsymbol{\mathcal{A}}^{k+1}$ correspond to the zero ones in $\boldsymbol{\mathcal{G}}^*$, $\widetilde{\boldsymbol{\mathcal{G}}}^{k+1}$ and $\boldsymbol{\mathcal{G}}^{k+1}$, respectively.

On the other hand, we denote by $\breve{S}$ the union set $\breve{S}_{k+1,\gamma} = S_k \cup S_{k+1} \cup S_\gamma$, where the subscripts of k and γ are suppressed when there is no confusion, and it holds that $|\breve{S}| \leqslant 3s$ since $s \geqslant s_\gamma$. Moreover, it can be verified that $\boldsymbol{\mathcal{A}}_{\breve{S}}^k = \boldsymbol{\mathcal{A}}^k$, $\boldsymbol{\mathcal{A}}_{\breve{S}}^{k+1} = \boldsymbol{\mathcal{A}}^{k+1}$ and

$$\boldsymbol{\mathcal{A}}^{k+1} = \mathrm{HT}(\widetilde{\boldsymbol{\mathcal{A}}}_{\breve{S}}^{\mathrm{k+1}}, \mathrm{s}) \text{ with } \widetilde{\boldsymbol{\mathcal{A}}}_{\breve{S}}^{k+1} = \widetilde{\boldsymbol{\mathcal{G}}}_{\breve{S}}^{k+1} \times_1 \boldsymbol{U}_1^{k+1} \times_2 \boldsymbol{U}_2^{k+1}.$$

The distances between $\widetilde{\boldsymbol{\mathcal{A}}}_{\breve{S}}^{k+1}$, $\boldsymbol{\mathcal{A}}^{k+1}$ and $\boldsymbol{\mathcal{A}}^*$ can be measured by

$$\widetilde{E}^{k+1} = \min_{\boldsymbol{R}_i \in \mathcal{O}^{r_i \times r_i}, 1 \leqslant i \leqslant 2} \sum_{i=1}^{2} \|\boldsymbol{U}_i^{k+1} - \boldsymbol{U}_i^* \boldsymbol{R}_i\|_{\mathrm{F}}^2 + \|\widetilde{\boldsymbol{\mathcal{G}}}_{\breve{S}}^{k+1} - \boldsymbol{\mathcal{G}}^* \times_1 \boldsymbol{R}_1' \times_2 \boldsymbol{R}_2'\|_{\mathrm{F}}^2,$$

$$E^{k+1} = \min_{\boldsymbol{R}_i \in \mathcal{O}^{r_i \times r_i}, 1 \leqslant i \leqslant 2} \sum_{i=1}^{2} \|\boldsymbol{U}_i^{k+1} - \boldsymbol{U}_i^* \boldsymbol{R}_i\|_{\mathrm{F}}^2 + \|\boldsymbol{\mathcal{G}}^{k+1} - \boldsymbol{\mathcal{G}}^* \times_1 \boldsymbol{R}_1' \times_2 \boldsymbol{R}_2'\|_{\mathrm{F}}^2,$$

respectively, and their optimizers are denoted by $(\widetilde{\boldsymbol{R}}_1^{k+1}, \widetilde{\boldsymbol{R}}_2^{k+1})$ and $(\boldsymbol{R}_1^{k+1}, \boldsymbol{R}_2^{k+1})$. We next introduce or rephrase the following list of conditions:

- Suppose that there exist $\alpha, \beta > 0$ such that

$$\langle \nabla \mathcal{L}(\boldsymbol{\mathcal{A}}) - \nabla \mathcal{L}(\boldsymbol{\mathcal{A}}^*), \boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}^* \rangle \geqslant \alpha \|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 + \beta \|\nabla \mathcal{L}(\boldsymbol{\mathcal{A}}) - \nabla \mathcal{L}(\boldsymbol{\mathcal{A}}^*)\|_{\mathrm{F}}^2, \tag{S4}$$

  holds for any $\boldsymbol{\mathcal{A}} \in \boldsymbol{\Theta}^{\mathrm{SP}}(r_1, r_2, 3s)$. Moreover, by Cauchy-Schwarz and the fact that $xy \leqslant \alpha x^2 + 0.25\alpha^{-1} y^2$, it can be further verified that $\alpha\beta \leqslant 0.25$.

- We assume that $b = \sigma_U^{1/4}$ for simplicity, and the proof can be easily adjusted if $c\sigma_U^{1/4} \leqslant b \leqslant C\sigma_U^{1/4}$ for two absolute constants $0 < c < C$. Moreover, for $0 \leqslant k \leqslant K$ and $i = 1$ and 2,

$$\|\boldsymbol{U}_i^k\|_{\mathrm{op}} \leqslant 1.1\sigma_U^{1/4}, \;\; \sigma_{\min}(\boldsymbol{U}_i^k) \geqslant 0.9\sigma_U^{1/4} \text{ and } \|\boldsymbol{\mathcal{G}}_{(i)}^k\|_{\mathrm{op}} \leqslant 1.1\sigma_U^{1/2}. \tag{S5}$$

  Finally, for all $0 \leqslant k \leqslant K$,

$$E^k \leqslant \frac{c_0 \sigma_L^{1/2}}{\kappa^{3/2}} =: C_1, \tag{S6}$$

  where $\kappa = \sigma_U / \sigma_L$, and $c_0 > 0$ is a small absolute constant (smaller than 1) determined later.

- Let $(\boldsymbol{\mathcal{A}}, \boldsymbol{U}_i, E)$ be $(\boldsymbol{\mathcal{A}}^k, \boldsymbol{U}_i^k, E^k)$, $(\widetilde{\boldsymbol{\mathcal{A}}}_{\check{S}}^{k+1}, \boldsymbol{U}_i^{k+1}, \widetilde{E}^{k+1})$ or $(\boldsymbol{\mathcal{A}}^0, \boldsymbol{U}_i^0, E^0)$, respectively. An important two-sided inequality is derived from Lemma C.2 by letting $b = \sigma_U^{1/4}$ and $c_e = 0.1$, namely

$$C_L\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 \leqslant E \leqslant C_{U,1}\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 + C_{U,2}\sum_{i=1}^{2}\|\boldsymbol{U}_i'\boldsymbol{U}_i - b^2\boldsymbol{I}_{r_i}\|_{\mathrm{F}}^2, \tag{S7}$$

where $C_L = [5(\sigma_U + 2\sigma_U^{3/2})]^{-1}$, $C_{U,1} = 3\sigma_U^{-1} + 8\sigma_L^{-2}\sigma_U^{-1/2} + 40\sigma_L^{-2}$ and $C_{U,2} = 2\sigma_U^{-1/2} + 10$.

Step 2 (Descent of $\widetilde{E}^{k+1}$) This step aims to establish

$$\widetilde{E}^{k+1} \leqslant E^k + \eta^2(Q_{\mathcal{G},1} + \sum_{i=1}^{2} Q_{i,1}) - 2\eta(Q_{\mathcal{G},2} + \sum_{i=1}^{2} Q_{i,2}), \tag{S8}$$

where $Q_{1,j}, Q_{2,j}$ and $Q_{\mathcal{G},j}$ with $j = 1$ and 2 are defined in (S15), (S16) and (S20), respectively. Note that, by the definition of $\widetilde{E}^{k+1}$,

$$\widetilde{E}^{k+1} \leqslant \sum_{i=1}^{2}\|\boldsymbol{U}_i^{k+1} - \boldsymbol{U}_i^*\boldsymbol{R}_i^k\|_{\mathrm{F}}^2 + \|\widetilde{\boldsymbol{\mathcal{G}}}_{\check{S}}^{k+1} - \boldsymbol{\mathcal{G}}^* \times_1 (\boldsymbol{R}_1^k)' \times_2 (\boldsymbol{R}_2^k)'\|_{\mathrm{F}}^2. \tag{S9}$$

For the first term of (S9), the gradient descent update of $\boldsymbol{U}_1^{k+1}$ gives

$$\begin{aligned}\|\boldsymbol{U}_1^{k+1} - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\|_{\mathrm{F}}^2 &= \|\boldsymbol{U}_1^k - \eta[\nabla_{U_1}\mathcal{L}(\boldsymbol{\mathcal{A}}^k) + a\boldsymbol{U}_1^k(\boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1})] - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\|_{\mathrm{F}}^2 \\ &= \|\boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\|_{\mathrm{F}}^2 + \eta^2\|\nabla_{U_1}\mathcal{L}(\boldsymbol{\mathcal{A}}^k) + a\boldsymbol{U}_1^k(\boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1})\|_{\mathrm{F}}^2 \\ &\quad - 2\eta\langle\nabla_{U_1}\mathcal{L}(\boldsymbol{\mathcal{A}}^k), \boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\rangle - 2a\eta\langle\boldsymbol{U}_1^k(\boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1}), \boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\rangle,\end{aligned} \tag{S10}$$

where $\nabla_{U_1}\mathcal{L}(\boldsymbol{\mathcal{A}}) = [\nabla\mathcal{L}(\boldsymbol{\mathcal{A}})]_{(1)}(\boldsymbol{I}_{T_0} \otimes \boldsymbol{U}_2)\boldsymbol{\mathcal{G}}_{(1)}'$ is the partial derivative of the loss function $\mathcal{L}(\boldsymbol{\mathcal{A}})$ with respect to $\boldsymbol{U}_1$. We will first handle the last three terms of (S10) one-by-one (without the scaling constants). Starting with the second term,

$$\|\nabla_{U_1}\mathcal{L}(\boldsymbol{\mathcal{A}}^k) + a\boldsymbol{U}_1^k(\boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1})\|_{\mathrm{F}}^2 \leqslant 2\|\nabla_{U_1}\mathcal{L}(\boldsymbol{\mathcal{A}}^k)\|_{\mathrm{F}}^2 + 2a^2\|\boldsymbol{U}_1^k(\boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1})\|_{\mathrm{F}}^2.$$

Let $C_2 = 1.5\sigma_U^{3/4}$ and, by the definition of dual norm, the conditions at (S5) and $\|\boldsymbol{\mathcal{G}}^k\|_0 = s$,

$$\begin{aligned}
\|\nabla_{U_1}\mathcal{L}(\boldsymbol{\mathcal{A}}^k)\|_{\mathrm{F}}^2 &= \sup_{\boldsymbol{M}\in\mathbb{R}^{N\times r_1},\|\boldsymbol{M}\|_{\mathrm{F}}=1} \langle \nabla_{U_1}\mathcal{L}(\boldsymbol{\mathcal{A}}^k), \boldsymbol{M}\rangle^2 \\
&= \sup_{\boldsymbol{M}\in\mathbb{R}^{N\times r_1},\|\boldsymbol{M}\|_{\mathrm{F}}=1} \langle \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k), \boldsymbol{\mathcal{G}}^k \times_1 \boldsymbol{M} \times_2 \boldsymbol{U}_2^k\rangle^2 \\
&\leqslant 2 \sup_{\boldsymbol{M}\in\mathbb{R}^{N\times r_1},\|\boldsymbol{M}\|_{\mathrm{F}}=1} \langle \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*), \boldsymbol{\mathcal{G}}^k \times_1 \boldsymbol{M} \times_2 \boldsymbol{U}_2^k\rangle^2 \\
&\quad + 2 \sup_{\boldsymbol{M}\in\mathbb{R}^{N\times r_1},\|\boldsymbol{M}\|_{\mathrm{F}}=1} \langle \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k) - \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*), \boldsymbol{\mathcal{G}}^k \times_1 \boldsymbol{M} \times_2 \boldsymbol{U}_2^k\rangle^2 \\
&\leqslant 2C_2^2\left(e_{\mathrm{stat}}^2 + \|\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k) - \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*)\|_{\mathrm{F}}^2\right),
\end{aligned}$$

which leads to

$$\begin{aligned}
&\|\nabla_{U_1}\mathcal{L}(\boldsymbol{\mathcal{A}}^k) + a\boldsymbol{U}_1^k(\boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1})\|_{\mathrm{F}}^2 \\
&\quad \leqslant 4C_2^2\left(e_{\mathrm{stat}}^2 + \|\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k) - \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*)\|_{\mathrm{F}}^2\right) + 3\sigma_U^{1/2}a^2\|\boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1}\|_{\mathrm{F}}^2 =: Q_{1,1}. \qquad \text{(S11)}
\end{aligned}$$

We next consider the third term at (S10). Let $\boldsymbol{\mathcal{A}}_{U_1} = \boldsymbol{\mathcal{G}} \times_1 (\boldsymbol{U}_1 - \boldsymbol{U}_1^*\boldsymbol{R}_1) \times_2 \boldsymbol{U}_2$ and, by the conditions at (S5), $\|\boldsymbol{\mathcal{A}}_{U_1}^k\|_{\mathrm{F}} \leqslant C_2\|\boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\|_{\mathrm{F}}$ for all $k \geqslant 1$. Moreover, by the conditions at (S5) and the fact that $xy \leqslant 0.5x^2 + 0.5y^2$,

$$\begin{aligned}
\langle \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*), \boldsymbol{\mathcal{A}}_{U_1}^k\rangle &\leqslant \sup_{\boldsymbol{\mathcal{M}}\in\Theta_1^{\mathrm{SP}}(r_1,r_2,s)} \langle \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*), \boldsymbol{\mathcal{M}}\rangle \cdot \|\boldsymbol{\mathcal{A}}_{U_1}^k\|_{\mathrm{F}} \\
&\leqslant 0.5C_2^2C_1^{-1}e_{\mathrm{stat}}^2 + 0.5C_1\|\boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\|_{\mathrm{F}}^2.
\end{aligned}$$

As a result,

$$\begin{aligned}
\langle \nabla_{U_1}\mathcal{L}(\boldsymbol{\mathcal{A}}^k), \boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\rangle &= \langle \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k), \boldsymbol{\mathcal{G}}^k \times_1 (\boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k) \times_2 \boldsymbol{U}_2^k\rangle \\
&= \langle \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k) - \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*), \boldsymbol{\mathcal{A}}_{U_1}^k\rangle + \langle \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*), \boldsymbol{\mathcal{A}}_{U_1}^k\rangle \\
&\geqslant \langle \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k) - \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*), \boldsymbol{\mathcal{A}}_{U_1}^k\rangle - 0.5C_1\|\boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\|_{\mathrm{F}}^2 \\
&\quad - 0.5C_2^2C_1^{-1}e_{\mathrm{stat}}^2. \qquad \text{(S12)}
\end{aligned}$$

