Change point detection in low-rank VAR processes

Farida Enikeeva, Olga Klopp and Mathilde Rousselot

April 26, 2023

Abstract

Vector autoregressive (VAR) models are widely used in multivariate time series analysis for describing the short-time dynamics of the data. The reduced-rank VAR models are of particular interest when dealing with high-dimensional and highly correlated time series. Many results for these models are based on the stationarity assumption that does not hold in several applications when the data exhibits structural breaks. We consider a low-rank piecewise stationary VAR model with possible changes in the transition matrix of the observed process. We develop a new test of presence of a change-point in the transition matrix and show its minimax optimality with respect to the dimension and the sample size. Our two-step change-point detection strategy is based on the construction of estimators for the transition matrices and using them in a penalized version of the likelihood ratio test statistic. The effectiveness of the proposed procedure is illustrated on synthetic data.

1 Introduction

Vector autoregression (VAR) is a classical model of multivariate time series analysis that has been successfully used to model data in epidemiology (Khan et al. 2020), economics and finance (Fan et al. 2010), medical research (Wild et al. 2010), econometrics (Stock and Watson 2001) and neuroscience (Gorrostieta et al. 2013). During the past two decades, a considerable interest has been focused on the problems of high-dimensional inference. High-dimensional multivariate time series have a large amount of variables but only a limited number of time steps. In such a situation the VAR model is ill-posed: it suffers from the over-parametrization issue as the number of parameters in the coefficient matrix is comparable to or much larger than the number of time series observations. To address this issue different structural assumptions on the transition matrix have been proposed. The idea is that one expects the system to be controlled primarily by a low-dimensional subset of variables. For example, in finance, the data can have a huge ambient dimension which includes financial instruments such as stocks, bonds, and etc. These financial instruments may be combined into a much smaller subset of macro-variables that actually govern the market.

A reduced-rank VAR model for multivariate time series analysis was first introduced by Velu et al. (1986) (see also (Reinsel and Velu 1998, Lütkepohl 2005) for detailed background). In (Velu et al. 1986) the authors show the consistency of the least squares estimator of the VAR transition matrix modeled by a product of two rectangular matrices. Ahn and Reinsel (1988) generalize these results to nested low-rank autoregressive models. Negahban and Wainwright (2011) propose a least-squares nuclear norm penalized estimator for the transition matrix and show its consistency in the Frobenius norm. Alquier et al. (2020) consider the problem of prediction for low-rank VAR modes. Their method is based on rank-penalized least-squares estimator. More recently Wang and Tsay (2021) proposed a method of the transition matrix estimation using the constrained Yule-Walker equations and showed its optimality under the β-mixing dependency condition.

Another popular assumption on the matrix structure is the entry-wise sparsity. Under the entry-wise sparsity assumptions, She et al. (2015) estimate jointly the transition matrix and the precision matrix of the process using joint regularization; Basu and Michailidis (2015) use an ℓ_1-penalized log-likelihood for estimating the transition matrix and Melnyk and Banerjee (2016) use a penalized log-likelihood estimator with ℓ_1-norm penalization and the group lasso penalty. Finally, Basu et al. (2019) consider the VAR model under both sparsity and low rank assumptions. They assume that the transition matrix can be written as the sum of a sparse matrix and a low rank matrix. The proposed penalized least-squares estimator uses the nuclear norm penalty combined with the ℓ_1-penalization.

Many of these results are based on the assumption of stationarity. However, in applications the data might exhibit structural breaks with several discontinuity points in the distribution and the stationarity assumption does not hold for the entire data set. In such situations a natural approach is to assume a piece-wise stationary model and apply existing methods to the intervals of stationarity. Then, the crucial question is that of the identification of change-points.

Detection of structural changes in multivariate time series is one of the major problems arising in applications. In neuroscience the breaks in the sequence of the electroencephalogram (EEG) signals correspond to changes in brain activity (Michel and Murray 2012). In financial analysis the volatility of the market indexes might change at certain time points due to some external event (Aue et al. 2009). Efficient detection of such breaks heavily relies on the underlying mechanism of the data temporal evolution. The goal of this paper is to propose a procedure that allows to detect the presence of changes in the matrix parameter of a low-rank VAR process. Note that classical methods of multidimensional change-point detection (see, for example, (Basseville and Nikiforov 1993)) are not applicable in the high-dimensional setting since the model dimension can be much larger than the sample size.

1.1 Low-rank VAR Model

We say that the process $\{X_t\}_{t\in\mathbb{Z}}$ is a p-dimensional vector autoregressive (VAR) process $\mathrm{VAR}_p(\Theta, \Sigma_Z)$ with Gaussian innovations, if

$$X_{t+1} = \Theta\, X_t + Z_{t+1} \quad \forall t \in \mathbb{Z}, \tag{1}$$

where $\Theta \in \mathbb{R}^{p\times p}$ is the transition matrix and $\{Z_t\}_{t\in\mathbb{Z}}$ is a p-dimensional centered Gaussian independent noise with the covariance matrix $\Sigma_Z \in \mathbb{R}^{p\times p}$, $Z_t \sim \mathcal{N}_p(0, \Sigma_Z)$. We assume that the operator norm of Θ satisfies $\|\Theta\|_{\mathrm{op}} = \gamma < 1$. Under this condition $\{X_t\}_{t\in\mathbb{Z}}$ is a stationary causal process (cf. (Lütkepohl 2005) for more details on stationary VAR processes) and the covariance matrix Σ of X_t satisfies the Lyapunov equation

$$\Sigma = \Theta\Sigma\Theta^{\mathrm{T}} + \Sigma_Z. \tag{2}$$

If $\|\Theta\|_{\text{op}} < 1$, then (2) has a unique positive-definite solution Σ.

In a more general setting the transition matrix may change over time. In our problem we observe a trajectory $\mathfrak{X} = (X_0, \dots, X_T)$ of a piece-wise stationary centered p-dimensional Gaussian $\text{VAR}_p(\Theta_1, \Theta_2, \Sigma_Z)$ process,

$$X_{t+1} = \Theta^t X_t + Z_{t+1} \quad t = 0, \dots, T-1, \tag{3}$$

where $Z_t \sim \mathcal{N}_p(0, \Sigma_Z)$ are i.i.d. and

$$\Theta^t = \Theta_1 \mathbf{1}_{\{0 \le t \le \tau\}} + \Theta_2 \mathbf{1}_{\{\tau+1 \le t \le T\}}.$$

Here τ stands for the change-point in transition matrix. We assume that the transition matrices are of the rank at most R, $\text{rank}(\Theta_j) \le R \; \forall j = 1, 2$. In general, Σ_Z can vary over segments, but we consider it to be fixed to avoid additional technicalities. In the following we assume that the system matrices Θ_j are stable:

Assumption 1. *For $j = 1, 2$ there exists $0 < \gamma_j < 1$ such that $\|\Theta_j\|_{\text{op}} \le \gamma_j$.*

Note that under this assumption the matrices Θ_j satisfy the Lyapunov equation (2), its unique solution gives the corresponding covariance matrix Σ_j of the VAR process.

1.2 Related work

The existing literature mainly considers the related question of change-point localization under entry-wise sparsity or entry-wise sparsity plus low-rank assumptions on the transition matrix. For example, under sparsity assumptions on the transition matrices Safikhani and Shojaie (2022) use a fused Lasso approach to estimate the breakpoints as well as the process parameters. They obtain the localization error bound under the assumption that the minimum distance between two change-points is a sufficiently large constant independent of the number of observations. In the same setting, Wang et al. (2019) obtain a stronger result allowing decreasing distance between the change-points. Their estimator is based on the combination of the Lasso and group Lasso methods. Bai et al. (2020) assume that the transition matrices can be decomposed into a constant low-rank component and a sparse time evolving component. They develop a strategy for identification of change-points in the sparse component and provide probabilistic guarantees for the accuracy of their identification. More recently, Bai et al. (2022) considered the "low rank plus sparse" VAR model where both matrix components may change with time under the assumption that the maximum absolute value of the entries of the low-rank component is bounded by $\left(\frac{\log(pT)}{T}\right)^{1/2}$, which goes to zero when the number of observations T is growing. They propose a method of multiple change-point estimation based on the plug-in estimators obtained in (Basu et al. 2019) and provide its theoretical guarantees.