For the last term at (S10), it holds that

$$
\begin{aligned}
\langle \boldsymbol{U}_1^k(\boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1}), \boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\rangle &= \langle \boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1}, \boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - \boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^*\boldsymbol{R}_1^k\rangle \\
&= 0.5\|\boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1}\|_{\mathrm{F}}^2 + 0.5\langle \boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1}, (\boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k)'(\boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k)\rangle \\
&\geqslant 0.25\|\boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1}\|_{\mathrm{F}}^2 - 0.25\|\boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\|_{\mathrm{F}}^4 \\
&\geqslant 0.25\|\boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1}\|_{\mathrm{F}}^2 - 0.25C_1\|\boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\|_{\mathrm{F}}^2,
\end{aligned}
\tag{S13}
$$

where the second equality is due to the fact that, for a symmetric matrix $\boldsymbol{W} \in \mathbb{R}^{r_1\times r_1}$,

$$
\langle \boldsymbol{W}, \boldsymbol{U}_1'\boldsymbol{U}_1 - \boldsymbol{U}_1'\boldsymbol{U}_1^*\boldsymbol{R}_1\rangle = 0.5\langle \boldsymbol{W}, (\boldsymbol{U}_1 - \boldsymbol{U}_1^*\boldsymbol{R}_1)'(\boldsymbol{U}_1 - \boldsymbol{U}_1^*\boldsymbol{R}_1)\rangle + 0.5\langle \boldsymbol{W}, \boldsymbol{U}_1'\boldsymbol{U}_1 - b^2\boldsymbol{I}_{r_1}\rangle,
$$

the first inequality is due to Cauchy-Schwarz inequality and the fact that $xy \leqslant 0.5x^2 + 0.5y^2$, and the last inequality is due to the fact that $\|\boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\|_{\mathrm{F}}^2 \leqslant E^k \leqslant C_1$ at (S6). We combine the last two terms of (S10), i.e. (S12) and (S13), with the scaling constant $a > 0$, and it leads to

$$
\begin{aligned}
&\langle \nabla_{U_1}\mathcal{L}(\boldsymbol{\mathcal{A}}^k), \boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\rangle + a\langle \boldsymbol{U}_1^k(\boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1}), \boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\rangle \\
&\geqslant \langle \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k) - \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*), \boldsymbol{\mathcal{A}}_{U_1}^k\rangle - (1+a)C_1\|\boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\|_{\mathrm{F}}^2 \\
&\quad + 0.25a\|\boldsymbol{U}_1^{k\prime}\boldsymbol{U}_1^k - b^2\boldsymbol{I}_{r_1}\|_{\mathrm{F}}^2 - 0.5C_2^2C_1^{-1}e_{\mathrm{stat}}^2 =: Q_{1,2}.
\end{aligned}
\tag{S14}
$$

By plugging (S11) and (S14) into (S10), it holds that

$$
\|\boldsymbol{U}_1^{k+1} - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\|_{\mathrm{F}}^2 \leqslant \|\boldsymbol{U}_1^k - \boldsymbol{U}_1^*\boldsymbol{R}_1^k\|_{\mathrm{F}}^2 + \eta^2 Q_{1,1} - 2\eta Q_{1,2}. \tag{S15}
$$

We can similarly define $Q_{2,1}$ and $Q_{2,2}$ such that

$$
\|\boldsymbol{U}_2^{k+1} - \boldsymbol{U}_2^*\boldsymbol{R}_2^k\|_{\mathrm{F}}^2 \leqslant \|\boldsymbol{U}_2^k - \boldsymbol{U}_2^*\boldsymbol{R}_2^k\|_{\mathrm{F}}^2 + \eta^2 Q_{2,1} - 2\eta Q_{2,2}. \tag{S16}
$$

We now return to deal with the last component in (S9). Let $\nabla_{\mathcal{G}}\mathcal{L}(\boldsymbol{\mathcal{A}})$ be the partial derivative of the loss function $\mathcal{L}(\boldsymbol{\mathcal{A}})$ with respect to $\boldsymbol{\mathcal{G}}$, and it holds that $\nabla_{\mathcal{G}}\mathcal{L}(\boldsymbol{\mathcal{A}}) = \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}) \times_1 \boldsymbol{U}_1' \times_2 \boldsymbol{U}_2'$. Then the gradient descent update of $\widetilde{\boldsymbol{\mathcal{G}}}_{\breve{S}}^{k+1}$ gives

$$
\begin{aligned}
\|\widetilde{\boldsymbol{\mathcal{G}}}_{\breve{S}}^{k+1} - \boldsymbol{\mathcal{G}}^* \times_1 (\boldsymbol{R}_1^k)' \times_2 (\boldsymbol{R}_2^k)'\|_{\mathrm{F}}^2 &= \|\boldsymbol{\mathcal{G}}^k - \eta[\nabla_{\mathcal{G}}\mathcal{L}(\boldsymbol{\mathcal{A}}^k)]_{\breve{S}} - \boldsymbol{\mathcal{G}}^* \times_1 (\boldsymbol{R}_1^k)' \times_2 (\boldsymbol{R}_2^k)'\|_{\mathrm{F}}^2 \\
&= \|\boldsymbol{\mathcal{G}}^k - \boldsymbol{\mathcal{G}}^* \times_1 (\boldsymbol{R}_1^k)' \times_2 (\boldsymbol{R}_2^k)'\|_{\mathrm{F}}^2 + \eta^2\|[\nabla_{\mathcal{G}}\mathcal{L}(\boldsymbol{\mathcal{A}}^k)]_{\breve{S}}\|_{\mathrm{F}}^2 \\
&\quad - 2\eta\langle \nabla_{\mathcal{G}}\mathcal{L}(\boldsymbol{\mathcal{A}}^k), \boldsymbol{\mathcal{G}}^k - \boldsymbol{\mathcal{G}}^* \times_1 (\boldsymbol{R}_1^k)' \times_2 (\boldsymbol{R}_2^k)'\rangle.
\end{aligned}
\tag{S17}
$$

For the second term at (S17), by the definition of dual norm, $|\breve{S}| \leqslant 3s$ and the conditions at (S5),

$$\begin{aligned}
\|[\nabla_{\mathcal{G}}\mathcal{L}(\boldsymbol{\mathcal{A}}^k)]_{\breve{S}}\|_{\mathrm{F}}^2 &= \sup_{\boldsymbol{\mathcal{M}}\in\mathbb{R}^{r_1\times r_2\times T_0},\|\boldsymbol{\mathcal{M}}\|_{\mathrm{F}}=1} \langle[\nabla_{\mathcal{G}}\mathcal{L}(\boldsymbol{\mathcal{A}}^k)]_{\breve{S}},\boldsymbol{\mathcal{M}}\rangle^2 \\
&\leqslant \sup_{\boldsymbol{\mathcal{M}}\in\mathbb{R}^{r_1\times r_2\times T_0},\|\boldsymbol{\mathcal{M}}\|_{\mathrm{F}}=1,\|\boldsymbol{\mathcal{M}}\|_0\leqslant 3s} \langle\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k),\boldsymbol{\mathcal{M}}\times_1\boldsymbol{U}_1^k\times_2\boldsymbol{U}_2^k\rangle^2 \\
&\leqslant 2\sup_{\boldsymbol{\mathcal{M}}\in\mathbb{R}^{r_1\times r_2\times T_0},\|\boldsymbol{\mathcal{M}}\|_{\mathrm{F}}=1,\|\boldsymbol{\mathcal{M}}\|_0\leqslant 3s} \langle\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*),\boldsymbol{\mathcal{M}}\times_1\boldsymbol{U}_1^k\times_2\boldsymbol{U}_2^k\rangle^2 \\
&\quad + 2\sup_{\boldsymbol{\mathcal{M}}\in\mathbb{R}^{r_1\times r_2\times T_0},\|\boldsymbol{\mathcal{M}}\|_{\mathrm{F}}=1,\|\boldsymbol{\mathcal{M}}\|_0\leqslant 3s} \langle\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k)-\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*),\boldsymbol{\mathcal{M}}\times_1\boldsymbol{U}_1^k\times_2\boldsymbol{U}_2^k\rangle^2 \\
&\leqslant 2C_3^2\left(e_{\mathrm{stat}}^2+\|\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k)-\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*)\|_{\mathrm{F}}^2\right) =: Q_{\mathcal{G},1},
\end{aligned} \tag{S18}$$

with $C_3 = 1.5\sigma_U^{1/2}$. For the last term at (S17), let $\boldsymbol{\mathcal{A}}_{\mathcal{G}} = (\boldsymbol{\mathcal{G}}-\boldsymbol{\mathcal{G}}^*\times_1\boldsymbol{R}_1'\times_2\boldsymbol{R}_2')\times_1\boldsymbol{U}_1\times_2\boldsymbol{U}_2$ and, by the conditions at (S5), $\|\boldsymbol{\mathcal{A}}_{\mathcal{G}}^k\|_{\mathrm{F}} \leqslant C_3\|\boldsymbol{\mathcal{G}}^k-\boldsymbol{\mathcal{G}}^*\times_1(\boldsymbol{R}_1^k)'\times_2(\boldsymbol{R}_2^k)'\|_{\mathrm{F}}$. From the definition of $\nabla_{\mathcal{G}}\mathcal{L}(\boldsymbol{\mathcal{A}}^k)$, it holds that

$$\langle\nabla_{\mathcal{G}}\mathcal{L}(\boldsymbol{\mathcal{A}}^k),\boldsymbol{\mathcal{G}}^k-\boldsymbol{\mathcal{G}}^*\times_1(\boldsymbol{R}_1^k)'\times_2(\boldsymbol{R}_2^k)'\rangle = \langle\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k)-\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*),\boldsymbol{\mathcal{A}}_{\mathcal{G}}^k\rangle + \langle\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*),\boldsymbol{\mathcal{A}}_{\mathcal{G}}^k\rangle.$$

By the conditions at (S5), $|\breve{S}| \leqslant 3s$ and that $xy \leqslant 0.5x^2+0.5y^2$,

$$\begin{aligned}
\langle\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*),\boldsymbol{\mathcal{A}}_{\mathcal{G}}^k\rangle &\leqslant \sup_{\boldsymbol{\mathcal{M}}\in\boldsymbol{\Theta}_1^{\mathrm{SP}}(r_1,r_2,2s)}\langle\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*),\boldsymbol{\mathcal{M}}\rangle\cdot\|\boldsymbol{\mathcal{A}}_{\mathcal{G}}^k\|_{\mathrm{F}} \\
&\leqslant 0.5C_3^2C_1^{-1}e_{\mathrm{stat}}^2 + 0.5C_1\|\boldsymbol{\mathcal{G}}^k-\boldsymbol{\mathcal{G}}^*\times_1(\boldsymbol{R}_1^k)'\times_2(\boldsymbol{R}_2^k)'\|_{\mathrm{F}}^2,
\end{aligned}$$

which leads to

$$\begin{aligned}
&\langle\nabla_{\mathcal{G}}\mathcal{L}(\boldsymbol{\mathcal{A}}^k),\boldsymbol{\mathcal{G}}^k-\boldsymbol{\mathcal{G}}^*\times_1(\boldsymbol{R}_1^k)'\times_2(\boldsymbol{R}_2^k)'\rangle \\
&\geqslant \langle\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k)-\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*),\boldsymbol{\mathcal{A}}_{\mathcal{G}}^k\rangle - C_1\|\boldsymbol{\mathcal{G}}^k-\boldsymbol{\mathcal{G}}^*\times_1(\boldsymbol{R}_1^k)'\times_2(\boldsymbol{R}_2^k)'\|_{\mathrm{F}}^2 - C_3^2C_1^{-1}e_{\mathrm{stat}}^2 =: Q_{\mathcal{G},2}.
\end{aligned} \tag{S19}$$

By plugging (S18) and (S19) into (S17), we can obtain that

$$\|\widetilde{\boldsymbol{\mathcal{G}}}_{\breve{S}}^{k+1}-\boldsymbol{\mathcal{G}}^*\times_1(\boldsymbol{R}_1^k)'\times_2(\boldsymbol{R}_2^k)'\|_{\mathrm{F}}^2 \leqslant \|\boldsymbol{\mathcal{G}}^k-\boldsymbol{\mathcal{G}}^*\times_1(\boldsymbol{R}_1^k)'\times_2(\boldsymbol{R}_2^k)'\|_{\mathrm{F}}^2 + \eta^2Q_{\mathcal{G},1} - 2\eta Q_{\mathcal{G},2}, \tag{S20}$$

which, together with (S14) and (S15), leads to (S8).

Step 3 This step aims to develop a lower bound for $Q_{\mathcal{G},2}+\sum_{i=1}^{2}Q_{i,2}$ at (S8). From (S14), a similar definition for $Q_{2,2}$, (S19) and the definition of E^k, we have

$$\begin{aligned}Q_{\mathcal{G},2}+\sum_{i=1}^{2}Q_{i,2}\geqslant&\langle\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k)-\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*),\boldsymbol{\mathcal{A}}_{\mathcal{G}}^k+\sum_{i=1}^{2}\boldsymbol{\mathcal{A}}_{U_i}^k\rangle-(1+a)C_1E^k\\&+0.25a\sum_{i=1}^{2}\|\boldsymbol{U}_i^{k\prime}\boldsymbol{U}_i^k-b^2\boldsymbol{I}_{r_i}\|_{\mathrm{F}}^2-C_4e_{\mathrm{stat}}^2,\end{aligned}\tag{S21}$$

where $C_4=(C_2^2+C_3^2)C_1^{-1}$. By the conditions at (S5) and (S6), Lemma C.3 holds with $B\leqslant(E^k)^{1/2}\leqslant(C_1E^k)^{1/4}$. Then, by plugging $c_e=0.1$ to Lemma C.3, the first term becomes

$$\begin{aligned}\langle\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k)-\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*),\boldsymbol{\mathcal{A}}_{\mathcal{G}}^k+\sum_{i=1}^{2}\boldsymbol{\mathcal{A}}_{U_i}^k\rangle=&\langle\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k)-\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*),\boldsymbol{\mathcal{A}}^k-\boldsymbol{\mathcal{A}}^*\rangle\\&+\langle\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k)-\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*),\boldsymbol{\mathcal{H}}_\epsilon^k\rangle,\end{aligned}$$

where $\|\boldsymbol{\mathcal{H}}_\epsilon^k\|_{\mathrm{F}}\leqslant3(\sigma_U^{1/2}+\sigma_U^{1/4})(C_1E^k)^{1/2}$. Given (S4) holds for some given $\alpha,\beta>0$, we can use $xy\leqslant0.5x^2+0.5y^2$ to further lower bound the above term,

$$\begin{aligned}\langle\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k)-\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*),\boldsymbol{\mathcal{A}}_{\mathcal{G}}^k+\sum_{i=1}^{2}\boldsymbol{\mathcal{A}}_{U_i}^k\rangle\geqslant&\,\alpha\|\boldsymbol{\mathcal{A}}^k-\boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2\\&+0.5\beta\|\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k)-\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*)\|_{\mathrm{F}}^2-0.5\beta^{-1}\|\boldsymbol{\mathcal{H}}_\epsilon^k\|_{\mathrm{F}}^2\\\geqslant&\,\frac{\alpha\sigma_L^{3/2}}{\kappa^{1/2}}C_5E^k-\alpha C_{U,1}^{-1}C_{U,2}\sum_{i=1}^{2}\|\boldsymbol{U}_i^{k\prime}\boldsymbol{U}_i^k-b^2\boldsymbol{I}_{r_i}\|_{\mathrm{F}}^2\\&+0.5\beta\|\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k)-\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*)\|_{\mathrm{F}}^2-0.5\beta^{-1}\|\boldsymbol{\mathcal{H}}_\epsilon^k\|_{\mathrm{F}}^2,\end{aligned}$$

where the second inequality is obtained from (S7) with $C_5=[\sigma_L^{3/2}\kappa^{-1/2}C_{U,1}]^{-1}$. Then, this jointly with $\|\boldsymbol{\mathcal{H}}_\epsilon^k\|_{\mathrm{F}}\leqslant3(\sigma_U^{1/2}+\sigma_U^{1/4})(C_1E^k)^{1/2}$ can be used to further lower bound the inequality at (S21), which becomes

$$\begin{aligned}Q_{\mathcal{G},2}+\sum_{i=1}^{2}Q_{i,2}\geqslant&\left(\frac{\alpha\sigma_L^{3/2}}{\kappa^{1/2}}C_5-[1+a+5\beta^{-1}(\sigma_U^{1/2}+\sigma_U^{1/4})^2]C_1\right)E^k\\&+[0.25a-\alpha C_{U,1}^{-1}C_{U,2}]\sum_{i=1}^{2}\|\boldsymbol{U}_i^{k\prime}\boldsymbol{U}_i^k-b^2\boldsymbol{I}_{r_i}\|_{\mathrm{F}}^2\\&+0.5\beta\|\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k)-\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*)\|_{\mathrm{F}}^2-C_4e_{\mathrm{stat}}^2.\end{aligned}$$