1.3 Our contributions

We consider the problem of testing the presence of a change in the low-rank transition matrix. Our testing procedure is based on the plug-in test statistic with the estimated low-rank matrices before and after an eventual change. We prove that our test allows for reliable change-point detection when the squared Frobenius norm of the change is larger then $\frac{Rp}{Tq^2(t/T)}$ (up to a logarithmic factor, for the precise statement see Theorem 1) where $q(t) = \sqrt{t(1-t)}$ for $t \in [0,1]$ controls the impact of the change-point location on the rate. An important point is that this result does not require any condition on the minimum spacing between the change-point and the boundaries of the interval of observations. We also show that our testing procedure is minimax-rate optimal both in terms of the dimension and the sample size (Theorem 2). As a bi-product of our analysis we provide a new result on the consistency of the nuclear norm-penalized estimator of the transition matrix in operator norm (see Proposition 6).

1.4 Notation

We start with basic notation used in this paper. For any matrix M, we denote by M_{ij} its entry in the ith row and jth column and by $M_{i\cdot}$ its ith row. The notation $\mathrm{diag}(M)$ stands for the diagonal of a square matrix M and M^{T} for the transpose of M. The column vector of dimension n with unit entries is denoted by $\mathbf{1}_n = (1,\dots,1)^{\mathrm{T}}$ and the column vector of dimension n with zero entries is denoted by $\mathbf{0}_n = (0,\dots,0)^{\mathrm{T}}$. The identity matrix of dimension n is denoted by id_n. For a set A, we denote by $\mathbf{1}_A$ its indicator function.

For any matrix $M \in \mathbb{R}^{d_1\times d_2}$, $\|M\|_2$ is its Frobenius norm, $\|M\|_{\mathrm{op}}$ is its operator norm (its largest singular value). We denote by $\sigma_j(M)$ the jth singular value and by $\sigma_{\max}(M)$ and $\sigma_{\min}(M)$ the largest and the smallest non-zero singular value of M. Assuming that matrix M has rank R we consider its ordered singular values $\sigma_1(M) \geq \sigma_2(M) \geq \cdots \geq \sigma_R(M) > 0$ and we denote the condition number of M by $\kappa(M) = \sigma_1(M)/\sigma_R(M)$.

For M, a $d_1 \times d_2$ matrix, let $u_j(M)$ and $v_j(M)$ be respectively the left and right orthonormal singular vectors of M, $S_1(M)$ be the linear span of $\{u_j(M)\}$, $S_2(M)$ be the linear span of $\{v_j(M)\}$. We denote by $S^{\perp}$ the orthogonal complement of S. For B, a $d_1\times d_2$ matrix, let $\mathbf{Pr}_M(B) = B - \mathcal{P}_{S_1^{\perp}(M)} B \mathcal{P}_{S_2^{\perp}(M)}$ and $\mathbf{Pr}_M^{\perp}(B) = \mathcal{P}_{S_1^{\perp}(M)} B \mathcal{P}_{S_2^{\perp}(M)}$, where $\mathcal{P}_S$ is the orthogonal projector on the linear vector subspace S.

We write $X \lesssim Y$ and $X \gtrsim Y$ if $X \leq CY$ and, respectively, $X \geq CY$ for some absolute constant $C > 0$.

We denote by $\mathcal{M}_p(R,\gamma)$ the set of all $p\times p$ real matrices of rank at most R with the operator norm bounded by γ:

$$\mathcal{M}_p(R,\gamma) = \Big\{ M \in \mathbb{R}^{p\times p} : \mathrm{rank}(M) \leq R \quad \text{and} \quad \|M\|_{\mathrm{op}} \leq \gamma \Big\}$$

For any $t \in \{1,\dots,T-1\}$, we introduce the following random matrices :

$$\begin{aligned}
\mathfrak{X}_{\geq t} &:= (X_t,\dots,X_{T-1}) \in \mathbb{R}^{p\times(T-t)}, \quad \mathfrak{X}_{<t} := (X_0,\dots,X_{t-1}) \in \mathbb{R}^{p\times t}\\
\mathfrak{Y}_{>t} &:= (X_{t+1},\dots,X_T) \in \mathbb{R}^{p\times(T-t)}, \quad \mathfrak{Y}_{\leq t} := (X_1,\dots,X_t) \in \mathbb{R}^{p\times t}\\
\mathfrak{Z}_{\geq t} &:= (Z_t,\dots,Z_{T-1}) \in \mathbb{R}^{p\times(T-t)}, \quad \mathfrak{Z}_{<t} := (Z_0,\dots,Z_{t-1}) \in \mathbb{R}^{p\times t}.
\end{aligned}$$

Here $\mathfrak{X}_{<t}$ ($\mathfrak{X}_{\geq t}$) contains our observations before (after) a given time point t and $\mathfrak{Z}_{\cdot}$ contains the innovation noise. $\mathfrak{Y}_{\cdot}$ is obtained from $\mathfrak{X}_{\cdot}$ by shifting our observations by one time step.

2 Change-point detection problem

We will consider the problem of detection of a single change-point in the VAR model (3) with the transition matrix Θ^t that can change at some unknown point $\tau \in \mathcal{D}_T \subseteq \{1, \dots, T-1\}$,

$$\Theta^t = \Theta_1 \mathbf{1}_{\{0 \le t \le \tau\}} + \Theta_2 \mathbf{1}_{\{\tau+1 \le t \le T\}}.$$

The difficulty of assessing the existence of a change-point can be quantified by what is called *energy* of the change point. It is defined as the product of the Frobenius norm of the jump in transition matrix and the function $q(t) = \sqrt{t(1-t)}$ for $t \in [0, 1]$. The function $q(t)$ quantifies the impact of change-point location to the difficulty of detecting the change. Thus, we write the detection problem as the problem of testing whether the jump energy

$$\mathcal{E}(t, \Theta_1, \Theta_2) = q(t/T) \|\Theta_1 - \Theta_2\|_2$$

is zero or not. To formulate the hypothesis testing problem, we define the set of all pairs of matrices with the operator norm bounded by $\gamma \in (0, 1)$ before (Θ_1) and after (Θ_2) the change at the location t such that the jump energy is at least $r > 0$,

$$\mathcal{V}_{p,t}(r) = \Big\{ \big(\Theta_1, \Theta_2\big) \in \mathcal{M}_p^{\otimes 2}(R, \gamma) : \ \mathcal{E}(t, \Theta_1, \Theta_2) \ge r \Big\}. \tag{4}$$

Let $\mathcal{V}_{p,0}$ denote the set without a jump:

$$\mathcal{V}_{p,0} = \Big\{ \big(\Theta_1, \Theta_2\big) \in \mathcal{M}_p^{\otimes 2}(R, \gamma) : \Theta_1 = \Theta_2 \Big\}$$

We will test the null hypothesis of no-change

$$\mathrm{H}_0 : \ (\Theta_1, \Theta_2) \in \mathcal{V}_{p,0} \tag{5}$$

against the alternative hypothesis of a change in the transition matrix:

$$\mathrm{H}_1 : \ (\Theta_1, \Theta_2) \in \mathcal{V}_{p,\tau}(\mathcal{R}_{p,\mathcal{D}_T}) \text{ for some } \tau \in \mathcal{D}_T, \tag{6}$$

where $\mathcal{R}_{p,\mathcal{D}_T} > 0$ is the minimal amount of energy that guarantees the change-point detection and $\mathcal{D}_T \subseteq \{1, \dots, T-1\}$.

We construct a change-point detection procedure based on the penalized least-squares minimization approach. Our procedure has two steps. In the first step, for each $t \in \mathcal{D}_T$, we compute estimators of the transition matrix at each of the two intervals $[0, t]$ and $[t, T]$. Once equipped with such estimators we use an information criteria to build the test statistic and the change-point estimator. Our estimators of transition matrices are based on the nuclear norm minimization criteria which is a convex relaxation of the rank-constrained minimization problem. Note that the estimation of the transition matrix is easier if t (respectively $T - t$) is large comparing to the matrix dimension p. For difficult cases of $t < p$ (respectively $T - t < p$) we use a slightly different penalization which allows us to cope with the lack of observations.

2.1 Transition matrix estimation

We start by giving a general construction of an estimator of the transition matrix Θ from the observations $\mathfrak{X}_{[t_1,t_2]} = (X_{t_1}, \ldots, X_{t_2}) \in \mathbb{R}^{p\times(t_2-t_1+1)}$ of a $\mathrm{VAR}_p(\Theta, \Sigma_Z)$ process observed at the consecutive time moments $t_1, t_1+1, \ldots, t_2$. Later on, we will use the estimators of the transition matrices obtained withing the intervals $[0,t]$ and $[t,T]$ in order to construct our change-point detection procedure.

For $t_2 - t_1 > p$, we set

$$\varphi(\mathfrak{X}_{[t_1,t_2]}, M) = \frac{1}{t_2-t_1}\sum_{i=t_1}^{t_2-1} \|X_{i+1} - MX_i\|_2^2 + \lambda\|M\|_* \tag{7}$$

and for $t_2 - t_1 \le p$,

$$\varphi(\mathfrak{X}_{[t_1,t_2]}, M) = \frac{1}{t_2-t_1}\sum_{i=t_1}^{t_2-1} \|X_{i+1} - MX_i\|_2^2 + \lambda\|M\mathfrak{X}_{[t_1,t_2]}\|_*. \tag{8}$$

Here $\|\cdot\|_*$ stands for the nuclear norm and the regularization parameter$\lambda_{[t_1,t_2]}$ is given by

$$\lambda = 6c_1^*\sqrt{\|\Sigma_Z\|_{\mathrm{op}}}\frac{\sqrt{p}}{t_2-t_1}\mathbf{1}_{\{t_2-t_1\le p\}} + 2c_2^*\frac{\|\Sigma\|_{\mathrm{op}}}{1-\gamma}\sqrt{\frac{p}{t_2-t_1}}\mathbf{1}_{\{t_2-t_1>p\}}, \tag{9}$$

where c_1^* and c_2^* are absolute constants provided in Lemmas 9 and 10.

We consider the following estimator of Θ within the interval $[t_1, t_2]$:

$$\widehat{\Theta} \in \underset{M\in\mathbb{R}^{p\times p}}{\arg\min}\, \varphi(\mathfrak{X}_{[t_1,t_2]}, M). \tag{10}$$

Estimator (7) was first introduced in (Negahban and Wainwright 2011). In the case of large number of observations $t_2 - t_1 \ge p$, with no change in the matrix of parameter, Negahban and Wainwright (2011) prove the estimator consistency in the Frobenius norm (see Proposition 5). Note that estimator provided by (10) is also consistent in the operator norm, see Proposition 6 in the Appendix.

2.2 Testing procedure

For any $t \in \mathcal{D}_T$, we will estimate the transition matrix on the intervals before and after the time t, $[0,t)$ and $[t,T)$ using, respectively, the observations $\mathfrak{X}_{\le t}$ and $\mathfrak{X}_{\ge t}$. These estimators are obtained as the solutions of the SDP (7) or (8) depending on the proximity of the point t to the endpoints of the observation interval $[0,T]$. Denote

$$\varphi_1^t(\mathfrak{X}, M) = \varphi(\mathfrak{X}_{[0,t]}, M) \quad \text{and} \quad \varphi_2^t(\mathfrak{X}, M) = \varphi(\mathfrak{X}_{[t,T]}, M).$$

Then the optimization programs for estimation of Θ_1 and Θ_2 can be written as

$$\widehat{\Theta}_1^t \in \underset{M\in\mathbb{R}^{p\times p}}{\arg\min}\, \varphi_1^t(\mathfrak{X}, M) \tag{11}$$

and

$$\widehat{\Theta}_2^t \in \underset{M\in\mathbb{R}^{p\times p}}{\arg\min}\, \varphi_2^t(\mathfrak{X}, M) \tag{12}$$

where

$$\varphi_1^t(\mathfrak{X}, M) = \frac{1}{t}\|\mathfrak{Y}_{\le t} - M\mathfrak{X}_{<t}\|_2^2 + \lambda_1^t\Big(\|M\mathfrak{X}_{<t}\|_*\mathbf{1}_{\{t\le p\}} + \|M\|_*\mathbf{1}_{\{t>p\}}\Big)$$

and

$$\varphi_2^t(\mathfrak{X}, M) = \frac{1}{T-t}\|\mathfrak{Y}_{>t} - M\mathfrak{X}_{\ge t}\|_2^2 + \lambda_2^t\Big(\|M\mathfrak{X}_{\ge t}\|_*\mathbf{1}_{\{T-t\le p\}} + \|M\|_*\mathbf{1}_{\{T-t>p\}}\Big)$$

with the penalties

$$\lambda_1^t := 6c_1^*\sqrt{\|\Sigma_Z\|_{\mathrm{op}}}\frac{\sqrt{p}}{t}\mathbf{1}_{\{t\le p\}} + 2c_2^*\frac{\|\Sigma_1\|_{\mathrm{op}}}{1-\gamma_1}\sqrt{\frac{p}{t}}\mathbf{1}_{\{t>p\}} \tag{13}$$

$$\lambda_2^t := 6c_1^*\sqrt{\|\Sigma_Z\|_{\mathrm{op}}}\frac{\sqrt{p}}{T-t}\mathbf{1}_{\{T-t\le p\}} + 2c_2^*\frac{\|\Sigma_2\|_{\mathrm{op}}}{1-\gamma_2}\sqrt{\frac{p}{T-t}}\mathbf{1}_{\{T-t>p\}}. \tag{14}$$

Finally we define the following test statistic:

$$\mathcal{F}(t) = \Big(\varphi_1^t(\mathfrak{X}, \widehat{\Theta}_2^t) - \varphi_1^t(\mathfrak{X}, \widehat{\Theta}_1^t)\Big)\mathbf{1}_{\{t<T/2\}} + \Big(\varphi_2^t(\mathfrak{X}, \widehat{\Theta}_1^t) - \varphi_2^t(\mathfrak{X}, \widehat{\Theta}_2^t)\Big)\mathbf{1}_{\{t\ge T/2\}}. \tag{15}$$

Let $\mathcal{T}$ be a subset of $\mathcal{D}_T$ that approximates the set of possible change-point locations. In case of testing against a simple alternative of change at a given point τ, we take $\mathcal{T} = \mathcal{D}_T = \{\tau\}$. In case of a composite alternative, we can choose $\mathcal{T} = \mathcal{D}_T$ or use an appropriate grid on $\mathcal{D}_T$. For a given significance level $\alpha \in (0,1)$ we introduce the test

$$\psi_\alpha(\mathfrak{X}) = \mathbf{1}\left\{\max_{t\in\mathcal{T}}\frac{\mathcal{F}(t)}{H_{\alpha,t}} > 1\right\} \tag{16}$$

where $H_{\alpha,t}$ is the threshold defined as follows:

$$H_{\alpha,t} = \begin{cases} C^*\frac{Rp}{Tq^2(t/T)}\left(1 + \frac{2}{1-\gamma}\frac{p+\log(8|\mathcal{T}|/\alpha)}{T}\right), & t\wedge(T-t)\le p \\ C^*\frac{Rp}{Tq^2(t/T)}\left(1 + \frac{2}{1-\gamma}\left(1 + \frac{\log(8|\mathcal{T}|/\alpha)}{T}\right)\right), & t\wedge(T-t)> p \end{cases} \tag{17}$$

with $C^* = C\max\left(\|\Sigma_Z\|_{\mathrm{op}}, \|\Sigma\|_{\mathrm{op}}\frac{\kappa^2(\Sigma)}{(1-\gamma)^2}\right)$ for an absolute constant C. Here Σ is the covariance matrix of the VAR process under the null. We can estimate its operator norm from the Lyapunov equation (2) as $\|\Sigma\|_{\mathrm{op}} \le \|\Sigma_Z\|_{\mathrm{op}}/(1-\gamma^2)$ and its condition number as $\kappa(\Sigma) \le \kappa(\Sigma_Z)\frac{1+\gamma^2}{1-\gamma^2}$ using Lemma 13.

Theorem 1. *Let $\alpha, \beta \in (0,1)$ be given significance levels such that $p > C\log\left(\frac{|\mathcal{T}|}{\alpha\beta}\right)$ for some absolute constant $C > 0$. Let $p \le T/2$ and $(\Theta_1, \Theta_2) \in \mathcal{M}_p^{\otimes 2}(R,\gamma)$ be a couple of transition matrices with the change-point energy satisfying*

$$\mathcal{E}^2(\tau, \Theta_1, \Theta_2) \ge \Xi\frac{Rp}{T}\left(1 + \frac{1}{T}\log\left(\frac{64|\mathcal{T}|}{\alpha\beta}\right)\right),$$

where the constant $\Xi = \Xi(\Sigma_1, \Sigma_2, \Sigma, \gamma)$ is defined in Lemma 3. Then, the α-level test ψ_α defined in (16) has the type II error smaller than β.