Then, we choose $c_0 > 0$ contained in the constant term C_1 small enough such that $[1 + a + 5\beta^{-1}(\sigma_U^{1/2} + \sigma_U^{1/4})^2]C_1 \leqslant 0.5\alpha\sigma_L^{3/2}\kappa^{-1/2}C_5$. From the definition right after (S7), we have $C_{U,1} \geqslant 3\sigma_U^{-1}$ and $C_{U,2} \leqslant 10(\sigma_U^{-1/2} + 1)$, and they lead to $C_{U,1}^{-1}C_{U,2} \leqslant 4(\sigma_U^{1/2} + \sigma_U)$. Let $a = 80\alpha(\sigma_U^{1/2} + \sigma_U)$ leading to $\alpha C_{U,1}^{-1}C_{U,2} \leqslant 0.05a$, and the above inequality takes the simple form

$$\begin{aligned} Q_{\mathcal{G},2} + \sum_{i=1}^{2} Q_{i,2} \geqslant & \frac{\alpha\sigma_L^{3/2}}{2\kappa^{1/2}}C_5E^k + 0.2a\sum_{i=1}^{2}\|\boldsymbol{U}_i^{k\prime}\boldsymbol{U}_i^k - b^2\boldsymbol{I}_{r_i}\|_{\mathrm{F}}^2 - C_4e_{\mathrm{stat}}^2 \\ & + 0.5\beta\|\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k) - \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*)\|_{\mathrm{F}}^2. \end{aligned} \tag{S22}$$

Step 4 This step uses the intermediate results at Steps 2 and 3 to bound $\widetilde{E}^{k+1}$ with E^k. From (S11), a similar definition for $Q_{2,1}$ and (S18), we have

$$\begin{aligned} Q_{\mathcal{G},1} + \sum_{i=1}^{2} Q_{i,1} = & (8C_2^2 + 2C_3^2)\left\{\|\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k) - \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*)\|_{\mathrm{F}}^2 + e_{\mathrm{stat}}^2\right\} \\ & + 3\sigma_U^{1/2}a^2\sum_{i=1}^{2}\|\boldsymbol{U}_i^{k\prime}\boldsymbol{U}_i^k - b^2\boldsymbol{I}_{r_i}\|_{\mathrm{F}}^2, \end{aligned}$$

which, together with (S22), implies that

$$\begin{aligned} \eta^2(Q_{\mathcal{G},1} + \sum_{i=1}^{2} Q_{i,1}) - 2\eta(Q_{\mathcal{G},2} + \sum_{i=1}^{2} Q_{i,2}) \leqslant & -\frac{\eta\alpha\sigma_L^{3/2}}{\kappa^{1/2}}C_5E^k \\ & + [\eta^2(8C_2^2 + 2C_3^2) + 2\eta C_4]e_{\mathrm{stat}}^2 \\ & + \left(3\eta^2\sigma_U^{1/2}a^2 - 0.4\eta a\right)\sum_{i=1}^{2}\|\boldsymbol{U}_i^{k\prime}\boldsymbol{U}_i^k - b^2\boldsymbol{I}_{r_i}\|_{\mathrm{F}}^2 \\ & + \left(\eta^2(8C_2^2 + 2C_3^2) - \eta\beta\right)\|\nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^k) - \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}^*)\|_{\mathrm{F}}^2. \end{aligned} \tag{S23}$$

Recall that $C_2 = 1.5\sigma_U^{3/4}$, $C_3 = 1.5\sigma_U^{1/2}$ and $a = 80\alpha(\sigma_U^{1/2} + \sigma_U)$, and it holds that $8C_2^2 + 2C_3^2 \leqslant 18\sigma_U(1 + \sigma_U^{1/2})$ and $\sigma_U^{1/2}a \leqslant 80\alpha\sigma_U(1 + \sigma_U^{1/2}) \leqslant 20\beta^{-1}\sigma_U(1 + \sigma_U^{1/2})$, where the second inequality comes from $\alpha\beta \leqslant 0.25$; see the discussion after (S4). Set $\eta = \eta_0\beta[(1 + \sigma_U)(1 + \sigma_U^{1/2})]^{-1}$ and $\eta_0 \leqslant 1/150$, and it can be verified that

$$3\eta^2\sigma_U^{1/2}a^2 - 0.4\eta a \leqslant 0 \ \text{ and } \ \eta^2(8C_2^2 + 2C_3^2) - \eta\beta \leqslant 0. \tag{S24}$$

Again, note that $\eta = \eta_0\beta[(1 + \sigma_U)(1 + \sigma_U^{1/2})]^{-1}$ with $\eta_0 \leqslant 1/150$ and $\eta(C_2^2 + C_3^2) \leqslant 0.02\beta$, leading to $\eta^2(8C_2^2 + 2C_3^2) \leqslant 0.02\beta^2$ and $\eta C_4 = \eta(C_2^2 + C_3^2)C_1^{-1} \leqslant 0.02\beta C_1^{-1}$. This, together with (S8), (S23) and (S24), implies that

$$\widetilde{E}^{k+1} \leqslant (1 - \eta_0 \alpha\beta C_6)\, E^k + 0.02\beta(\beta + C_1^{-1}) e_{\text{stat}}^2, \tag{S25}$$

where $C_6 = \sigma_L^{3/2}\kappa^{-1/2}(1+\sigma_U)^{-1}(1+\sigma_U^{1/2})^{-1}C_5$, and

$$\eta_0 \alpha\beta C_6 \leqslant \frac{\eta_0 \alpha\beta}{C_{U,1}} < \frac{\eta_0 \sigma_U}{204} < 1$$

since $\eta_0 < 204\sigma_U^{-1}$ and $\alpha\beta \leqslant 0.25$. Moreover,

$$\alpha\beta C_6 = \frac{\alpha\beta}{(1+\sigma_U)(1+\sigma_U^{1/2})C_{U,1}} \geqslant \frac{\alpha\beta}{204\kappa^2} := \delta_{\alpha,\beta},$$

which, together with (S25), implies that

$$\widetilde{E}^{k+1} \leqslant (1 - \eta_0 \delta_{\alpha,\beta})\, E^k + 0.02\beta(\beta + C_1^{-1}) e_{\text{stat}}^2. \tag{S26}$$

Step 5 This step upper bounds E^{k+1} by $\widetilde{E}^{k+1}$, and hence by E^k. To begin with, we first establish the inequality between E^{k+1} and $\widetilde{E}^{k+1}$,

$$E^{k+1} \leqslant \sum_{i=1}^{2} \|\boldsymbol{U}_i^{k+1} - \boldsymbol{U}_i^* \widetilde{\boldsymbol{R}}_i^{k+1}\|_{\text{F}}^2 + \|\boldsymbol{\mathcal{G}}^{k+1} - \boldsymbol{\mathcal{G}}^* \times_1 (\widetilde{\boldsymbol{R}}_1^{k+1})' \times_2 (\widetilde{\boldsymbol{R}}_2^{k+1})'\|_{\text{F}}^2, \tag{S27}$$

and $\widetilde{E}^{k+1} = \sum_{i=1}^2 \|\boldsymbol{U}_i^{k+1} - \boldsymbol{U}_i^* \widetilde{\boldsymbol{R}}_i^{k+1}\|_{\text{F}}^2 + \|\widetilde{\boldsymbol{\mathcal{G}}}_{\breve{S}}^{k+1} - \boldsymbol{\mathcal{G}}^* \times_1 (\widetilde{\boldsymbol{R}}_1^{k+1})' \times_2 (\widetilde{\boldsymbol{R}}_2^{k+1})'\|_{\text{F}}^2$, where $\breve{S} = S_k \cup S_{k+1} \cup S_\gamma$. Denote by $\breve{s}$ the cardinality of $\breve{S}$, i.e., $\breve{s} = |S_{k+1} \cup S_k \cup S_\gamma|$. Note that $2xy \leqslant \mu x^2 + \mu^{-1} y^2$ for any $\mu > 0$ and $x, y \in \mathbb{R}$, and then the second term at the right hand side of (S27) satisfies the following inequality,

$$\begin{aligned} &\|\boldsymbol{\mathcal{G}}^{k+1} - \boldsymbol{\mathcal{G}}^* \times_1 (\widetilde{\boldsymbol{R}}_1^{k+1})' \times_2 (\widetilde{\boldsymbol{R}}_2^{k+1})'\|_{\text{F}}^2 \\ &\quad \leqslant (1+\mu_s)\|\widetilde{\boldsymbol{\mathcal{G}}}_{\breve{S}}^{k+1} - \boldsymbol{\mathcal{G}}^* \times_1 (\widetilde{\boldsymbol{R}}_1^{k+1})' \times_2 (\widetilde{\boldsymbol{R}}_2^{k+1})'\|_{\text{F}}^2 + (1 + \frac{1}{\mu_s})\|\boldsymbol{\mathcal{G}}^{k+1} - \widetilde{\boldsymbol{\mathcal{G}}}_{\breve{S}}^{k+1}\|_{\text{F}}^2, \end{aligned} \tag{S28}$$

where $\mu_s = \sqrt{(\breve{s} - s)/(\breve{s} - s_\gamma)} < 1$.

Moreover, $\boldsymbol{\mathcal{A}}^{k+1} = \boldsymbol{\mathcal{G}}^{k+1} \times_1 \boldsymbol{U}_1^{k+1} \times_2 \boldsymbol{U}_2^{k+1}$, $\widetilde{\boldsymbol{\mathcal{A}}}^{k+1} = \widetilde{\boldsymbol{\mathcal{G}}}_{\breve{S}}^{k+1} \times_1 \boldsymbol{U}_1^{k+1} \times_2 \boldsymbol{U}_2^{k+1}$, and $\boldsymbol{\mathcal{A}}^{k+1} = \text{HT}(\widetilde{\boldsymbol{\mathcal{A}}}_{\breve{\text{S}}}^{\text{k+1}}, \text{s})$. Since $s_\gamma < s \leqslant \breve{s}$, it can be verified that

$$\begin{aligned} \|\boldsymbol{\mathcal{G}}^{k+1} - \widetilde{\boldsymbol{\mathcal{G}}}_{\breve{S}}^{k+1}\|_{\text{F}}^2 &\leqslant \left[\prod_{i=1}^{2} \sigma_{\min}(\boldsymbol{U}_i^{k+1})\right]^{-2} \|\boldsymbol{\mathcal{A}}^{k+1} - \widetilde{\boldsymbol{\mathcal{A}}}_{\breve{S}}^{k+1}\|_{\text{F}}^2 \\ &\overset{\text{(Lemma C.4)}}{\leqslant} \mu_s^2 \left[\prod_{i=1}^{2} \sigma_{\min}(\boldsymbol{U}_i^{k+1})\right]^{-2} \|\widetilde{\boldsymbol{\mathcal{A}}}_{\breve{S}}^{k+1} - \boldsymbol{\mathcal{A}}^*\|_{\text{F}}^2 \overset{\text{(S5)\&(S7)}}{\leqslant} C_7 \mu_s^2 \widetilde{E}^{k+1}, \end{aligned} \tag{S29}$$

which, together with (S27) and (S28), implies that

$$E^{k+1} \leqslant [1 + (1 + 2C_7)\mu_s]\widetilde{E}^{k+1}, \tag{S30}$$

where $C_7 = 2\sigma_U^{-1}C_L^{-1} = 10(1 + 2\sigma_U^{1/2})$.

Note that μ_s is a function of $\breve{s}$ with $\breve{s} \geqslant s$ and, by the condition of $|S_k \cup S_{k+1}| \leqslant (1+\nu)s$, it holds that $\breve{s} \leqslant (1+\nu_k)s + s_\gamma$. Note that $s_\gamma \leqslant \nu_k s$, and it can be verified that $\mu_s \leqslant \sqrt{2\nu_k/(1+\nu_k)} < \sqrt{2\nu_k}$, and hence $(1+2C_7)\mu_s \leqslant \eta_0\delta_{\alpha,\beta}$ as long as $\nu_k \leqslant (1/128)(3 + 5\sigma_U^{1/2})^{-2}\eta_0^2\delta_{\alpha,\beta}^2$. As a result, from (S26) and (S30),

$$E^{k+1} \leqslant \left(1 - \eta_0^2\delta_{\alpha,\beta}^2\right) E^k + C_8 e_{\text{stat}}^2, \tag{S31}$$

where $C_8 = 0.02[1 + (1 + 2C_7)\mu_s]\beta(\beta + C_1^{-1})$. By unfolding this iteration, we can obtain that

$$E^K \leqslant \left(1 - \eta_0^2\delta_{\alpha,\beta}^2\right)^K E^0 + \eta_0^{-2}\delta_{\alpha,\beta}^{-2}C_8 e_{\text{stat}}^2,$$

which, together with (S7) and $\boldsymbol{U}_i^{0\prime}\boldsymbol{U}_i^0 = b^2\boldsymbol{I}_{r_i}$ for $1 \leqslant i \leqslant 2$, leads to

$$\begin{aligned}\|\boldsymbol{\mathcal{A}}^K - \boldsymbol{\mathcal{A}}^*\|_{\text{F}}^2 &\leqslant C_L^{-1}E^K \leqslant C_L^{-1}\left(1 - \eta_0^2\delta_{\alpha,\beta}^2\right)^K E^0 + C_L^{-1}C_8\eta_0^{-2}\delta_{\alpha,\beta}^{-2}e_{\text{stat}}^2 \\ &\leqslant C_{U,1}C_L^{-1}\left(1 - \eta_0^2\delta_{\alpha,\beta}^2\right)^K \|\boldsymbol{\mathcal{A}}^0 - \boldsymbol{\mathcal{A}}^*\|_{\text{F}}^2 + C_{U,1}C_L^{-1}C_8\eta_0^{-2}\delta_{\alpha,\beta}^{-2}e_{\text{stat}}^2. \end{aligned}\tag{S32}$$

Step 6 (Verifying conditions at (S4), (S5) and (S6)) We first show that the conditions at (S4) hold with certain values of α and β. From Lemmas C.1 and C.5, if $T_1 \gtrsim s^2(N + \log T_0)$ and we choose

$$\alpha = \frac{3\kappa_{\text{RSC}}\kappa_{\text{RSS}}}{\kappa_{\text{RSC}} + 3\kappa_{\text{RSS}}} \quad \text{and} \quad \beta = \frac{1}{\kappa_{\text{RSC}} + 3\kappa_{\text{RSS}}},$$

then the inequality at (S4) holds with probability at least $1 - Ce^{-N-\log T_0}$. Note that $\kappa_{\text{RSC}} \leqslant \kappa_{\text{RSS}}$ and $\alpha\beta \leqslant 0.25$, and it can be further verified that

$$\beta \leqslant 0.25\kappa_{\text{RSC}}^{-1} \quad \text{and} \quad \frac{3\kappa_{\text{RSC}}}{16\kappa_{\text{RSS}}} \leqslant \alpha\beta \leqslant \frac{1}{4}, \tag{S33}$$

which can be used to update quantities, $\delta_{\alpha,\beta}$, a and ν_k in Steps 4, 3 and 5, respectively. Specifically,

$$\delta_{\alpha,\beta} = \frac{\alpha\beta}{204\kappa^2} \geqslant \frac{\kappa_{\text{RSC}}}{1088\kappa_{\text{RSS}}\kappa^2} := \delta,$$

$$a = 80\alpha(\sigma_U^{1/2} + \sigma_U) = \frac{240(\sigma_U^{1/2} + \sigma_U)}{\kappa_{\text{RSS}}^{-1} + 3\kappa_{\text{RSC}}^{-1}} \quad \text{and} \quad \nu_k \leqslant 10^{-10} \cdot \frac{\eta_0^2 \kappa_{\text{RSC}}^2}{\kappa_{\text{RSS}}^2 \kappa^4},$$

where the final inequality ensures that $\nu_k \leqslant (1/128)\eta_0^2 \delta_{\alpha,\beta}^2 (3 + 5\sigma_U^{1/2})^{-2}$. Moreover, since $\kappa_{\text{RSC}} < 1$ and $c_0 < 1$,

$$C_8 = 0.02[1 + (1 + 2C_7)\mu_s]\beta(\beta + C_1^{-1}) \leqslant 0.62 c_0^{-1} \kappa^2 \kappa_{\text{RSC}}^{-2}, \tag{S34}$$

which, together with the fact that $C_{U,1} C_L^{-1} \lesssim \kappa^{3/2} \sigma_L^{-1/2}$, can be used to rewrite (S32) into

$$\|\boldsymbol{\mathcal{A}}^K - \boldsymbol{\mathcal{A}}^*\|_{\text{F}}^2 \lesssim \kappa^{3/2} \sigma_L^{-1/2} \left(1 - \eta_0^2 \delta^2\right)^K \|\boldsymbol{\mathcal{A}}^0 - \boldsymbol{\mathcal{A}}^*\|_{\text{F}}^2 + \kappa^{7/2} \sigma_L^{-1/2} \kappa_{\text{RSC}}^{-2} \eta_0^{-2} \delta^{-2} e_{\text{stat}}^2.$$

We next verify (S6). Note that $\|\boldsymbol{U}_i^{0\prime} \boldsymbol{U}_i^0 - b^2 \boldsymbol{I}_{r_i}\|_{\text{F}}^2 = 0$ for $1 \leqslant i \leqslant 2$ and, by the initialization error bound and (S7), it holds that

$$E^0 \leqslant C_{U,1} \|\boldsymbol{\mathcal{A}}^0 - \boldsymbol{\mathcal{A}}^*\|_{\text{F}}^2 \leqslant c_0 \frac{\sigma_L^{1/2}}{\kappa^{3/2}}. \tag{S35}$$

Suppose that the above inequality holds for E^k. Then, by (S31) and (S34),

$$\begin{aligned}
E^{k+1} &\leqslant \left(1 - \eta_0^2 \delta^2\right) E^k + 0.62 c_0^{-1} \kappa^2 \kappa_{\text{RSC}}^{-2} e_{\text{stat}}^2 \\
&\leqslant \left(1 - \eta_0^2 \delta^2\right) \cdot c_0 \frac{\sigma_L^{1/2}}{\kappa^{3/2}} + 0.62 c_0^{-1} \kappa^2 \kappa_{\text{RSC}}^{-2} e_{\text{stat}}^2 \\
&= c_0 \frac{\sigma_L^{1/2}}{\kappa^{3/2}} - \left(c_0 \frac{\eta_0^2 \delta^2 \sigma_L^{1/2}}{\kappa^{3/2}} - 0.62 c_0^{-1} \kappa^2 \kappa_{\text{RSC}}^{-2} e_{\text{stat}}^2 \right) \leqslant c_0 \frac{\sigma_L^{1/2}}{\kappa^{3/2}}
\end{aligned}$$

since, when $e_{\text{stat}}^2 \leqslant (1/800) c_0^2 \eta_0^2 \kappa^{-8} \kappa_{\text{RSS}}^{-4} \kappa_{\text{RSC}}^4$,

$$c_0 \frac{\eta_0^2 \delta^2 \sigma_L^{1/2}}{\kappa^{3/2}} \geqslant 0.62 c_0^{-1} \kappa^2 \kappa_{\text{RSC}}^{-2} e_{\text{stat}}^2.$$

Hence the inequality at (S6) holds.

Finally, we verify (S5). Since $\kappa \geqslant 1$ and $b = \sigma_U^{1/4}$, it holds $E^k \leqslant c_0 \sigma_L^{1/2} \kappa^{-3/2} \leqslant c_0 \sigma_U^{1/2}$ where $c_0 < 0.01$ is a very small number. From the definition of E^k, we can verify that, for $i = 1$ or 2:

$$\begin{aligned}
\|\boldsymbol{U}_i^k\|_{\text{op}} &\leqslant \|\boldsymbol{U}_i^* \boldsymbol{R}_i^k\|_{\text{op}} + \|\boldsymbol{U}_i^k - \boldsymbol{U}_i^* \boldsymbol{R}_i^k\|_{\text{op}} \leqslant \sigma_U^{1/4} + \|\boldsymbol{U}_i^k - \boldsymbol{U}_i^* \boldsymbol{R}_i^k\|_{\text{F}} \leqslant 1.1 \sigma_U^{1/4}, \\
\sigma_{\min}(\boldsymbol{U}_i^k) &\geqslant \|\boldsymbol{U}_i^* \boldsymbol{R}_i^k\|_{\text{op}} - \|\boldsymbol{U}_i^k - \boldsymbol{U}_i^* \boldsymbol{R}_i^k\|_{\text{op}} \geqslant \sigma_U^{1/4} - \|\boldsymbol{U}_i^k - \boldsymbol{U}_i^* \boldsymbol{R}_i^k\|_{\text{F}} \geqslant 0.9 \sigma_U^{1/4}, \text{ and} \\
\|\boldsymbol{\mathcal{G}}_{(i)}^k\|_{\text{op}} &\leqslant \|\boldsymbol{R}_i^k \boldsymbol{\mathcal{G}}_{(i)}^* (\boldsymbol{I}_{T_0} \otimes \boldsymbol{R}_{i-1}^k)'\|_{\text{op}} + \|\boldsymbol{\mathcal{G}}_{(i)}^k - \boldsymbol{R}_i^k \boldsymbol{\mathcal{G}}_{(i)}^* (\boldsymbol{I}_{T_0} \otimes \boldsymbol{R}_{i-1}^k)'\|_{\text{op}} \\
&\leqslant \sigma_U^{1/2} + \sqrt{c_0} \sigma_U^{1/4} \leqslant 1.1 \sigma_U^{1/2}.
\end{aligned}$$

We hence accomplish the whole proof.

S4.3 Proof of Corollary 2

Note that, by Assumption 2 and the low-rank conditions at (2.4), $\|\boldsymbol{A}_j^*\|_{\mathrm{F}} \leqslant C\sqrt{r_1 \wedge r_2}\rho^j$ for $j \geqslant 1$. Let Q_γ be the smallest integer such that $C\sqrt{r_1 \wedge r_2}\rho^j \leqslant \gamma$ for all $j \geqslant Q_\gamma$, and it can be verified that $Q_\gamma = \lceil \log(C\sqrt{r_1 \wedge r_2}/\gamma)/\log(1/\rho) \rceil$ and $\tau^2 Q_\gamma \lesssim 1$. As a result, by a method similar to (S14), we can show that

$$\|\boldsymbol{\mathcal{A}}_{S_\gamma^c}^*\|_{\mathrm{F}}^2 \lesssim \gamma^2 Q_\gamma \ \text{ and } \ \tau^2 \|\boldsymbol{\mathcal{A}}_{S_\gamma^c}^*\|_{\ddagger}^2 \lesssim \gamma^2 Q_\gamma. \tag{S36}$$

Moreover, by choosing $\gamma \asymp \sqrt{\{(r_1 \wedge r_2)N + \log T_0\}/T_1}$, we have $Q_\gamma \asymp \log T_1/\log(1/\rho)$. Let $s \asymp Q_\gamma \asymp \log T_1/\log(1/\rho)$, and it is implied by Theorem 3 that

$$e_{\text{stat}}^2 \lesssim \frac{[(r_1 \wedge r_2)N + \log T_0]s}{T_1}. \tag{S37}$$

On the other hand, by plugging the values of γ and Q_γ into the first term in (S36), we have $\|\boldsymbol{\mathcal{A}}_{S_\gamma^c}^*\|_{\mathrm{F}}^2 \leqslant C e_{\text{stat}}^2$ for an absolute constant C. Note that $\|\boldsymbol{\mathcal{A}}^K - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 \leqslant 2\|\boldsymbol{\mathcal{A}}^K - \boldsymbol{\mathcal{A}}_{S_\gamma}^*\|_{\mathrm{F}}^2 + 2\|\boldsymbol{\mathcal{A}}_{S_\gamma^c}^*\|_{\mathrm{F}}^2$ and, from Theorem 4,

$$\|\boldsymbol{\mathcal{A}}^K - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 \leqslant 2D_1 (1 - D_2)^K \|\boldsymbol{\mathcal{A}}^0 - \boldsymbol{\mathcal{A}}_{S_\gamma}^*\|_{\mathrm{F}}^2 + 2(C + D_3)e_{\text{stat}}^2,$$

where $D_1 = \kappa^{3/2}\sigma_L^{-1/2}$, $D_2 = \eta_0^2\delta^2$, $D_3 = \kappa^{7/2}\sigma_L^{-1/2}\kappa_{\text{RSC}}^{-2}\eta_0^{-2}\delta^{-2}$, and $C \lesssim D_3$.

As a result, when

$$K \geqslant \frac{\log(2D_3) + \log e_{\text{stat}}^2 - \log D_1 - \log\|\boldsymbol{\mathcal{A}}^0 - \boldsymbol{\mathcal{A}}_{S_\gamma}^*\|_{\mathrm{F}}^2}{\log(1 - D_2)}, \tag{S38}$$

the optimization error can be shown to be dominated by the statistical error, i.e.

$$\|\boldsymbol{\mathcal{A}}^K - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 \lesssim \frac{[(r_1 \wedge r_2)N + \log T_0]s}{T_1}.$$

Moreover, from (S37), $\log e_{\text{stat}}^2$ can be upper-bounded by some absolute positive constant when $T_1 \gtrsim \{(r_1 \wedge r_2) + s^2\}N + s^2 \log T_0$. Recall that the absolute constant $\eta_0 \leqslant 1/150$ and $\|\boldsymbol{\mathcal{A}}^0 - \boldsymbol{\mathcal{A}}_{S_\gamma}^*\|_{\mathrm{F}}^2 \lesssim \sigma_L^{5/2}\kappa^{-3/2}$, and the bound at (S38) can be further simplified into

$$K \gtrsim \frac{\log(\kappa^{7/2}\sigma_L^{-5/2}\kappa_{\text{RSC}}^{-2}\delta^{-2})}{\log(1 - \eta_0^2\delta^2)}.$$

Hence, the proof of this corollary is accomplished.

S4.4 Five auxiliary lemmas

This subsection gives five auxiliary lemmas used in the proof of Theorem 4.

Lemma C.1 (Restricted strong convexity and smoothness conditions). *Suppose that Assumptions 1–4 are satisfied. If $T_1 \gtrsim s^2(N + \log T_0)$, then for any $\boldsymbol{\mathcal{A}}_1, \boldsymbol{\mathcal{A}}_2 \in \boldsymbol{\Theta}^{\mathrm{SP}}(r_1, r_2, s)$,*

$$0.5\kappa_{\mathrm{RSC}}\|\boldsymbol{\mathcal{A}}_1 - \boldsymbol{\mathcal{A}}_2\|_{\mathrm{F}}^2 \leqslant \mathcal{L}(\boldsymbol{\mathcal{A}}_1) - \mathcal{L}(\boldsymbol{\mathcal{A}}_2) - \langle \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}_2), \boldsymbol{\mathcal{A}}_1 - \boldsymbol{\mathcal{A}}_2\rangle \leqslant 1.5\kappa_{\mathrm{RSS}}\|\boldsymbol{\mathcal{A}}_1 - \boldsymbol{\mathcal{A}}_2\|_{\mathrm{F}}^2$$

holds with probability at least $1 - Ce^{-N-\log T_0}$, where κ_{RSC} and κ_{RSS} are defined in Theorem 1.

Proof. Recall that

$$\mathcal{L}(\boldsymbol{\mathcal{A}}_1) - \mathcal{L}(\boldsymbol{\mathcal{A}}_2) - \langle \nabla\mathcal{L}(\boldsymbol{\mathcal{A}}_2), \boldsymbol{\mathcal{A}}_1 - \boldsymbol{\mathcal{A}}_2\rangle = \frac{1}{2T_1}\|(\boldsymbol{\mathcal{A}}_1 - \boldsymbol{\mathcal{A}}_2)_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2.$$

Let $\boldsymbol{\Delta} = \boldsymbol{\mathcal{A}}_1 - \boldsymbol{\mathcal{A}}_2$, and it holds that $\boldsymbol{\Delta} \in \boldsymbol{\Theta}^{\mathrm{SP}}(2r_1, 2r_2, 2s)$. Using this notation and ignoring the constant scaling, we have the inequalities

$$\frac{1}{T_1}\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2 \leqslant \frac{\mathbb{E}\left(\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2\right)}{T_1} + \frac{\left|\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2 - \mathbb{E}\left(\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2\right)\right|}{T_1}, \tag{S39}$$

and

$$\frac{1}{T_1}\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2 \geqslant \frac{\mathbb{E}\left(\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2\right)}{T_1} - \frac{\left|\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2 - \mathbb{E}\left(\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2\right)\right|}{T_1}. \tag{S40}$$

First, by (S34), Basu and Michailidis (2015) and Assumptions 2 & 3, we have

$$\kappa_{\mathrm{RSC}}\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2 \leqslant \frac{\mathbb{E}\left(\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2\right)}{T_1} = \mathrm{tr}\left(\boldsymbol{\Delta}_{(1)}\boldsymbol{\Sigma}_{T_0}\boldsymbol{\Delta}_{(1)}'\right) \leqslant \lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon)\mu_{\max}(\boldsymbol{\Psi}_*) \leqslant \kappa_{\mathrm{RSS}}\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2.$$

Moreover, by a method similar to the proof of Lemma B.3,

$$\begin{aligned}
\frac{\left|\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2 - \mathbb{E}\left(\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2\right)\right|}{T_1} &= \left|\mathrm{tr}\left\{\boldsymbol{\Delta}_{(1)}\left(\frac{\boldsymbol{X}\boldsymbol{X}'}{T_1} - \boldsymbol{\Sigma}_{T_0}\right)\boldsymbol{\Delta}_{(1)}'\right\}\right| \\
&\leqslant \sum_{i=1}^{T_0}\sum_{j=1}^{T_0}\left|\mathrm{tr}\left[\boldsymbol{\Delta}_i\left\{\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1} - \boldsymbol{\Gamma}(i-j)\right\}\boldsymbol{\Delta}_j'\right]\right| \\
&\leqslant \sum_{i=1}^{T_0}\sum_{j=1}^{T_0}\|\boldsymbol{\Delta}_i\|_{\mathrm{F}}\|\boldsymbol{\Delta}_j\|_{\mathrm{F}}\left\|\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1} - \boldsymbol{\Gamma}(i-j)\right\|_{\mathrm{op}} \\
&\leqslant \|\boldsymbol{\Delta}\|_{\ddagger}^2 \max_{1\leqslant i\leqslant T_0}\max_{1\leqslant j\leqslant T_0}\left\|\frac{\boldsymbol{X}_i\boldsymbol{X}_j'}{T_1} - \boldsymbol{\Gamma}(i-j)\right\|_{\mathrm{op}}.
\end{aligned}$$