Proof. The proof of Theorem 1 is based on Lemmas 2 and 3. Lemma 2 implies that $\alpha(\psi_\alpha) \leq \alpha$. By Lemma 3, the type II error smaller than β is guaranteed if $R_{p,\mathcal{D}_T}$ satisfies (24) which completes the proof of the theorem. $\square$

Remark 1. *Note that the detection rate is of the order $\sqrt{Rp/T}$ in the case of testing at a given point τ and when $T > \log\left(\frac{T}{\alpha\beta}\right)$ in the case of undefined change-point location with $\mathcal{T} = \mathcal{D}_T = \{1, \ldots, T-1\}$. For smaller T, we can replace $\log T$ by $\log\log T$ by taking the dyadic grid (see, for example, (Liu et al. 2021)) defined in the following way: $\mathcal{T} = \mathcal{T}^L \cup \mathcal{T}^R$, where*

$$\mathcal{T}^L = \left\{2^k,\ k = 0, \ldots, \lfloor \log_2(T/2) \rfloor\right\}, \quad \mathcal{T}^R = \left\{T - 2^k,\ k = 0, \ldots, \lfloor \log_2(T/2) \rfloor\right\}. \tag{18}$$

Remark 2. *The test defined in (16) detects the change-points τ located in any time point. If we know that the change-point belongs to the interval $(p, T-p)$, where the consistent estimation of the transition matrices is possible, then, we can define a penalized likelihood ratio statistic defined as*

$$\mathcal{G}(t) = \frac{t}{T}\left(\varphi_1^t(\mathfrak{X}, \widehat{\Theta}_2^t) - \varphi_1^t(\mathfrak{X}, \widehat{\Theta}_1^t)\right) + \frac{T-t}{T}\left(\varphi_2^t(\mathfrak{X}, \widehat{\Theta}_1^t) - \varphi_2^t(\mathfrak{X}, \widehat{\Theta}_2^t)\right). \tag{19}$$

We can show, using exactly the same technique as in Lemmas 2 and 3 that the procedure will detect the change-point with the same detection rate Rp/T up to the logarithmic term and with a slightly different constant.

We have the following optimality result on the minimal detectable change-point energy.

Theorem 2. *Let $\alpha \in (0,1)$ be given significance level. Assume that the change-point location satisfies the following condition:*

$$\min(\tau/T, 1 - \tau/T) \geq h_* > 0.$$

Let the change-point energy $\mathcal{E}(\tau, \Delta\Theta) = q(\tau/T)\|\Delta\Theta\|_2$ satisfy

(a) $\lim\limits_{p,T\to\infty} \mathcal{E}(\tau, \Theta_1, \Theta_2)\left(\frac{T}{p}\right)^{1/2} = 0$ *if* $p \lesssim \sqrt{T}$

(b) $\lim\limits_{p,T\to\infty} \mathcal{E}(\tau, \Theta_1, \Theta_2)T^{1/4} = 0$ *if* $p \gtrsim \sqrt{T}$.

Then, the type II error of any α-level test ψ_α satisfies $\liminf\limits_{p,T\to\infty} \beta(\psi) \geq 1 - \alpha$ and the lower bound on the testing rate is $\mathcal{R}_{p,\tau} = o\left(\sqrt{\frac{p}{T}} \wedge \frac{1}{T^{1/4}}\right)$.

For the matrices of dimension $p \lesssim \sqrt{T}$, Theorems 1 and 2 imply that the minimal detectable energy satisfies the condition

$$\sqrt{\frac{p}{T}} \lesssim \mathcal{E}(\tau, \Theta_1, \Theta_2) \lesssim \sqrt{\frac{Rp}{T}}$$

and our testing procedure is minimax rate optimal up to a possible loss of order $\sqrt{R}$ and a log term.

Remark 3. *We can construct a change-point estimator based on he statistic $\mathcal{G}(t)$ defined in* (19)*:*

$$\widehat{\tau} = \arg\max_{t \in \mathcal{D}_T} \mathcal{G}(t).$$

The consistency of the change-point localization can be proven under the condition that $\mathcal{E}(\tau, \Theta_1, \Theta_2) \gtrsim \sqrt{Rp(\log T)/T}$. Note that using the same technique as in Theorem 2, it can be shown that the consistent estimation is impossible if $\mathcal{E}(\tau, \Theta_1, \Theta_2) \lesssim \sqrt{p/T}$ for $p \lesssim \sqrt{T}$.

3 Simulations

We suppose that an eventual change-point is located within the interval $[Th, T - Th]$, where $p/T < h < 1/2$ is given. The process is stationary within the intervals $[1, Th]$ and $[T - Th, T]$ that will be used for calibration of the quantiles and estimation of the transition matrices. We have implemented the testing procedure based on the test statistic $\mathcal{G}(t)$ defined in (19)

$$\mathcal{G}(t) = \frac{t}{T}\Big(\varphi_1^t(\mathfrak{X}, \widehat{\Theta}_2^t) - \varphi_1^t(\mathfrak{X}, \widehat{\Theta}_1^t)\Big) + \frac{T-t}{T}\Big(\varphi_2^t(\mathfrak{X}, \widehat{\Theta}_1^t) - \varphi_2^t(\mathfrak{X}, \widehat{\Theta}_2^t)\Big)$$

with the corresponding test

$$\psi_\alpha(\mathfrak{X}) = \mathbf{1}\left\{\max_{t \in \mathcal{T}} \frac{\mathcal{G}(t)}{q_{\alpha,t}} > 1\right\},$$

where $\mathcal{T}$ is a grid approximating the set of possible change-points $\mathcal{D}_T \subseteq [Th, T - Th]$.

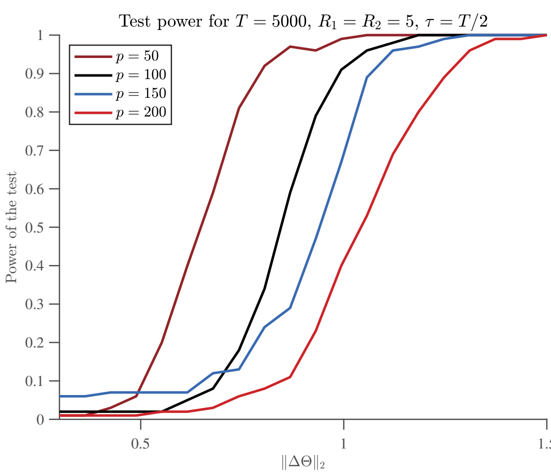

Figure 1: Test power depending on the dimension of transition matrices

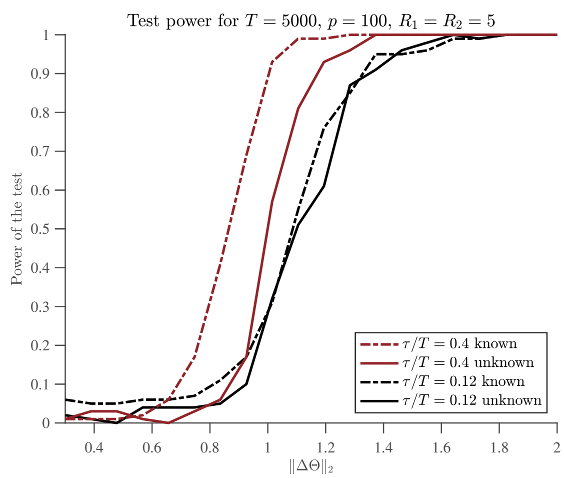

Figure 2: Test power for known or unknown location of the change-point

To calculate the test statistic $\mathcal{G}(t)$ we need to find the solutions $\widehat{\Theta}_1^t$ and $\widehat{\Theta}_2^t$ of the SDPs (11) and (12). The estimation quality of the transition matrices depends on the choice of constants in the regularization parameters $\lambda_1^t := \tilde{c}_1\sqrt{p/t}$ and $\lambda_2^t := \tilde{c}_2\sqrt{p/(T-t)}$. We tune the constants $\tilde{c}_i$ by cross-validation in the following way. We estimate the transition matrices Θ_1 and Θ_2 using, respectively, the first and the last $\lfloor \delta Th \rfloor$ observations $X_1, \ldots, X_{\lfloor \delta Th \rfloor}$ and $X_{T - \lfloor Th \rfloor}, \ldots, X_{T - \lfloor (1-\delta)Th \rfloor}$, where $\delta \in (0,1)$ is chosen a priori. Let

$\widehat{\Theta}_{1,\tilde{c}_1}$ and $\widehat{\Theta}_{2,\tilde{c}_2}$ be the corresponding solutions of SDPs (11) and (12) with the regularization constants $\tilde{c}_1$ and $\tilde{c}_2$ varying within some appropriately chosen grid. The optimal constants $\tilde{c}_1^*$ and $\tilde{c}_2^*$ will minimize the least-squares criteria for $\widehat{\Theta}_{1,\tilde{c}_1}$ and $\widehat{\Theta}_{2,\tilde{c}_2}$ calculated over the interval of size $(1-\delta)Th$:

$$\tilde{c}_1^* = \underset{\tilde{c}_1 \in \text{grid}}{\arg\min} \sum_{t=\lfloor \delta Th \rfloor}^{\lfloor Th \rfloor - 1} \|X_{t+1} - \widehat{\Theta}_{1,\tilde{c}_1} X_t\|_2^2, \quad \tilde{c}_2^* = \underset{\tilde{c}_2 \in \text{grid}}{\arg\min} \sum_{t=T-\lfloor (1-\delta)Th \rfloor}^{T-1} \|X_{t+1} - \widehat{\Theta}_{2,\tilde{c}_2} X_t\|_2^2.$$

To obtain the plug-in estimates $\widehat{\Theta}_1^t$ and $\widehat{\Theta}_2^t$ used in $\mathcal{G}(t)$, we take the regularization parameters λ_1^t and λ_2^t with the constants $\tilde{c}_1^*$ and $\tilde{c}_2^*$. The numerical solution to the SDPs is calculated using the accelerated gradient decent algorithm of Ji and Ye (2009).