Note that $\|\boldsymbol{\Delta}\|_{\ddagger} \leqslant \sqrt{2s}\|\boldsymbol{\Delta}\|_{\mathrm{F}}$ and then, by Lemma B.6, it can be verified that, when $T_1 \gtrsim s^2(N + \log T_0)$,

$$\frac{\left|\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2 - \mathbb{E}\left(\|\boldsymbol{\Delta}_{(1)}\boldsymbol{X}\|_{\mathrm{F}}^2\right)\right|}{T_1} \leqslant s\tau^2\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2 \leqslant 0.5\kappa_{\mathrm{RSC}}\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2$$

holds with probability at least $1 - Ce^{-N-\log T_0}$, where $\tau^2 = C\sqrt{(N + \log T_0)/T_1}$. This, together with (S39) and (S40), accomplishes the proof of this lemma. □

Lemma C.2. *Consider two tensors $\boldsymbol{\mathcal{A}}^* = \boldsymbol{\mathcal{G}}^* \times_1 \boldsymbol{U}_1^* \times_2 \boldsymbol{U}_2^* \in \mathbb{R}^{N\times N\times T_0}$ and $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \in \mathbb{R}^{N\times N\times T_0}$, where $\boldsymbol{\mathcal{G}}^*$ and $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{r_1\times r_2\times T_0}$ are core tensors, and $\boldsymbol{U}_i^*, \boldsymbol{U}_i \in \mathbb{R}^{N\times r_i}$ with $i = 1$ and 2 are factor matrices. We define their distance below,*

$$E = \min_{\boldsymbol{R}_i \in \mathcal{O}^{r_i\times r_i},\, 1\leqslant i\leqslant 2} \left\{\sum_{i=1}^{2}\|\boldsymbol{U}_i - \boldsymbol{U}_i^*\boldsymbol{R}_i\|_{\mathrm{F}}^2 + \|\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{G}}^* \times_1 \boldsymbol{R}_1' \times_2 \boldsymbol{R}_2'\|_{\mathrm{F}}^2\right\}.$$

Suppose that, for $i = 1$ and 2, $\|\boldsymbol{U}_i\|_{\mathrm{op}} \leqslant (1 + c_e)\sigma_U^{1/4}$, $\|\boldsymbol{\mathcal{G}}_{(i)}\|_{\mathrm{op}} \leqslant (1 + c_e)\sigma_U^{1/2}$, $\boldsymbol{U}_i^{*\prime}\boldsymbol{U}_i^* = b^2\boldsymbol{I}_{r_i}$, and $\sigma_L \leqslant \sigma_{\min}(\boldsymbol{\mathcal{A}}_{(i)}^*) \leqslant \|\boldsymbol{\mathcal{A}}_{(i)}^*\|_{\mathrm{op}} \leqslant \sigma_U$, where $c_e > 0$, $b > 0$, $0 < \sigma_L \leqslant \sigma_U$, and $\sigma_{\min}(\boldsymbol{A})$ denotes the smallest nonzero singular value of matrix $\boldsymbol{A}$. It then holds that

$$c_3\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 \leqslant E \leqslant c_2\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 + 2(1 + c_1)b^{-2}\sum_{i=1}^{2}\|\boldsymbol{U}_i'\boldsymbol{U}_i - b^2\boldsymbol{I}_{r_i}\|_{\mathrm{F}}^2,$$

where $c_1 = 3(1 + c_e)^4\sigma_U^{3/2}b^{-4}$, $c_2 = 3b^{-4} + 8(1 + c_1)\sigma_L^{-2}b^{-2}$, and $c_3 = [3(1 + c_e)^4(\sigma_U + 2\sigma_U^{3/2})]^{-1}$.

Proof. We first prove the upper bound. For any $\boldsymbol{R}_i \in \mathcal{O}^{r_i\times r_i}$ with $i = 1$ and 2,

$$\|\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{G}}^* \times_1 \boldsymbol{R}_1' \times_2 \boldsymbol{R}_2'\|_{\mathrm{F}}^2 = b^{-4}\|\boldsymbol{\mathcal{G}} \times_1 \boldsymbol{U}_1^*\boldsymbol{R}_1 \times_2 \boldsymbol{U}_2^*\boldsymbol{R}_2 - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2,$$

and, by the fact that $(x + y + z)^2 \leqslant 3x^2 + 3y^2 + 3z^2$,

$$\begin{aligned}
&\|\boldsymbol{\mathcal{G}} \times_1 \boldsymbol{U}_1^*\boldsymbol{R}_1 \times_2 \boldsymbol{U}_2^*\boldsymbol{R}_2 - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 \\
&\quad\leqslant 3\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 + 3\|\boldsymbol{\mathcal{G}} \times_1 (\boldsymbol{U}_1 - \boldsymbol{U}_1^*\boldsymbol{R}_1) \times_2 \boldsymbol{U}_2\|_{\mathrm{F}}^2 + 3\|\boldsymbol{\mathcal{G}} \times_1 \boldsymbol{U}_1^*\boldsymbol{R}_1 \times_2 (\boldsymbol{U}_2 - \boldsymbol{U}_2^*\boldsymbol{R}_2)\|_{\mathrm{F}}^2 \\
&\quad\leqslant 3\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 + 3(1 + c_e)^4\sigma_U^{3/2}\sum_{i=1}^{2}\|\boldsymbol{U}_i - \boldsymbol{U}_i^*\boldsymbol{R}_i\|_{\mathrm{F}}^2.
\end{aligned}$$

As a result,

$$\|\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{G}}^* \times_1 \boldsymbol{R}_1' \times_2 \boldsymbol{R}_2'\|_{\mathrm{F}}^2 \leqslant 3b^{-4}\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 + c_1\sum_{i=1}^{2}\|\boldsymbol{U}_i - \boldsymbol{U}_i^*\boldsymbol{R}_i\|_{\mathrm{F}}^2, \tag{S41}$$

with $c_1 = 3(1+c_e)^4\sigma_U^{3/2}b^{-4}$, which implies that

$$E \leqslant 3b^{-4}\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 + (1+c_1)\sum_{i=1}^{2} \min_{\boldsymbol{R}_i \in \mathcal{O}^{r_i\times r_i}} \|\boldsymbol{U}_i - \boldsymbol{U}_i^*\boldsymbol{R}_i\|_{\mathrm{F}}^2. \tag{S42}$$

We next handle the second term of (S42). For $i=1$ and 2, consider an SVD form $\boldsymbol{U}_i = \bar{\boldsymbol{U}}_i\bar{\boldsymbol{\Sigma}}_i\bar{\boldsymbol{V}}_i'$. Note that $\|\boldsymbol{U}_i - \boldsymbol{U}_i^*\boldsymbol{R}_i\|_{\mathrm{F}}^2 \leqslant 2\|\boldsymbol{U}_i - b\bar{\boldsymbol{U}}_i\bar{\boldsymbol{V}}_i'\|_{\mathrm{F}}^2 + 2\|b\bar{\boldsymbol{U}}_i\bar{\boldsymbol{V}}_i' - \boldsymbol{U}_i^*\boldsymbol{R}_i\|_{\mathrm{F}}^2$, and then

$$\min_{\boldsymbol{R}_i \in \mathcal{O}^{r_i\times r_i}} \|\boldsymbol{U}_i - \boldsymbol{U}_i^*\boldsymbol{R}_i\|_{\mathrm{F}}^2 \leqslant 2\|\bar{\boldsymbol{\Sigma}}_i - b\boldsymbol{I}_{r_i}\|_{\mathrm{F}}^2 + 2\min_{\boldsymbol{R}_i \in \mathcal{O}^{r_i\times r_i}} \|b\bar{\boldsymbol{U}}_i - \boldsymbol{U}_i^*\boldsymbol{R}_i\|_{\mathrm{F}}^2, \tag{S43}$$

where $\|\bar{\boldsymbol{\Sigma}}_i - b\boldsymbol{I}_{r_i}\|_{\mathrm{F}}^2 \leqslant b^{-2}\|\boldsymbol{U}_i'\boldsymbol{U}_i - b^2\boldsymbol{I}_{r_i}\|_{\mathrm{F}}^2$; see (E.3) in Han et al. (2022). On the other hand, $\bar{\boldsymbol{U}}_i$ and $b^{-1}\boldsymbol{U}_i^*$ have orthonormal columns and they span the left singular subspaces of $\boldsymbol{\mathcal{A}}_{(i)}$ and $\boldsymbol{\mathcal{A}}_{(i)}^*$, respectively. Then, from Lemma 1 in Cai and Zhang (2018),

$$\min_{\boldsymbol{R}_i \in \mathcal{O}^{r_i\times r_i}} \|b\bar{\boldsymbol{U}}_i - \boldsymbol{U}_i^*\boldsymbol{R}_i\|_{\mathrm{F}}^2 \leqslant 2b^2\|\bar{\boldsymbol{U}}_{i\perp}'(b^{-1}\boldsymbol{U}_i^*)\|_{\mathrm{F}}^2 \leqslant 2b^{-2}\sigma_L^{-2}\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2, \tag{S44}$$

where $\bar{\boldsymbol{U}}_{i\perp} \in \mathcal{O}^{N\times(N-r_i)}$ lies in the orthogonal complementary subspace of $\bar{\boldsymbol{U}}_i$ and the last inequality is due to

$$\begin{aligned}\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 &= \|\boldsymbol{\mathcal{A}}_{(i)} - \boldsymbol{\mathcal{A}}_{(i)}^*\|_{\mathrm{F}}^2 \geqslant \|\bar{\boldsymbol{U}}_{i\perp}'\boldsymbol{\mathcal{A}}_{(i)}^*\|_{\mathrm{F}}^2 = b^4\|\bar{\boldsymbol{U}}_{i\perp}'(b^{-1}\boldsymbol{U}_i^*)(b^{-1}\boldsymbol{U}_i^{*\prime})\boldsymbol{\mathcal{A}}_{(i)}^*\|_{\mathrm{F}}^2\\ &\geqslant b^4\sigma_L^2\|\bar{\boldsymbol{U}}_{i\perp}'(b^{-1}\boldsymbol{U}_i^*)\|_{\mathrm{F}}^2.\end{aligned}$$

From (S43) and (S44), we have

$$\min_{\boldsymbol{R}_i \in \mathcal{O}^{r_i\times r_i}} \|\boldsymbol{U}_i - \boldsymbol{U}_i^*\boldsymbol{R}_i\|_{\mathrm{F}}^2 \leqslant 2b^{-2}\|\boldsymbol{U}_i'\boldsymbol{U}_i - b^2\boldsymbol{I}_{r_i}\|_{\mathrm{F}}^2 + 4b^{-2}\sigma_L^{-2}\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2,$$

which, together with (S42), implies that

$$E \leqslant c_2\|\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 + 2(1+c_1)b^{-2}\sum_{i=1}^{2}\|\boldsymbol{U}_i'\boldsymbol{U}_i - b^2\boldsymbol{I}_{r_i}\|_{\mathrm{F}}^2,$$

where $c_2 = 3b^{-4} + 8(1+c_1)\sigma_L^{-2}b^{-2}$.

We next prove the lower bound. Note that $\|\boldsymbol{\mathcal{G}}_{(i)}^*\|_{\mathrm{op}} = b^{-2}\|\boldsymbol{\mathcal{A}}_{(i)}^*\|_{\mathrm{op}} \leqslant \sigma_U b^{-2}$ for $1 \leqslant i \leqslant 2$, and then it holds that

$$
\begin{aligned}
\|\boldsymbol{\mathcal{A}}-\boldsymbol{\mathcal{A}}^*\|_{\mathrm{F}}^2 \leqslant\, & 3\|(\boldsymbol{\mathcal{G}}-\boldsymbol{\mathcal{G}}^* \times_1 \boldsymbol{R}_1' \times_2 \boldsymbol{R}_2') \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2\|_{\mathrm{F}}^2 \\
& + 3\|(\boldsymbol{\mathcal{G}}^* \times_1 \boldsymbol{R}_1' \times_2 \boldsymbol{R}_2') \times_1 (\boldsymbol{U}_1-\boldsymbol{U}_1^*\boldsymbol{R}_1) \times_2 \boldsymbol{U}_2\|_{\mathrm{F}}^2 \\
& + 3\|(\boldsymbol{\mathcal{G}}^* \times_2 \boldsymbol{R}_2') \times_1 \boldsymbol{U}_1^* \times_2 (\boldsymbol{U}_2-\boldsymbol{U}_2^*\boldsymbol{R}_2)\|_{\mathrm{F}}^2 \\
& \leqslant c_3^{-1} E,
\end{aligned}
$$

where $c_3 = [3(1+c_e)^4(\sigma_U + 2\sigma_U^{3/2})]^{-1}$. □

Lemma C.3. *Consider the two tensors,* $\boldsymbol{\mathcal{A}}^* = \boldsymbol{\mathcal{G}}^* \times_1 \boldsymbol{U}_1^* \times_2 \boldsymbol{U}_2^* \in \mathbb{R}^{N\times N\times T_0}$ *and* $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \in \mathbb{R}^{N\times N\times T_0}$*, in Lemma C.2, and it holds that, for* $i=1$ *and 2,* $\|\boldsymbol{U}_i\|_{\mathrm{op}} \leqslant (1+c_e)\sigma_U^{1/4}$*,* $\|\boldsymbol{\mathcal{G}}_{(i)}\|_{\mathrm{op}} \leqslant (1+c_e)\sigma_U^{1/2}$*,* $\boldsymbol{U}_i^{*\prime}\boldsymbol{U}_i^* = b^2\boldsymbol{I}_{r_i}$*, and* $\|\boldsymbol{\mathcal{A}}_{(i)}^*\|_{\mathrm{op}} \leqslant \sigma_U$*, where* $c_e>0$*,* $b>0$*, and* $0<\sigma_U$*. Define three tensors below,*

$$
\boldsymbol{\mathcal{A}}_{\mathcal{G}} = (\boldsymbol{\mathcal{G}}-\boldsymbol{\mathcal{G}}^* \times_1 \boldsymbol{R}_1' \times_2 \boldsymbol{R}_2') \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2, \quad \boldsymbol{\mathcal{A}}_{U_1} = \boldsymbol{\mathcal{G}} \times_1 (\boldsymbol{U}_1-\boldsymbol{U}_1^*\boldsymbol{R}_1) \times_2 \boldsymbol{U}_2,
$$

and $\boldsymbol{\mathcal{A}}_{U_2} = \boldsymbol{\mathcal{A}}_{U_1} = \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{U}_1 \times_2 (\boldsymbol{U}_2-\boldsymbol{U}_2^*\boldsymbol{R}_2)$*. If there exists a* $B\geqslant 0$ *such that* $\|\boldsymbol{\mathcal{G}}-\boldsymbol{\mathcal{G}}^* \times_1 \boldsymbol{R}_1' \times_2 \boldsymbol{R}_2'\|_{\mathrm{F}} \leqslant B$ *and* $\|\boldsymbol{U}_i-\boldsymbol{U}_i^*\boldsymbol{R}_i\|_{\mathrm{F}} \leqslant B$ *for* $\boldsymbol{R}_i \in \mathcal{O}^{r_i\times r_i}$ *with* $i=1$ *and 2, then*

$$
\|\boldsymbol{\mathcal{H}}_\epsilon\|_{\mathrm{F}} \leqslant [(1+c_e)\sigma_U^{1/2} + (2+c_e)\sigma_U^{1/4}]B^2 \ \text{ with } \boldsymbol{\mathcal{H}}_\epsilon = \boldsymbol{\mathcal{A}}_{\mathcal{G}} + \sum_{i=1}^{2}\boldsymbol{\mathcal{A}}_{U_i} - (\boldsymbol{\mathcal{A}}-\boldsymbol{\mathcal{A}}^*).
$$

Proof. Note that

$$
\boldsymbol{\mathcal{A}}_{\mathcal{G}} + \sum_{i=1}^{2}\boldsymbol{\mathcal{A}}_{U_i} = \boldsymbol{\mathcal{A}}-\boldsymbol{\mathcal{A}}^* + \underbrace{\boldsymbol{\mathcal{H}}_\epsilon^{(1)} + \boldsymbol{\mathcal{H}}_\epsilon^{(2)} + \boldsymbol{\mathcal{H}}_\epsilon^{(3)}}_{=\boldsymbol{\mathcal{H}}_\epsilon},
$$

where

$$
\begin{aligned}
\boldsymbol{\mathcal{H}}_\epsilon^{(1)} &= \boldsymbol{\mathcal{G}} \times_1 (\boldsymbol{U}_1-\boldsymbol{U}_1^*\boldsymbol{R}_1) \times_2 (\boldsymbol{U}_2-\boldsymbol{U}_2^*\boldsymbol{R}_2), \\
\boldsymbol{\mathcal{H}}_\epsilon^{(2)} &= (\boldsymbol{\mathcal{G}}-\boldsymbol{\mathcal{G}}^* \times_1 \boldsymbol{R}_1' \times_2 \boldsymbol{R}_2') \times_1 (\boldsymbol{U}_1-\boldsymbol{U}_1^*\boldsymbol{R}_1) \times_2 \boldsymbol{U}_2, \qquad \text{(S45)} \\
\boldsymbol{\mathcal{H}}_\epsilon^{(3)} &= (\boldsymbol{\mathcal{G}}-\boldsymbol{\mathcal{G}}^* \times_1 \boldsymbol{R}_1' \times_2 \boldsymbol{R}_2') \times_1 \boldsymbol{U}_1^*\boldsymbol{R}_1 \times_2 (\boldsymbol{U}_2-\boldsymbol{U}_2^*\boldsymbol{R}_2).
\end{aligned}
$$

It can be easily verified that

$$
\|\boldsymbol{\mathcal{H}}_\epsilon^{(1)}\|_{\mathrm{F}} \leqslant (1+c_e)\sigma_U^{1/2}B^2, \quad \|\boldsymbol{\mathcal{H}}_\epsilon^{(2)}\|_{\mathrm{F}} \leqslant (1+c_e)\sigma_U^{1/4}B^2 \quad \text{and} \quad \|\boldsymbol{\mathcal{H}}_\epsilon^{(1)}\|_{\mathrm{F}} \leqslant \sigma_U^{1/4}B^2.
$$

Hence, the proof of this lemma is accomplished. □

Lemma C.4 (Contractive projection property (CPP)). *Consider a tensor* $\mathcal{X} \in \mathbb{R}^{N\times N\times T_0}$ *with frontal slices* $\{\boldsymbol{X}_i, 1 \leqslant i \leqslant T_0\}$*, and let* $S_1 = \{1 \leqslant j \leqslant T_0, \|\boldsymbol{X}_j\|_{\mathrm{F}} > 0\}$ *be the collection of nonzero frontal slices. Moreover,* $\mathcal{X}^* \in \mathbb{R}^{N\times N\times T_0}$ *is another tensor with* S_2 *being the collection of nonzero frontal slices. Denote by* s_j *the cardinality of* S_j *with* $j = 1$ *and 2. If* $s_2 < s \leqslant s_1$ *and* $S_2 \subset S_1$*, then*

$$\|\mathrm{HT}(\mathcal{X}, \mathrm{s}) - \mathcal{X}\|_{\mathrm{F}}^2 \leqslant \frac{s_1 - s}{s_1 - s_2}\|\mathcal{X} - \mathcal{X}^*\|_{\mathrm{F}}^2.$$

Proof. This lemma is a trivial extension of Lemma 1.1 in Jain et al. (2014) to the case with tensors, and the proof is also similar. □

Lemma C.5. *Let* f *be a continuously differentiable function and, for any tensors* $\mathcal{A}$ *and* $\mathcal{B}$*,*

$$\frac{m}{2}\|\mathcal{A} - \mathcal{B}\|_{\mathrm{F}}^2 \leqslant f(\mathcal{A}) - f(\mathcal{B}) - \langle \nabla f(\mathcal{B}), \mathcal{A} - \mathcal{B}\rangle \leqslant \frac{M}{2}\|\mathcal{A} - \mathcal{B}\|_{\mathrm{F}}^2,$$

where $0 < m \leqslant M < \infty$*. It then holds that*

$$\langle \nabla f(\mathcal{A}) - \nabla f(\mathcal{B}), \mathcal{A} - \mathcal{B}\rangle \geqslant \frac{mM}{m+M}\|\mathcal{A} - \mathcal{B}\|_{\mathrm{F}}^2 + \frac{1}{m+M}\|\nabla f(\mathcal{A}) - \nabla f(\mathcal{B})\|_{\mathrm{F}}^2.$$

Proof. This lemma is from Theorem 2.1.11 in Nesterov (2003) and is provided here to make the proof self-contained. □

S5 Additional details in the simulation studies

In this section, we first present an additional simulation study to compare the sensitivity of the soft- and hard-thresholding algorithm with respect to T_0. Next, we present more details regarding the generation of $\boldsymbol{B}_j$ and $\boldsymbol{C}_j$ in our VAR data generating processes.

S5.1 Sensitivity of soft- and hard-thresholding

This subsection presents two additional simulations to compare the finite-sample performance of the soft- and hard-thresholding methods from Section 3.1 at various values of T and T_0. Specifically, the goal is to demonstrate that the tuning parameter λ in the soft-thresholding algorithm is more sensitive to changes in T and T_0 than the sparsity level s in the hard-thresholding algorithm, thus supporting our preference for the more stable hard-thresholding algorithm.

The first experiment aims to show how the parameter estimation errors vary with respect to T under different λ's and s's while holding T_0 fixed. We generate the data using the VARMA process with $(N, r) = (20, 4)$, and four sample sizes are considered with $T = 400$, 600, 800 and 1000. The running order is fixed at $T_0 = 100$, and the effective sample size is $T_1 = T - T_0$. There are 500 replications for each sample size, and the hard-thresholding method, i.e. Algorithm 1, is first considered to search for estimates with the sparsity level s varying from 4 to 18. Figure 1 gives the parameter estimation errors, averaged over 500 replications, and it can be seen that the optimal sparsity level grows slowly from nine to eleven as the sample size T increases. In the meanwhile, Algorithm 1 modified with the soft-thresholding method in Section 3.1 is also applied, and the tuning parameter λ varies among the values of $\{0.5j \times 10^{-3}, 1 \leqslant j \leqslant 33\}$. The averaged parameter estimation errors are presented in Figure 1, and the optimal values of λ change a lot when the sample size increases from $T = 400$ to 1000. Moreover, since the fitted models at each fixed λ may have different sparsity levels, Figure 1 also plots the sparsity levels of 100 fitted models at the optimal value of λ for each sample size. The sparsity level varies roughly from 20 to 30, and the variation decreases as the sample size increases. Finally, although parameter estimation errors for both methods become larger as sample sizes decrease, those for the soft-thresholding method may vary dramatically, especially when the sample size is as small as $T = 400$.

The second experiment aims to show how the parameter estimation errors vary with respect to T_0 under different λ's and s's while holding T fixed. We fix the sample size at $T = 800$, while four running orders are considered with $T_0 = 25$, 50, 100 and 200. The sparsity level for the hard-thresholding method varies from $s = 5$ to 14, and the tuning parameter for the soft-thresholding method takes the values of $\lambda \in \{0.5j \times 10^{-3}, 4 \leqslant j \leqslant 20\}$. All the other settings are the same as those in the first experiment, and the estimation results are also given in Figure 1. The stability of the hard-thresholding method is confirmed again, while the optimal values of λ are more sensitive to the change in T_0 for the soft-thresholding method. In fact, when the running order is as large as

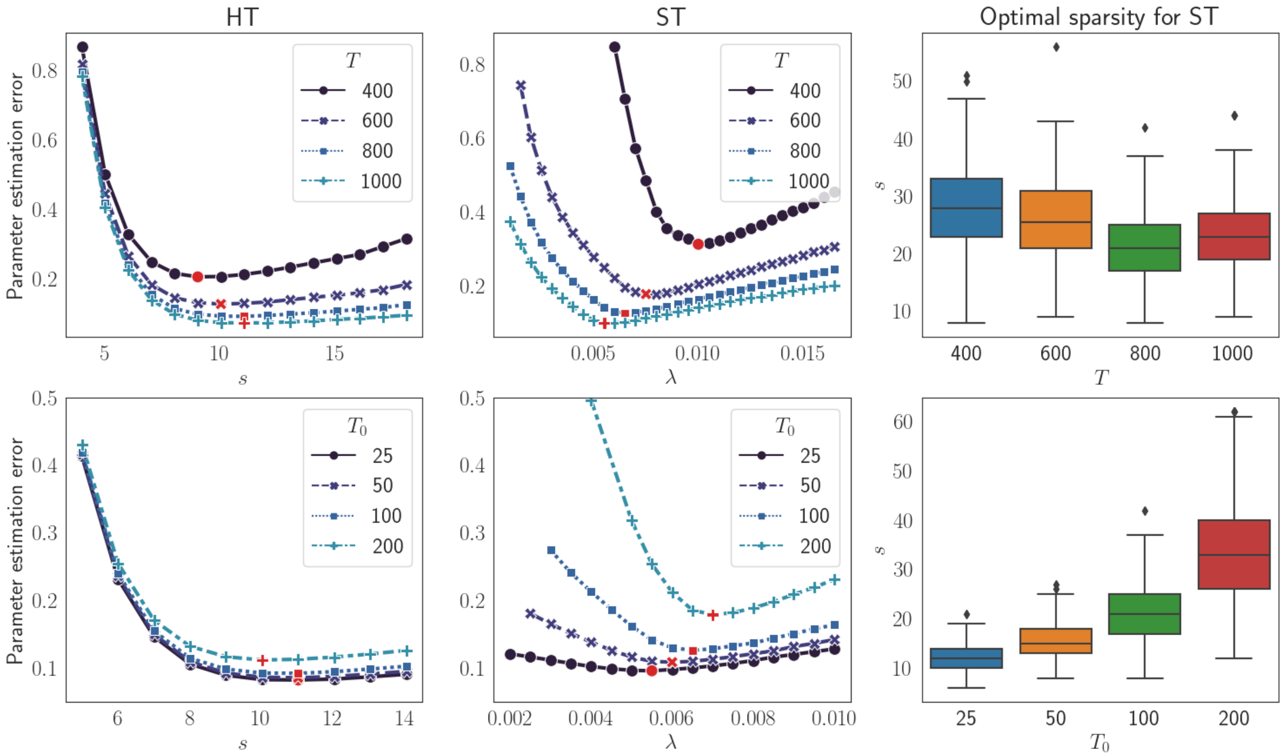

Figure 1: Plots of parameter estimation errors against sparsity levels s (left panel) for the hard-thresholding (HT) method and tuning parameters λ (middle panel) for the soft-thresholding (ST) method, where the optimal setting on each curve with the minimum error is highlighted in red, and boxplots (right panel) for sparsity levels of 100 fitted models at the optimal values of λ. Two cases are considered: varying sample size T but fixed running order T_0 (upper panel) and varying running order T_0 but fixed sample size T (lower panel).

$T_0 = 200$, the parameter estimation errors from the soft-thresholding method vary dramatically, and we can even observe a much higher variation in sparsity levels of the fitted models at the optimal values of λ. This further undermines the stability of the soft-thresholding method.

S5.2 Generation of $\boldsymbol{B}_j$ and $\boldsymbol{C}_j$

We illustate with a simple case where $r_2 < r_1$ and subsequently $r^{(4)} = r_1 - r_2$ and $r^{(j)} = r_2$ for $j \in \{1, 5, 8, 9\}$. The generation details of $\boldsymbol{B}_j$'s and $\boldsymbol{C}_j$'s are as follows. We first generate an $N \times N$ orthonormal matrix $\boldsymbol{O} = (\boldsymbol{o}_1, \ldots, \boldsymbol{o}_N)$ using python function *scipy.stats.ortho_group.rvs(N)*.

Then we set $\boldsymbol{B}_1 = (\boldsymbol{o}_1, \ldots, \boldsymbol{o}_{r_2})$, $\boldsymbol{B}_4 = (\boldsymbol{o}_{r_2+1}, \ldots, \boldsymbol{o}_{r_1})$ and randomly draw with replacement r_2 columns from the set $\{\boldsymbol{o}_1, \ldots, \boldsymbol{o}_{r_1}\}$ to form $\boldsymbol{B}_5, \boldsymbol{B}_8, \boldsymbol{B}_9 \in \mathbb{R}^{N \times r_2}$, respectively. Meanwhile, we generate another $N \times N$ orthogonal matrix $\widetilde{\boldsymbol{O}} = (\widetilde{\boldsymbol{o}}_1, \ldots, \widetilde{\boldsymbol{o}}_N)$ and set $\boldsymbol{C}_1 = (\widetilde{\boldsymbol{o}}_1, \ldots, \widetilde{\boldsymbol{o}}_{r_2})$. Then we randomly sample $r_1 - r_2$ columns from $\boldsymbol{C}_1$ without replacement to form $\boldsymbol{C}_4$ and r_2 columns from $\boldsymbol{C}_1$ with replacement to form $\boldsymbol{C}_5, \boldsymbol{C}_8, \boldsymbol{C}_9 \in \mathbb{R}^{N \times r_2}$, respectively.

S6 Additional information on the macro-economic dataset

In Tables 2–9, we provide more detailed descriptions of the variables of the macro-economic dataset in Section 6 and their transformations.

Figure 2 plots projection matrices of the estimated response loading, i.e. $\widehat{\boldsymbol{U}}_1 \widehat{\boldsymbol{U}}_1^T$, and the predictor loadings, $\widehat{\boldsymbol{U}}_2 \widehat{\boldsymbol{U}}_2^T$, for the small size dataset. We reorder the variables the same way ass described in Section 6 of the main paper. It can be observed that the loading matrices are not only low-rank but also sparse, which can explain why the sparsity-based methods also achieve good forecasting performance. Moreover, the loading matrices for the predictor and response factors differ from each other. The predictor factor summarizes the dynamics from CPI of all items (CPIAUCSL), all items less food&energy (CPIFESL) and apparel (CPIAPPSL), while the response factor mainly depicts how the CPI of all items responds to the changes in the predictor variables.

Figure 3 plots projection matrices of the estimated response loading, i.e. $\widehat{\boldsymbol{U}}_1 \widehat{\boldsymbol{U}}_1^T$, and the predictor loadings, $\widehat{\boldsymbol{U}}_2 \widehat{\boldsymbol{U}}_2^T$, for the medium size dataset. We reorder the variables the same way ass described in Section 6 of the main paper. Again, the loading matrices for the predictor and response factors differ from each other. The predictor factor mainly summarizes the personal consumption expenditure, overall goods prices or durable goods prices, and producer price indices. In the meantime, the response factor additionally captures how durable goods and consumer price indices respond to the changes in the predictor variables.