The theoretical quantile of the change-point detection procedure depends on an unknown universal constant. In this study we use Monte-Carlo simulated quantiles $q_{\alpha,t}$. The details about the quantile simulation are provided in the Appendix.

We have performed 100 simulations of $T = 5000$ observations of $\mathrm{VAR}_p(\Theta_1, \Theta_2, \Sigma_Z)$ process with independent noise, $\Sigma_Z = \mathrm{id}_p$ and transition matrices Θ_1 and Θ_2 of the same rank $R_1 = R_2 = 5$. The significance level α is fixed to 0.05. We have chosen the intervals of size $Th = 5p$ for the estimation of matrices Θ_1 and Θ_2 with $\delta = 4/5$ for the constant calibration interval. We consider the problems of testing for a change-point at a given point $\tau \in \mathcal{D}_T$ and at an unknown point within the set $\mathcal{D}_T$. In the second case the dyadic grid defined in (18) was chosen as $\mathcal{T}$.

Fig. 1 shows the dependence of the test power on the Frobenius norm of the change in transition matrices, $\|\Delta\Theta\|_2$, for different values of p varying from 50 to 200. Here we test at a given point located in the middle. We see that the power decreases if the dimension increases and the detection problem becomes harder for the matrices of larger dimension. This confirms our theoretical detection rate $C\sqrt{Rp/T}$. In Fig. 2 we compare our test's performance for a given change-point versus an unknown change-point. In this simulation the dimension is fixed, $p = 100$. The bold line corresponds to the test power within the dyadic grid for the case of unknown change-point location. We see that the adaptive test performs quite well with respect to testing at a known change-point location. We can also see the impact of the change-point location to the detection rate: the detection is harder near the boundary of $\mathcal{D}_T$ and easier in the middle of the interval. Additional simulation results can be found in the Appendix.

References

Ahn, S. K. and Reinsel, G. C. (1988). Nested reduced-rank autogressive models for multiple time series. Journal of the American Statistical Association, 83(403):849–856.

Alquier, P., Bertin, K., Doukhan, P., and Garnier, R. (2020). High-dimensional VAR with low-rank transition. Stat. Comput., 30(4):1139–1153.

Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. The Annals of Statistics, 37(6B):4046 – 4087.

Bai, P., Safikhani, A., and Michailidis, G. (2020). Multiple Change Points Detection in Low Rank and Sparse High Dimensional Vector Autoregressive Models. IEEE Transactions on Signal Processing, 68:3074–3089.

Bai, P., Safikhani, A., and Michailidis, G. (2022). Multiple change point detection in reduced rank high dimensional vector autoregressive models. Journal of the American Statistical Association, 0(ja):1–42.

Basseville, M. and Nikiforov, I. V. (1993). Detection of Abrupt Changes: Theory and Application. Prentice-Hall, Inc., USA.

Basu, S., Li, X., and Michailidis, G. (2019). Low rank and structured modeling of high-dimensional vector autoregressions. IEEE Transactions on Signal Processing, 67(5):1207–1222.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. The Annals of Statistics, 43(4):1535 – 1567.

Carpentier, A. and Nickl, R. (2015). On signal detection and confidence sets for low rank inference problems. Electronic Journal of Statistics, 9(2):2675 – 2688.

Fan, J., Lv, J., and Qi, L. (2010). Sparse High Dimensional Models in Economics. SSRN Electronic Journal.

Gorrostieta, C., Fiecas, M., Ombao, H., Burke, E., and Cramer, S. (2013). Hierarchical vector auto-regressive models and their applications to multi-subject effective connectivity. Frontiers in Computational Neuroscience, 7.

Ingster, Y. I. and Suslina, I. A. (2003). Nonparametric Goodness-of-Fit Testing Under Gaussian Models. Springer New York.

Ji, S. and Ye, J. (2009). An accelerated gradient method for trace norm minimization. In International Conference on Machine Learning, pages 457 – 464. ACM.

Khan, F., Saeed, A., and Ali, S. (2020). Modelling and forecasting of new cases, deaths and recover cases of COVID-19 by using Vector Autoregressive model in Pakistan. Chaos, Solitons & Fractals, 140:110189.

Liu, H., Gao, C., and Samworth, R. J. (2021). Minimax rates in sparse, high-dimensional change point detection. The Annals of Statistics, 49(2):1081–1112.

Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. Springer.

Melnyk, I. and Banerjee, A. (2016). Estimating structured vector autoregressive models. In Balcan, M. F. and Weinberger, K. Q., editors, Proceedings of The 33rd International Conference on Machine Learning, volume 48 of Proceedings of Machine Learning Research, pages 830–839, New York, New York, USA. PMLR.

Michel, C. M. and Murray, M. M. (2012). Towards the utilization of eeg as a brain imaging tool. NeuroImage, 61(2):371–385.

Negahban, S. and Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. Ann. Statist., 39(2):1069–1097.

Reinsel, G. C. and Velu, R. P. (1998). Multivariate Reduced-Rank Regression: Theory and Applications. ISBN: 9781475728538 OCLC: 1159215349.

Rump, S. M. (2018). Estimates of the determinant of a perturbed identity matrix. Linear Algebra and its Applications, 558:101–107.

Safikhani, A. and Shojaie, A. (2022). Joint structural break detection and parameter estimation in high-dimensional nonstationary var models. Journal of the American Statistical Association, 117(537):251–264.

She, Y., He, Y., Li, S., and Wu, D. (2015). Joint association graph screening and decomposition for large-scale linear dynamical systems. IEEE Transactions on Signal Processing, 63(2):389–401.

Stock, J. H. and Watson, M. W. (2001). Vector autoregressions. Journal of Economic Perspectives, 15(4):101–115.

Tippett, M. K., Cohn, S. E., Todling, R., and Marchesin, D. (2000). Conditioning of the stable, discrete-time lyapunov operator. SIAM Journal on Matrix Analysis and Applications, 22(1):56–65.

Velu, R. P., Reinsel, G. C., and Wichern, D. W. (1986). Reduced rank models for multiple time series. Biometrika, 73(1):105–118.

Vershynin, R. (2018). High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wang, D. and Tsay, R. S. (2021). Rate-optimal robust estimation of high-dimensional vector autoregressive models.

Wang, D., Yu, Y., Rinaldo, A., and Willett, R. (2019). Localizing changes in high-dimensional vector autoregressive processes.

Wild, B., Eichler, M., Friederich, H.-C., Hartmann, M., Zipfel, S., and Herzog, W. (2010). A graphical vector autoregressive modelling approach to the analysis of electronic diary data. BMC Medical Research Methodology, 10(1):28.