Table 1: Variable Descriptions and FRED MNEMONICS

Short Name	Description	FRED MNEMONIC
PCED	Personal Consumption Expenditures: Chain-type Price Index (Index 2017=100)	PCEPI
GPDI_Defl	Gross Private Domestic Investment: Chain-type Price Index (Index 2017=100)	GPDIDEF
GS1_TB3M	1-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market (Percent)	GS1TB3M
Nonborrowed_Reserves_Depository	Reserves Of Depository Institutions, Nonborrowed (Millions of Dollars)	NONBORRES
PPI_FinConsGds_Food	Producer Price Index by Commodity for Finished Consumer Foods (Index 1982=100)	PPIFCG
PCED_OtherServices	Personal consumption expenditures: Other services (chain-type price index)	PCEDOTHERSERV
Total_Reserves_Depository	Total Reserves of Depository Institutions (Billions of Dollars)	TOTRESNS
Real_Price_Oil	Producer Price Index by Commodity for Fuels and Related Products and Power: Crude Petroleum (Domestic Production) (Index 1982=100)	PPIORLY
BAA_GS10	Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity (Percent)	BAAGS10
PPI_Crude_Materials	Producer Price Index: Crude Materials for Further Processing (Index 1982-84=100)	PPICRMAT
PCED_FIRE	Personal consumption expenditures: Financial services and insurance (chain-type price index)	PCEDFIRE
Fed_Funds	Effective Federal Funds Rate (Percent)	FEDFUNDS
PPI_Metals_Nonferrous	Producer Price Index: Commodities: Metals and metal products: Primary nonferrous metals (Index 1982=100)	PPINMET
Nikkei_Stock_Avg	Nikkei Stock Average	NIKKEI225
Real_NonRevCredit	Total Real Nonrevolving Credit Owned and Securitized, Outstanding (Billions of Dollars), deflated by Core PCE	NONREVSL
Real_LoansRE	Real Real Estate Loans, All Commercial Banks (Billions of 2017 U.S. Dollars), deflated by Core PCE	REALLNS
TM_3M_FedFunds	3-Month Treasury Constant Maturity Minus Federal Funds Rate	TM3MFED
Real_HHW_RESA	Real Real Estate Assets of Households and Nonprofit Organizations (Billions of 2017 Dollars), deflated by Core PCE	HHWRENSA
Real_AHE_MFG	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing (2017 Dollars per Hour), deflated by Core PCE	AHEMAN
Real_ConsLoans	Real Consumer Loans at All Commercial Banks (Billions of 2017 U.S. Dollars), deflated by Core PCE	CONLOANSNSA
Real_Nonfin_NCorp_NW	Real Nonfinancial Noncorporate Business Sector Net Worth (Billions of 2017 Dollars), Deflated by Implicit Price Deflator for Business Sector	NFCNWNSA
Real_Nonfin_NCorp_Assets	Real Nonfinancial Noncorporate Business Sector Assets (Billions of 2017 Dollars), Deflated by Implicit Price Deflator for Business Sector	NONCORPNSA
Real_ConsuCred	Total Consumer Credit Outstanding, deflated by Core PCE	TOTALSL
Cons_Expectations	University of Michigan: Consumer Sentiment (Index 1st Quarter 1966=100)	UMCSENT
Real_NonRevCredit	Total Real Nonrevolving Credit Owned and Securitized, Outstanding (Billions of Dollars), deflated by Core PCE	NONREVSL
Real_Nonfin_NCorp_Assets	Real Nonfinancial Noncorporate Business Sector Assets (Billions of 2017 Dollars), Deflated by Implicit Price Deflator for Business Sector	NONCORPNSA
Real_LoansRE	Real Real Estate Loans, All Commercial Banks (Billions of 2017 U.S. Dollars), deflated by Core PCE	REALLNS
Real_HHW_LiabSA	Real Total Liabilities of Households and Nonprofit Organizations (Billions of 2017 Dollars), deflated by Core PCE	HHWTOTNSA
Real_Nonfin_NCorp_NW	Real Nonfinancial Noncorporate Business Sector Net Worth (Billions of 2017 Dollars), Deflated by Implicit Price Deflator for Business Sector	NFCNWNSA
Real_HHW_RESA	Real Real Estate Assets of Households and Nonprofit Organizations (Billions of 2017 Dollars), deflated by Core PCE	HHWRENSA
Real_Nonfin_NCorp_Liab	Real Nonfinancial Noncorporate Business Sector Liabilities (Billions of 2017 Dollars), Deflated by Implicit Price Deflator for Business Sector	NFCLNSA
Fed_Funds	Effective Federal Funds Rate (Percent)	FEDFUNDS
Real_CL_Loans	Real Commercial and Industrial Loans, All Commercial Banks (Billions of 2017 U.S. Dollars), deflated by Core PCE	BUSLOANS
CPI_Services	Consumer Price Index for All Urban Consumers: Services (Index 1982-84=100)	CUSR0000SAS
CPI	Consumer Price Index for All Urban Consumers: All Items (Index 1982-84=100)	CPIAUCSL
CPI_LessMedCare	Consumer Price Index for All Urban Consumers: All items less medical care (Index 1982-84=100)	CUSR0000SAM
Real_NCorp_Assets	Real Nonfinancial Corporate Business Sector Assets (Billions of 2017 Dollars), Deflated by Implicit Price Deflator for Business Sector	NFCBAA
CPI_LessFood	Consumer Price Index for All Urban Consumers: All Items Less Food (Index 1982-84=100)	CUSR0000SA0L1
Nonrev_CC_PersInc	Nonrevolving consumer credit to Personal Income	NONREVCCPI
TM_6MTH	6-Month Treasury Bill: Secondary Market Rate (Percent)	DGS6MO
GS10_TB3M	10-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market (Percent)	GS10TB3
BAA_GS10	Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity (Percent)	BAAGS10

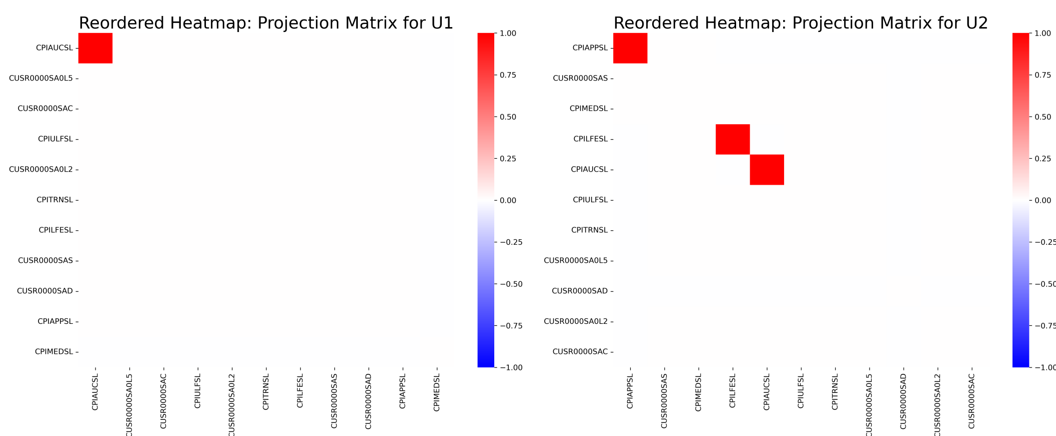

Figure 2: Projection matrices of the estimated response and the predictor loadings for the small size dataset.

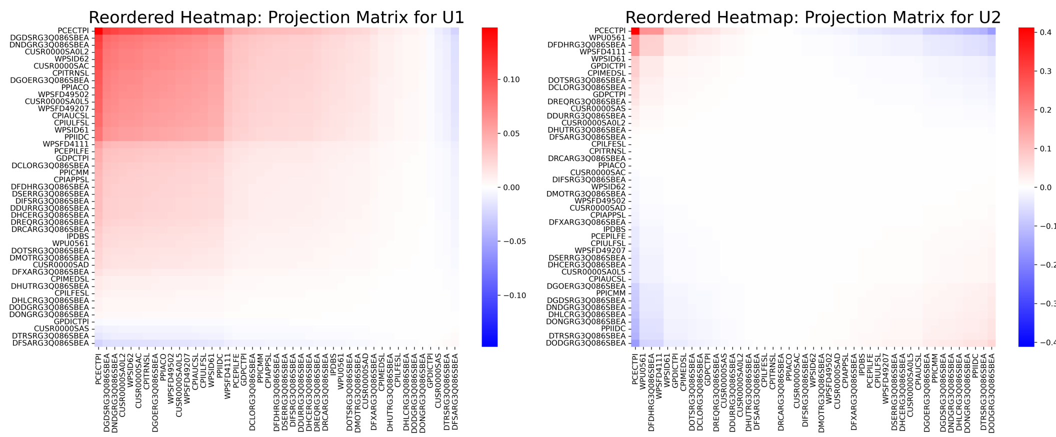

Figure 3: Projection matrices of the estimated response and the predictor loadings for the medium size dataset.

Table 2: Price Indices. FRED MNEMONIC: mnemonic for data in FRED-QD. SW MNEMONIC: mnemonic in Stock and Watson (2012). T: data transformation code, where 6 = second difference of log series. DESCRIPTION: brief definition of the data. G: Group code, where 6 = price index variables.

FRED MNEMONIC	SW MNEMONIC	T	DESCRIPTION	G
IPDBS	BusSec Defl	6	Business Sector: Implicit Price Deflator (Index 2017=100)	6
GDPCTPI	GDP Defl	6	Gross Domestic Product: Chain-type Price Index (Index 2017=100)	6
GPDICTPI	GPDI Defl	6	Gross Private Domestic Investment: Chain-type Price Index (Index 2017=100)	6
PCEPILFE	PCED_LFE	6	Personal Consumption Expenditures Excluding Food and Energy (Chain-Type Price Index) (Index 2017=100)	6
PCECTPI	PCED	6	Personal Consumption Expenditures: Chain-type Price Index (Index 2017=100)	6
DDURRG3Q086SBEA	PCED_DurGoods	6	Personal consumption expenditures: Durable goods (chain-type price index)	6
DFDHRG3Q086SBEA	PCED_DurHousehold	6	Personal consumption expenditures: Durable goods: Furnishings and durable household equipment (chain-type price index)	6
DMOTRG3Q086SBEA	PCED_MotorVec	6	Personal consumption expenditures: Durable goods: Motor vehicles and parts (chain-type price index)	6
DODGRG3Q086SBEA	PCED_OthDurGds	6	Personal consumption expenditures: Durable goods: Other durable goods (chain-type price index)	6
DREQRG3Q086SBEA	PCED_Recreation	6	Personal consumption expenditures: Durable goods: Recreational goods and vehicles (chain-type price index)	6
DIFSRG3Q086SBEA	PCED_FIRE	6	Personal consumption expenditures: Financial services and insurance (chain-type price index)	6
DGDSRG3Q086SBEA	PCED_Goods	6	Personal consumption expenditures: Goods (chain-type price index)	6
DNDGRG3Q086SBEA	PCED_NDurGoods	6	Personal consumption expenditures: Nondurable goods (chain-type price index)	6
DCLORG3Q086SBEA	PCED_Clothing	6	Personal consumption expenditures: Nondurable goods: Clothing and footwear (chain-type price index)	6
DFXARG3Q086SBEA	PCED_Food_Bev	6	Personal consumption expenditures: Nondurable goods: Food and beverages purchased for off-premises consumption (chain-type price index)	6
DGOERG3Q086SBEA	PCED_Gas_Enrgy	6	Personal consumption expenditures: Nondurable goods: Gasoline and other energy goods (chain-type price index)	6
DONGRG3Q086SBEA	PCED_OthNDurGds	6	Personal consumption expenditures: Nondurable goods: Other nondurable goods (chain-type price index)	6
DOTSRG3Q086SBEA	PCED_OtherServices	6	Personal consumption expenditures: Other services (chain-type price index)	6
DRCARG3Q086SBEA	PCED_RecServices	6	Personal consumption expenditures: Recreation services (chain-type price index)	6
DSERRG3Q086SBEA	PCED_Serv	6	Personal consumption expenditures: Services (chain-type price index)	6
DFSARG3Q086SBEA	PCED_FoodServ_Acc.	6	Personal consumption expenditures: Services: Food services and accommodations (chain-type price index)	6
DHLCRG3Q086SBEA	PCED_HealthCare	6	Personal consumption expenditures: Services: Health care (chain-type price index)	6
DHCERG3Q086SBEA	PCED_HouseholdServ.	6	Personal consumption expenditures: Services: Household consumption expenditures (chain-type price index)	6
DHUTRG3Q086SBEA	PCED_Housing-Utilities	6	Personal consumption expenditures: Services: Housing and utilities (chain-type price index)	6
DTRSRG3Q086SBEA	PCED_TransSvg	6	Personal consumption expenditures: Transportation services (chain-type price index)	6
WPSFD4111	PPI:FinConsGds(Food)	6	Producer Price Index by Commodity for Finished Consumer Foods (Index 1982=100)	6
WPSFD49502	PPI:FinConsGds	6	Producer Price Index by Commodity for Finished Consumer Goods (Index 1982=100)	6
WPSFD49207	PPI:FinGds	6	Producer Price Index by Commodity for Finished Goods (Index 1982=100)	6
WPU0561	Real Price:Oil	6	Producer Price Index by Commodity for Fuels and Related Products and Power: Crude Petroleum (Domestic Production) (Index 1982=100)	6
PPIIDC	PPI:IndCom	6	Producer Price Index by Commodity Industrial Commodities (Index 1982=100)	6
WPSID61	PPI:IntMat	6	Producer Price Index by Commodity Intermediate Materials: Supplies & Components (Index 1982=100)	6
PPIACO	PPI	6	Producer Price Index for All Commodities (Index 1982=100)	6
PPICMM		6	Producer Price Index: Commodities: Metals and metal products: Primary nonferrous metals (Index 1982=100)	6
WPSID62		6	Producer Price Index: Crude Materials for Further Processing (Index 1982=100)	6

Table 3: (Price Indices continued) Consumer price index-related variables from the price group. T: data transformation code, where 6 = second difference of log series. G: Group code, where 6 = price index variables.

FRED MNEMONIC	SW MNEMONIC	T	DESCRIPTION	G
CPIAUCSL	CPI	6	Consumer Price Index for All Urban Consumers: All Items (Index 1982-84=100)	6
CPIULFSL		6	Consumer Price Index for All Urban Consumers: All Items Less Food (Index 1982-84=100)	6
CPILFESL	CPI_LFE	6	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy (Index 1982-84=100)	6
CUSR0000SA0L5		6	Consumer Price Index for All Urban Consumers: All items less medical care (Index 1982-84=100)	6
CUSR0000SA0L2		6	Consumer Price Index for All Urban Consumers: All items less shelter (Index 1982-84=100)	6
CPIAPPSL		6	Consumer Price Index for All Urban Consumers: Apparel (Index 1982-84=100)	6
CUSR0000SAC		6	Consumer Price Index for All Urban Consumers: Commodities (Index 1982-84=100)	6
CUSR0000SAD		6	Consumer Price Index for All Urban Consumers: Durables (Index 1982-84=100)	6
CPIMEDSL		6	Consumer Price Index for All Urban Consumers: Medical Care (Index 1982-84=100)	6
CUSR0000SAS		6	Consumer Price Index for All Urban Consumers: Services (Index 1982-84=100)	6
CPITRNSL		6	Consumer Price Index for All Urban Consumers: Transportation (Index 1982-84=100)	6

Table 4: Earnings and productivity variables. T: data transformation code, where 5 = first difference of log series, 6 = second difference of log series. G: Group code, where 7 = earnings and productivity variables.

FRED MNEMONIC	**SW MNEMONIC**	**T**	**DESCRIPTION**	**G**
CES0600000008		6	Average Hourly Earnings of Production and Nonsupervisory Employees: Goods-Producing (Dollars per Hour)	7
RCPHBS	CPH:Bus	5	Business Sector: Real Compensation Per Hour (Index 2017=100)	7
OPHPBS	OPH:Bus	5	Business Sector: Real Output Per Hour of All Persons (Index 2017=100)	7
ULCBS	ULC:Bus	5	Business Sector: Unit Labor Cost (Index 2017=100)	7
COMPRNFB	CPH:NFB	5	Nonfarm Business Sector: Real Compensation Per Hour (Index 2017=100)	7
OPHNFB	OPH:nfb	5	Nonfarm Business Sector: Real Output Per Hour of All Persons (Index 2017=100)	7
ULCNFB	ULC:NFB	5	Nonfarm Business Sector: Unit Labor Cost (Index 2017=100)	7
UNLPNBS	UNLPay:nfb	5	Nonfarm Business Sector: Unit Nonlabor Payments (Index 2017=100)	7
CES2000000008x	Real AHE:Const	5	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Construction (2017 Dollars per Hour), deflated by Core PCE	7
CES3000000008x	Real AHE:MFG	5	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing (2017 Dollars per Hour), deflated by Core PCE	7

Table 5: Interest rates variables. T: data transformation code, where 1 = no transformation, 2 = first difference. G: Group code, where 8 = interest rate variables.