A Definitions from minimax testing theory

Let $\mathfrak{X} = (X_0, \ldots, X_T)$ be observed data satisfying $\mathrm{VAR}_p(\Theta_1, \Theta_2, \Sigma_Z)$ model (3) with the change at point τ and the transition matrices Θ_1 and Θ_2 before and after the change. Denote by $\mathrm{P}^{\tau}_{\Theta_1,\Theta_2}$ the distribution of $\mathfrak{X}$. Let $\psi : \mathfrak{X} \to \{0, 1\}$ be a test for the presence of a change. Define the testing errors of $\psi(\mathfrak{X})$. The type I error is given by

$$\alpha(\psi) = \sup_{(\Theta_1,\Theta_2)\in\mathcal{V}_{p,0}} \mathrm{P}_{\Theta_1,\Theta_2}\{\psi(\mathfrak{X}) = 1\},$$

the type II error of testing at the given point τ is defined as

$$\beta(\psi, \mathcal{R}^{\tau}_{p,T}) = \sup_{(\Theta_1,\Theta_2)\in\mathcal{V}_{p,\tau}(\mathcal{R}^{\tau}_{p,T})} \mathrm{P}^{\tau}_{\Theta_1,\Theta_2}\{\psi(\mathfrak{X}) = 0\},$$

and the type II error of testing at an unknown change-point $\tau \in \mathcal{D}_T$ is defined as

$$\beta(\psi, \mathcal{R}_{p,T}) = \sup_{\tau\in\mathcal{D}_T} \sup_{(\Theta_1,\Theta_2)\in\mathcal{V}_{p,\tau}(\mathcal{R}_{p,T})} \mathrm{P}^{\tau}_{\Theta_1,\Theta_2}\{\psi(\mathfrak{X}) = 0\}.$$

Let $\alpha \in (0, 1)$ be a given significance level. Denote by Ψ_α the set of all tests of level at most α:

$$\Psi_\alpha = \{\psi : \ \alpha(\psi) \le \alpha\}.$$

It is important to know what are the conditions on the jump matrix $\Delta\Theta$ and the radius $\mathcal{R}_{p,T}$ that allow to detect the change-point with a given significance level and reasonable type II error. These conditions are formulated in terms of the minimax separation rate.

Definition 1. *Let $\alpha, \beta \in (0, 1)$ be given. We say that the radius $\mathcal{R}^*_{p,T}$ is (α, β)-minimax detection boundary in problem of testing the hypothesis of no change against the alternative $\mathcal{V}_p, \tau(\mathcal{R}_{p,\mathcal{D}_T})$ if*

$$\mathcal{R}^*_{p,\mathcal{D}_T} = \inf_{\psi\in\Psi_\alpha} \mathcal{R}_{p,\mathcal{D}_T}(\alpha, \psi),$$

where

$$\mathcal{R}_{p,\mathcal{D}_T}(\alpha, \psi) = \inf\Big\{\mathcal{R} > 0 : \ \beta(\psi, \mathcal{R}) \le \beta\Big\}.$$

The minimax detection boundary is often written as the product $\mathcal{R}^*_{p,\mathcal{D}_T} = C\varphi_{p,\mathcal{D}_T}$, where $\varphi_{p,\mathcal{D}_T}$ is called *minimax detection rate* and C is a constant independent of p and T. We say that the radius $\mathcal{R}_{p,\mathcal{D}_T}$ satisfies the upper bound condition if there exists a constant $C^* > 0$ and a test $\psi^* \in \Psi_\alpha$ such that $\forall C > C^*$ $\beta(\psi^*, \mathcal{R}_{p,\mathcal{D}_T}) \le \beta$. We say that $\mathcal{R}_{p,\mathcal{D}_T}$ satisfies the lower bound condition if for any $0 < C \le C_*$ there is no test of level α with type II error smaller than β. Our goal is to find the minimax detection rate $\varphi_{p,\mathcal{D}_T}$ and two constants C_* and C^* such that

$$C_*\varphi_{p,\mathcal{D}_T} \le \mathcal{R}^*_{p,\mathcal{D}_T} \le C^*\varphi_{p,\mathcal{D}_T}.$$

B Proof of the upper bound

Denote

$$\mathcal{F}_1(t) = \varphi_1^t(\mathfrak{X}, \widehat{\Theta}_2^t) - \varphi_1^t(\mathfrak{X}, \widehat{\Theta}_1^t) \quad \text{and} \quad \mathcal{F}_2(t) = \varphi_2^t(\mathfrak{X}, \widehat{\Theta}_1^t) - \varphi_2^t(\mathfrak{X}, \widehat{\Theta}_2^t)$$

Lemma 2 (Type I error)**.** *Let $\alpha \in (0,1)$ be a given significance level. Assume that $T/2 \ge p > \log(8|\mathcal{T}|/\alpha)$. If we solve (11)–(12) with λ_1^t and λ_2^t defined in (13)–(14), then the type I error of test (16) is bounded by α.*

Proof. By the union bound, the type I error of the test ψ_α can be bounded as follows,

$$\alpha(\psi_\alpha) = \sup_{(\Theta_1,\Theta_2)\in\mathcal{V}_p(\tau,0)} \mathrm{P}_{\Theta_1,\Theta_2}\left(\max_{t\in\mathcal{T}} \frac{\mathcal{F}(t)}{H_{\alpha,t}} > 1\right) \le \sum_{t\in\mathcal{T}} \sup_{(\Theta_1,\Theta_2)\in\mathcal{V}_p(\tau,0)} \mathrm{P}_{\Theta_1,\Theta_2}\Big\{\mathcal{F}(t) \ge H_{\alpha,t}\Big\}.$$

Denote by $\Theta := \Theta_1 = \Theta_2$ the transition matrix of the process under the null hypothesis. To provide an α-level test, we need to show that the type I error of testing at each point $t \in \mathcal{T}$ is bounded by $\alpha/|\mathcal{T}|$:

$$\sup_{\Theta\in\mathcal{M}_p(R,\gamma)} \mathrm{P}_\Theta\Big\{\mathcal{F}(t) \ge H_{\alpha,t}\Big\} \le \alpha/|\mathcal{T}|,$$

For any two matrices A and B we have

$$\begin{aligned}
&\sum_{i=t}^{T-1}\Big(\|X_{i+1} - AX_i\|_2^2 - \|X_{i+1} - BX_i\|_2^2\Big) \\
&\quad= \sum_{i=t}^{T-1}\Big(\|(\Theta - A)X_i + Z_{i+1}\|_2^2 - \|(\Theta - B)X_i + Z_{i+1}\|_2^2\Big) \\
&\quad= \sum_{i=t}^{T-1}\Big(\|(\Theta - A)X_i\|_2^2 - \|(\Theta - B)X_i\|_2^2\Big) + 2\sum_{i=t}^{T-1}\Big(\langle(\Theta - A)X_i, Z_{i+1}\rangle - \langle(\Theta - B)X_i, Z_{i+1}\rangle\Big) \\
&\quad= \|(\Theta - A)\mathfrak{X}_{\ge t}\|_2^2 - \|(\Theta - B)\mathfrak{X}_{\ge t}\|_2^2 + 2\langle(B - A)\mathfrak{X}_{\ge t}, \mathfrak{Z}_{\ge t}\rangle. \qquad (20)
\end{aligned}$$

A similar calculation provides

$$\begin{aligned}
&\sum_{i=0}^{t-1}\left(\|X_{i+1} - AX_i\|_2^2 - \|X_{i+1} - BX_i\|_2^2\right) \\
&\qquad\qquad= \|(\Theta - A)\mathfrak{X}_{<t}\|_2^2 - \|(\Theta - B)X_i\|_2^2 + 2\langle(B - A)\mathfrak{X}_{\ge t}, \mathfrak{Z}_{\ge t}\rangle. \qquad (21)
\end{aligned}$$

Since the test statistic $\mathcal{F}(t)$ depends on the location of t with respect to the boundary, we will consider four following cases: (a) $t \geq T/2$ and $T-t \leq p$, $t \geq p$, (b) $t \geq T/2$ and $T-t \geq p$, (c) $t \leq T/2$ and $t \leq p$, $T-t \geq T-p$, and (d) $t \leq T/2$ and $t \geq p$.

Note that the assumption $p \leq T/2$ is important. It implies, for example, that $t \geq p$ if $T-t \leq p$ and $t \geq T/2$.

By symmetry, it is sufficient to consider the cases (a) and (b). Consider the case (a) when $t \geq T/2 \geq p$ and $T-t \leq p$. Using (20) and (21) we have

$$\begin{aligned}\sum_{i=t}^{T-1} &\left(\|X_{i+1} - \widehat{\Theta}_1^t X_i\|_2^2 - \|X_{i+1} - \widehat{\Theta}_2^t X_i\|_2^2\right) \\ &= \|(\Theta - \widehat{\Theta}_1^t)\mathfrak{X}_{\geq t}\|_2^2 - \|(\Theta - \widehat{\Theta}_2^t)\mathfrak{X}_{\geq t}\|_2^2 + 2\left\langle (\widehat{\Theta}_2^t - \widehat{\Theta}_1^t)\mathfrak{X}_{\geq t}, \mathfrak{Z}_{\geq t}\right\rangle.\end{aligned}$$

Thus $\mathcal{F}(t)$ can be written as the sum of three terms,

$$\begin{aligned}\mathcal{F}(t) = \frac{1}{T-t}&\left(\|(\Theta - \widehat{\Theta}_1^t)\mathfrak{X}_{\geq t}\|_2^2 - \|(\Theta - \widehat{\Theta}_2^t)\mathfrak{X}_{\geq t}\|_2^2\right) + \frac{2}{T-t}\left\langle (\widehat{\Theta}_2^t - \widehat{\Theta}_1^t)\mathfrak{X}_{\geq t}, \mathfrak{Z}_{\geq t}\right\rangle. \\ &+ \lambda_2^t\left(\|\widehat{\Theta}_1^t \mathfrak{X}_{\geq t}\|_* - \|\Theta\mathfrak{X}_{\geq t}\|_* + \|\Theta\mathfrak{X}_{\geq t}\|_* - \|\widehat{\Theta}_2^t\mathfrak{X}_{\geq t}\|_*\right) := A + B + C. \end{aligned} \tag{22}$$

Introduce the following random events,

$$\mathcal{A} = \left\{\|\mathfrak{Z}_{\geq t}\|_{\text{op}} \leq c_1^* \sqrt{\|\Sigma_Z\|_{\text{op}}}\left(\sqrt{T-t} + \sqrt{p} + \sqrt{\log\frac{8|\mathcal{T}|}{\alpha}}\right)\right\},$$

$$\mathcal{B}_1 = \left\{\|\widehat{\Theta}_1^t - \Theta\|_2^2 \leq c^* \frac{\kappa^2(\Sigma)}{(1-\gamma)^2}\frac{Rp}{t}\right\} \quad \text{and} \quad \mathcal{B}_2 = \left\{\|\widehat{\Theta}_2^t - \Theta\|_2^2 \leq c^* \frac{\kappa^2(\Sigma)}{(1-\gamma)^2}\frac{Rp}{T-t}\right\}$$

Lemma 9 implies that $\mathrm{P}(\mathcal{A}) \geq 1 - \dfrac{\alpha}{4|\mathcal{T}|}$ and Proposition 5 gives $\mathrm{P}(\mathcal{B}_i) \geq 1 - c_2\exp(-c_3 p) \geq 1 - \dfrac{\alpha}{4|\mathcal{T}|}$ for sufficiently large p. Thus, we have

$$\mathrm{P}_\Theta\{\mathcal{F}(t) > H_{\alpha,t}\} \leq \mathrm{P}_\Theta\Big\{\{\mathcal{F}(t) > H_{\alpha,t}\} \cap \{\mathcal{A} \cap \mathcal{B}_1 \cap \mathcal{B}_2\}\Big\} + \mathrm{P}(\mathcal{A}^c) + \mathrm{P}_\Theta(\mathcal{B}_1^c) + \mathrm{P}_\Theta(\mathcal{B}_2^c).$$

To show that the type I error is bounded by α, we have to show that the first probability is bounded by $\alpha/(4|\mathcal{T}|)$.

Consider first the scalar product term B in (22). We have

$$\begin{aligned}\left(\widehat{\Theta}_2^t - \widehat{\Theta}_1^t\right)\mathfrak{X}_{\geq t} &= \left(\Theta - \widehat{\Theta}_1^t\right)\mathfrak{X}_{\geq t} + \left(\widehat{\Theta}_2^t - \Theta\right)\mathfrak{X}_{\geq t} \\ &= \left(\Theta - \widehat{\Theta}_1^t\right)\mathfrak{X}_{\geq t} - \mathbf{Pr}_{\Theta\mathfrak{X}_{\geq t}}\left[\Theta\mathfrak{X}_{\geq t}\right] + \mathbf{Pr}_{\Theta\mathfrak{X}_{\geq t}}\left[\widehat{\Theta}_2^t\mathfrak{X}_{\geq t}\right] + \mathbf{Pr}^{\perp}_{\Theta\mathfrak{X}_t}\left[\widehat{\Theta}_2^t\mathfrak{X}_{\geq t}\right].\end{aligned}$$

Using this identity and the Hölder inequality, we obtain

$$\begin{aligned}\left\langle\left(\widehat{\Theta}_2^t - \widehat{\Theta}_1^t\right)\mathfrak{X}_{\geq t}, \mathfrak{Z}_{\geq t}\right\rangle &\leq \|\left(\widehat{\Theta}_1^t - \Theta\right)\mathfrak{X}_{\geq t}\|_*\|\mathfrak{Z}_{\geq t}\|_{\text{op}} \\ &+ \|\mathbf{Pr}_{\Theta\mathfrak{X}_{\geq t}}\left[\left(\Theta - \widehat{\Theta}_2^t\right)\mathfrak{X}_{\geq t}\right]\|_*\|\mathfrak{Z}_{\geq t}\|_{\text{op}} + \|\mathbf{Pr}^{\perp}_{\Theta\mathfrak{X}_{\geq t}}\left[\widehat{\Theta}_2^t\mathfrak{X}_{\geq t}\right]\|_*\|\mathfrak{Z}_{\geq t}\|_{\text{op}}.\end{aligned}$$

By Lemma 9, taking into account the fact that $p \geq T-t$ and $p > \log(8|\mathcal{T}|/\alpha)$, we obtain that with probability at least $1-\alpha/(4|\mathcal{T}|)$,

$$\frac{2}{T-t}\|\mathfrak{Z}_{\geq t}\|_{\mathrm{op}} \leq 6c_1^* \sqrt{\|\Sigma_Z\|_{\mathrm{op}}}\frac{\sqrt{p}}{T-t} = \lambda_2^t.$$

Thus, given $\mathcal{A}$, we have

$$B \leq \lambda_2^t \|(\widehat{\Theta}_1^t - \Theta)\mathfrak{X}_{\geq t}\|_* + \lambda_2^t \|\mathbf{Pr}_{\Theta\mathfrak{X}_{\geq t}}\left[(\Theta - \widehat{\Theta}_2^t)\mathfrak{X}_{\geq t}\right]\|_* + \|\mathbf{Pr}^{\perp}_{\Theta\mathfrak{X}_{\geq t}}\left[\widehat{\Theta}_2^t \mathfrak{X}_{\geq t}\right]\|_*$$

Using the triangle inequality, $\|\widehat{\Theta}_1^t\mathfrak{X}_{\geq t}\|_* - \|\Theta\mathfrak{X}_{\geq t}\|_* \leq \|\left(\widehat{\Theta}_1^t - \Theta\right)\mathfrak{X}_{\geq t}\|_*$ and the fact that by (41), $\|\Theta\mathfrak{X}_{\geq t}\|_* - \|\widehat{\Theta}_2^t\mathfrak{X}_{\geq t}\|_* \leq \|\mathbf{Pr}_{\Theta\mathfrak{X}_{\geq t}}\left[\left(\Theta - \widehat{\Theta}_2^t\right)\mathfrak{X}_{\geq t}\right]\|_* - \|\mathbf{Pr}^{\perp}_{\Theta\mathfrak{X}_{\geq t}}\left[\left(\Theta - \widehat{\Theta}_2^t\right)\mathfrak{X}_{\geq t}\right]\|_*$, we obtain for the term C,

$$C \leq \lambda_2^t \|\left(\widehat{\Theta}_1^t - \Theta\right)\mathfrak{X}_{\geq t}\|_* + \lambda_2^t\|\mathbf{Pr}_{\Theta\mathfrak{X}_{\geq t}}\left[\left(\Theta - \widehat{\Theta}_2^t\right)\mathfrak{X}_{\geq t}\right]\|_* - \lambda_2^t\|\mathbf{Pr}^{\perp}_{\Theta\mathfrak{X}_{\geq t}}\left[\widehat{\Theta}_2^t\mathfrak{X}_{\geq t}\right]\|_*.$$

Gathering these bounds, we obtain that given $\mathcal{A}$,

$$\begin{aligned}\mathcal{F}(t) \leq &\frac{1}{T-t}\|\left(\Theta - \widehat{\Theta}_1^t\right)\mathfrak{X}_{\geq t}\|_2^2 - \frac{1}{T-t}\|\left(\Theta - \widehat{\Theta}_2^t\right)\mathfrak{X}_{\geq t}\|_2^2 \\ &+ 2\lambda_2^t\|\left(\widehat{\Theta}_1^t - \Theta\right)\mathfrak{X}_{\geq t}\|_* + 2\lambda_2^t\|\mathbf{Pr}_{\Theta\mathfrak{X}_{\geq t}}\left[\left(\Theta - \widehat{\Theta}_2^t\right)\mathfrak{X}_{\geq t}\right]\|_*. \end{aligned} \tag{23}$$