FRED MNEMONIC	SW MNEMONIC	T	DESCRIPTION	G
GS1TB3Mx	GS1_tb3m	1	1-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market (Percent)	8
GS1	TB-1YR	2	1-Year Treasury Constant Maturity Rate (Percent)	8
GS10TB3Mx	GS10_tb3m	1	10-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market (Percent)	8
GS10	TB-10YR	2	10-Year Treasury Constant Maturity Rate (Percent)	8
TB3MS	TB-3Mth	2	3-Month Treasury Bill: Secondary Market Rate (Percent)	8
TB3SMFFM		1	3-Month Treasury Constant Maturity Minus Federal Funds Rate	8
T5YFFM		1	5-Year Treasury Constant Maturity Minus Federal Funds Rate	8
GS5		2	5-Year Treasury Constant Maturity Rate	8
TB6M3Mx	tb6m_tb3m	1	6-Month Treasury Bill Minus 3-Month Treasury Bill, secondary market (Percent)	8
TB6MS	TM-6MTH	2	6-Month Treasury Bill: Secondary Market Rate (Percent)	8
FEDFUNDS	FedFunds	2	Effective Federal Funds Rate (Percent)	8
AAAFFM		1	Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate	8
AAA	AAA Bond	2	Moody's Seasoned Aaa Corporate Bond Yield© (Percent)	8
BAA10YM	BAA_GS10	1	Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity (Percent)	8
BAA	BAA Bond	2	Moody's Seasoned Baa Corporate Bond Yield (Percent)	8

Table 6: Money and credit variables. T: data transformation code, where 5 = first difference of log series, 6 = second difference of log series, 7 = the difference in return series, i.e. $\Delta(x_t/x_{t-1} - 1)$. G: Group code, where 9 = money and credit variables.

FRED MNEMONIC	SW MNEMONIC	T	DESCRIPTION	G
DTCOLNVHFNM		6	Consumer Motor Vehicle Loans Outstanding Owned by Finance Companies (Millions of Dollars)	9
BUSLOANSx	Real C&Lloand	5	Real Commercial and Industrial Loans, All Commercial Banks (Billions of 2017 U.S. Dollars), deflated by Core PCE	9
CONSUMERx	Real ConsLoans	5	Real Consumer Loans at All Commercial Banks (Billions of 2017 U.S. Dollars), deflated by Core PCE	9
M1REAL	Real m1	5	Real M1 Money Stock (Billions of 1982-84 Dollars), deflated by CPI	9
M2REAL	Real m2	5	Real M2 Money Stock (Billions of 1982-84 Dollars), deflated by CPI	9
REALLNx	Real LoansRealEst	5	Real Real Estate Loans, All Commercial Banks (Billions of 2017 U.S. Dollars), deflated by Core PCE	9
NONBORRES		7	Reserves Of Depository Institutions, Nonborrowed (Millions of Dollars)	9
INVEST		6	Securities in Bank Credit at All Commercial Banks (Billions of Dollars)	9
BOGMBASEREALx	Real Mbase	5	St. Louis Adjusted Monetary Base (Billions of 1982-84 Dollars), deflated by CPI	9
TOTALSLx	Real ConsuCred	5	Total Consumer Credit Outstanding, deflated by Core PCE	9
DTCTHFNM		6	Total Consumer Loans and Leases Outstanding Owned and Securitized by Finance Companies (Millions of Dollars)	9
NONREVSLx	Real NonRevCredit	5	Total Real Nonrevolving Credit Owned and Securitized, Outstanding (Billions of Dollars), deflated by Core PCE	9
TOTRESNS		6	Total Reserves of Depository Institutions (Billions of Dollars)	9

Table 7: Household and non-household balance sheet variables. T: data transformation code, where 1 = no transformation, 2 = first difference, and 5 = first difference of log series. G: Group code, where 10 = household balance sheet variables and 14 = non-household balance sheet variables

FRED MNEMONIC	SW MNEMONIC	T	DESCRIPTION	G
LIABPIx	liab_PDISA	5	Liabilities of Households and Nonprofit Organizations Relative to Personal Disposable Income (Percent)	10
NWPIx	W_PDISA	1	Net Worth of Households and Nonprofit Organizations Relative to Disposable Personal Income (Percent)	10
CONSPIx		2	Nonrevolving consumer credit to Personal Income	10
TARESAx	Real HHW:TA_RESA	5	Real Assets of Households and Nonprofit Organizations excluding Real Estate Assets (Billions of 2017 Dollars), deflated by Core PCE	10
TNWBSHNOx	Real HHW:WSA	5	Real Net Worth of Households and Nonprofit Organizations (Billions of 2017 Dollars), deflated by Core PCE	10
HNOREMQ027Sx	Real HHW:RESA	5	Real Real Estate Assets of Households and Nonprofit Organizations (Billions of 2017 Dollars), deflated by Core PCE	10
TABSHNOx	Real HHW:TASA	5	Real Total Assets of Households and Nonprofit Organizations (Billions of 2017 Dollars), deflated by Core PCE	10
TFAABSHNOx	Real HHW:FinSA	5	Real Total Financial Assets of Households and Nonprofit Organizations (Billions of 2017 Dollars), deflated by Core PCE	10
TLBSHNOx	Real HHW:LiabSA	5	Real Total Liabilities of Households and Nonprofit Organizations (Billions of 2017 Dollars), deflated by Core PCE	10
TLBSNNCBBDIx		1	Nonfinancial Corporate Business Sector Liabilities to Disposable Business Income (Percent)	14
TNWMVBSNNCBBDIx		2	Nonfinancial Corporate Business Sector Net Worth to Disposable Business Income (Percent)	14
TLBSNNBBDIx		1	Nonfinancial Noncorporate Business Sector Liabilities to Disposable Business Income (Percent)	14
TNWBSNNBBDIx		2	Nonfinancial Noncorporate Business Sector Net Worth to Disposable Business Income (Percent)	14
CNCFx		5	Real Disposable Business Income, Billions of 2017 Dollars (Corporate cash flow with IVA minus taxes on corporate income, deflated by Implicit Price Deflator for Business Sector IPDBS)	14
TTAABSNNCBx		5	Real Nonfinancial Corporate Business Sector Assets (Billions of 2017 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS	14
TLBSNNCBx		5	Real Nonfinancial Corporate Business Sector Liabilities (Billions of 2017 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS	14
TNWMVBSNNCBx		5	Real Nonfinancial Corporate Business Sector Net Worth (Billions of 2017 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS	14
TABSNNBx		5	Real Nonfinancial Noncorporate Business Sector Assets (Billions of 2017 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS	14
TLBSNNBx		5	Real Nonfinancial Noncorporate Business Sector Liabilities (Billions of 2017 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS	14
TNWBSNNBx		5	Real Nonfinancial Noncorporate Business Sector Net Worth (Billions of 2017 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS	14

Table 8: Exchange rate variables. T: data transformation code, where 5 = first difference of log series. G: Group code, where 11 = exchange rate variables.

FRED MNEMONIC	SW MNEMONIC	T	DESCRIPTION	G
EXCAUSx	EX rate:Canada	5	Canada / U.S. Foreign Exchange Rate	11
EXJPUSx	Ex rate:Japan	5	Japan / U.S. Foreign Exchange Rate	11
EXSZUSx	Ex rate:Switz	5	Switzerland / U.S. Foreign Exchange Rate	11
EXUSUKx	Ex rate:UK	5	U.S. / U.K. Foreign Exchange Rate	11

Table 9: Stock market variables and consumer sentiment. T: data transformation code, where 1 = no transformation, 2 = first difference, and 5 = first difference of log series. G: Group code, where 12 = consumer sentiment variable, and 13 = stock market variables.

FRED MNEMONIC	SW MNEMONIC	T	DESCRIPTION	G
UMCSENTx	Cons. Expectations	1	University of Michigan: Consumer Sentiment (Index 1st Quarter 1966=100)	12
NIKKEI225		5	Nikkei Stock Average	13
S&P 500		5	S&P's Common Stock Price Index: Composite	13
S&P div yield		2	S&P's Composite Common Stock: Dividend Yield	13
S&P PE ratio		5	S&P's Composite Common Stock: Price-Earnings Ratio	13

S7 Additional empirical experiment on realized volatility

It is an important task in finance to predict realized volatility (Chen et al., 2010), and this subsection attempts to tackle this problem by considering the daily realized volatility for $N = 46$ stocks from January 2, 2012 to December 31, 2013, with sample size $T = 495$. The stocks are from S&P 500 companies with the largest trading volumes on the first day of 2013, and they cover a wide range of sectors, including communication service, information technology, consumer, finance, healthcare, materials and energy. The tick-by-tick data are downloaded from the Wharton Research Data Service (WRDS), and the daily realized volatility is calculated based on five-minute returns (Andersen et al., 2006). The stationarity can be confirmed for these series by checking their sample autocorrelation functions, and each sequence is standardized to have zero mean and unit variance; see the Table 11 information on the 46 stocks.

Algorithm 1 is applied again to search for the VAR sieve estimates, and the running order T_0 is initially set to $\lfloor\sqrt{495}\rfloor$. The AIC chooses the sparsity level $s = 4$ for most running orders T_0 and $s = 3$ for the others, while the selected lags are always among $\{1, 3, 4, 8\}$. We may argue that the movement of market volatility is largely driven by intra-week information, with a bit spillover effect from the previous week. As a result, the lag order is fixed at $T_0 = 10$, providing two more lags as a buffer. The ranks and sparsity level selected by the AIC are $(r_1, r_2, s) = (3, 3, 4)$. The rolling forecast procedure, used in the macroeconomic application of the main paper, is employed here as well, reserving the last 10% of the observations for one-step-ahead predictions. Table 2 presents the mean squared forecast error (MSFE) and model confidence set (MCS) p-values from the proposed method, alongside the competing methods introduced in Section 5 of the main paper. For VAR-based models, we consider a finite lag order of $p = 4$. At the suggestion of a reviewer, we include the Heterogeneous AutoRegressive (HAR) model (Corsi, 2009) as an additional benchmark for this application. In contrast to the macroeconomic analysis in the main paper, our preliminary analysis of the realized volatility dataset suggests that the underlying response and predictor factors lie within approximately the same low-dimensional space. Therefore, the proposed model is fitted using the approach outlined in Remark 5, which enforces $\boldsymbol{U}_1 = \boldsymbol{U}_2$. As shown in Table 10, the proposed model performs best among all models. This superior performance, compared to VARMA-based methods, may be due to the non-consecutive nature of the non-zero lags, which requires greater flexibility in capturing diverse temporal patterns. Additionally, the realized volatility data exhibit longer temporal dependence, meaning that truncating the lags may introduce bias, resulting in weaker performance from finite-order VAR models with $p = 4$.

Table 10: Mean squared forecast errors (MSFE) and mean absolute forecast errors (MAFE) of our methods and other models on the realized volatility dataset. The best result in each column is highlighted in bold black font.

Models	MSFE	p_{MCS}
RW	6.46	0.00
AR(1)	5.00	0.20
AR(2)	4.95	0.20
VAR(1)	5.18	0.03
VAR(2)	5.77	0.03
VAR (ℓ_1)	5.24	0.03
VAR (MLR)	5.06	0.20
VAR (SHORR)	5.12	0.20
BVAR	5.41	0.01
VARMA (ℓ_1)	5.21	0.03
VARMA (HLag)	5.21	0.03
Approx VARMA	5.19	0.03
FactorAug Reg	4.82	0.20
HAR	4.75	0.20
Ours	**4.68**	**1.00**

Table 11: Forty six selected S&P 500 stocks. CODE: stock code in the New York Stock Exchange. NAME: name of company. G: group code, where 1 = communication service, 2 = information technology, 3 = consumer, 4 = financials, 5 = healthcare, 6 = materials and industrials, and 7 = energy and utilities.

CODE	NAME	G	CODE	NAME	G
T	AT&T Inc.	1	JPM	JPMorgan Chase & Co.	4
NWSA	News Corp	1	WFC	Wells Fargo & Company	4
FTR	Frontier Communications Parent Inc	1	MS	Morgan Stanley	4
VZ	Verizon Communications Inc.	1	AIG	American International Group Inc.	4
IPG	Interpublic Group of Companies Inc	1	MET	MetLife Inc.	4
MSFT	Microsoft Corporation	2	RF	Regions Financial Corp	4
HPQ	HP Inc	2	PGR	Progressive Corporation	4
INTC	Intel Corporation	2	SCHW	Charles Schwab Corporation	4
EMC	EMC Instytut Medyczny SA	2	FITB	Fifth Third Bancorp	4
ORCL	Oracle Corporation	2	PFE	Pfizer Inc.	5
MU	Micron Technology Inc.	2	ABT	Abbott Laboratories	5
AMD	Advanced Micro Devices Inc.	2	MRK	Merck & Co. Inc.	5
AAPL	Apple Inc.	2	RAD	Rite Aid Corporation	5
YHOO	Yahoo! Inc.	2	JNJ	Johnson & Johnson	5
QCOM	Qualcomm Inc	2	AA	Alcoa Corp	
GLW	Corning Incorporated	2	FCX	Freeport-McMoRan Inc.	
AMAT	Applied Materials Inc.	2	X	United States Steel Corporation	
F	Ford Motor Company	3	GE	General Electric Company	
LVS	Las Vegas Sands Corp.	3	CSX	CSX Corporation	
EBAY	eBay Inc.	3	ANR	Alpha Natural Resources	7
KO	Coca-Cola Company	3	XOM	Exxon Mobil Corporation	7
BAC	Bank of America Corp	4	CHK	Chesapeake Energy	7
C	Citigroup Inc.	4	EXC	Exelon Corporation	7

References

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2006). Volatility and correlation forecasting. *Handbook of economic forecasting*, 1:777–878.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43:1535–1567.

Cai, T. T. and Zhang, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics*, 46:60–89.

Candes, E. J. and Plan, Y. (2011). Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Transactions on Information Theory*, 57:2342–2359.

Chen, Y., Härdle, W. K., and Pigorsch, U. (2010). Localized realized volatility modeling. *Journal of the American Statistical Association*, 105:1376–1393.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Han, R., Willett, R., and Zhang, A. R. (2022). An optimal statistical and computational framework for generalized tensor estimation. *The Annals of Statistics*, 50:1–29.

Jain, P., Tewari, A., and Kar, P. (2014). On iterative hard thresholding methods for high-dimensional m-estimation. In *Advances in Neural Information Processing Systems*, volume 27.

Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, 51:455–500.

Lewis, R. and Reinsel, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis*, 16:393–411.

Nesterov, Y. (2003). *Introductory Lectures on Convex optimization: A Basic Course*, volume 87. Springer Science & Business Media, New York.

Simchowitz, M., Mania, H., Tu, S., Jordan, M. I., and Recht, B. (2018). Learning without mixing: Towards a sharp analysis of linear system identification. In *Conference On Learning Theory*, pages 439–473. PMLR.

Stock, J. H. and Watson, M. W. (2012). Disentangling the channels of the 2007-2009 recession. Working Paper 18094, National Bureau of Economic Research.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311.

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*.