Using Lemma 7, we get the bound

$$\|\widehat{\Theta}_1^t - \Theta\|_* \leq \|\mathbf{Pr}^{\perp}_{\Theta}\left[\widehat{\Theta}_1^t - \Theta\right]\|_* + \|\mathbf{Pr}_{\Theta}\left[\widehat{\Theta}_1^t - \Theta\right]\|_* \leq 6\|\mathbf{Pr}_{\Theta}\left[\widehat{\Theta}_1^t - \Theta\right]\|_*.$$

Taking into account that $\mathrm{rank}\left(\mathbf{Pr}_{\Theta}\left[\widehat{\Theta}_1^t - \Theta\right]\right) \leq 2\mathrm{rank}(\Theta) = 2R$, using $\|A\|_*^2 \leq \mathrm{rank}(A)\|A\|_2^2$ and the trivial inequality $2ab \leq a^2 + b^2$, we get

$$\begin{aligned}2\lambda_2^t\|\left(\widehat{\Theta}_1^t - \Theta\right)\mathfrak{X}_{\geq t}\|_* &\leq 2\lambda_2^t\|\widehat{\Theta}_1^t - \Theta\|_*\|\mathfrak{X}_{\geq t}\|_{\mathrm{op}} \\ &\leq 12\lambda_2^t\|\mathbf{Pr}_{\Theta}\left[\widehat{\Theta}_1^t - \Theta\right]\|_*\|\mathfrak{X}_{\geq t}\|_{\mathrm{op}} \\ &\leq 12\lambda_2^t\sqrt{2R}\|\widehat{\Theta}_1^t - \Theta\|_2\|\mathfrak{X}_{\geq t}\|_{\mathrm{op}} \\ &\leq 2\left(\lambda_2^t\right)^2 R(T-t) + \frac{36}{T-t}\|\widehat{\Theta}_1^t - \Theta\|_2^2\|\mathfrak{X}_{\geq t}\|_{\mathrm{op}}^2.\end{aligned}$$

Thus by Proposition 5, given $\mathcal{B}_1$ with probability at least $1 - c_2\exp(-c_3 p) \geq 1 - \alpha/(4|\mathcal{T}|)$, we have

$$2\lambda_2^t\|\left(\widehat{\Theta}_1^t - \Theta\right)\mathfrak{X}_{\geq t}\|_* \leq 2\left(\lambda_2^t\right)^2 R(T-t) + \frac{36c^*Rp}{t(T-t)}\frac{\kappa^2(\Sigma)}{(1-\gamma)^2}\|\mathfrak{X}_{\geq t}\|_{\mathrm{op}}^2$$

Since $\mathrm{rank}\left(\Theta\mathfrak{X}_t\right) \leq \mathrm{rank}(\Theta)$, we have, using the same argument as above, we have

$$\begin{aligned}2\lambda_2^t\|\mathbf{Pr}_{\Theta\mathfrak{X}_{\geq t}}\left[(\Theta - \widehat{\Theta}_2^t)\mathfrak{X}_{\geq t}\right]\|_* &\leq 2\lambda_2^t\sqrt{2R}\|(\Theta - \widehat{\Theta}_2^t)\mathfrak{X}_{\geq t}\|_2 \\ &\leq 2\left(\lambda_2^t\right)^2 R(T-t) + \frac{1}{T-t}\|(\Theta - \widehat{\Theta}_2^t)\mathfrak{X}_{\geq t}\|_2^2.\end{aligned}$$

Similarly, by Proposition 5, given $\mathcal{B}_2$ we have

$$\frac{1}{T-t}\|(\Theta-\widehat{\Theta}_1^t)\mathfrak{X}_{\geq t}\|_2^2 \leq \frac{c^* Rp}{t(T-t)}\frac{\kappa^2(\Sigma)}{(1-\gamma)^2}\|\mathfrak{X}_{\geq t}\|_{\mathrm{op}}^2$$

Finally, for sufficiently large p, with the probability at least $1-\alpha/(4|\mathcal{T}|)$, (23) and the above inequalities and the definition of λ_2^t imply, given the event $\mathcal{A}\cap\mathcal{B}_1\cap\mathcal{B}_2$,

$$\mathcal{F}(t) \leq \frac{37c^* Rp}{t(T-t)}\frac{\kappa^2(\Sigma)}{(1-\gamma)^2}\|\mathfrak{X}_{\geq t}\|_{\mathrm{op}}^2 + \frac{144(c_1^*)^2\|\Sigma_Z\|_{\mathrm{op}} Rp}{T-t}.$$

Thus, for any $\Theta\in\mathcal{M}_p(R,\gamma)$,

$$\begin{aligned}
\mathrm{P}_\Theta&\Big\{\{\mathcal{F}(t)>H_{\alpha,t}\}\cap\{\mathcal{A}\cap\mathcal{B}_1\cap\mathcal{B}_2\}\Big\}\\
&\leq \mathrm{P}\left\{\frac{37c^* Rp}{t(T-t)}\frac{\kappa^2(\Sigma)}{(1-\gamma)^2}\|\mathfrak{X}_{\geq t}\|_{\mathrm{op}}^2 + \frac{144(c^*)^2 Rp\|\Sigma_Z\|_{\mathrm{op}}}{T-t} > H_{\alpha,t}\right\}\\
&\leq \mathrm{P}\left\{\frac{1}{T-t}\|\mathfrak{X}_{\geq t}\|_{\mathrm{op}}^2 \geq 2\|\Sigma\|_{\mathrm{op}}\Big(1+\frac{c_0}{1-\gamma}\frac{p+\log(8|\mathcal{T}|/\alpha)}{T-t}\Big)\right\} \leq \frac{\alpha}{4|\mathcal{T}|},
\end{aligned}$$

where we used Lemma 8 with $\delta=\dfrac{\alpha}{8|\mathcal{T}|}$, $T-t\leq p$, $t\geq p$ and the fact that for a sufficiently large universal constant C depending on c^*, c_0 and c_1^*, we have the bound

$$\begin{aligned}
H_{\alpha,t} &= C\max\Big(\|\Sigma_Z\|_{\mathrm{op}}, \frac{\kappa^2(\Sigma)}{(1-\gamma)^2}\|\Sigma\|_{\mathrm{op}}\Big)\frac{Rp}{Tq^2(t/T)}\left(1+\frac{1}{1-\gamma}\frac{p+\log(8|\mathcal{T}|/\alpha)}{T}\right)\\
&\geq \frac{144(c_1^*)^2 Rp\|\Sigma_Z\|_{\mathrm{op}}}{T-t} + \frac{74c^* Rp}{t}\frac{\kappa^2(\Sigma)}{(1-\gamma)^2}\|\Sigma\|_{\mathrm{op}}\Big(1+\frac{c_0}{1-\gamma}\frac{p+\log(8|\mathcal{T}|/\alpha)}{T-t}\Big).
\end{aligned}$$

Let us turn to the case (b) of the candidate change-point located far from the endpoints of the interval, $T-t\geq p$ and $t\geq T/2$. Calculations similar to the case (a) imply

$$\begin{aligned}
\mathcal{F}(t) &= \frac{1}{T-t}\sum_{i=t}^{T-1}\left(\|\left(\Theta-\widehat{\Theta}_1^t\right)X_i\|_2^2 - \|\left(\Theta-\widehat{\Theta}_2^t\right)X_i\|_2^2\right) + \frac{2}{T-t}\left\langle\left(\widehat{\Theta}_2^t-\widehat{\Theta}_1^t\right), \mathfrak{Z}_{\geq t}\mathfrak{X}_{\geq t}^{\mathrm{T}}\right\rangle\\
&\quad + \lambda_2^t\left(\|\widehat{\Theta}_1^t\|_* - \|\Theta\|_* + \|\Theta\|_* - \|\left(\widehat{\Theta}_2^t\right)\|_*\right) = A+B+C.
\end{aligned}$$

Introduce the random events

$$\mathcal{A}' = \left\{\frac{1}{T-t}\|\mathfrak{Z}_{\geq t}\mathfrak{X}_{\geq t}^{\mathrm{T}}\|_{\mathrm{op}} \leq \frac{c_2^*\|\Sigma\|_{\mathrm{op}}}{1-\gamma}\sqrt{\frac{p}{T-t}}\right\} \quad \text{and} \quad \mathcal{D} = \left\{\frac{1}{T-t}\sigma_{\min}(\mathfrak{X}_{\geq t}^{\mathrm{T}}\mathfrak{X}_{\geq t}) \geq \frac{\sigma_{\min}(\Sigma_2)}{4}\right\}.$$

Lemma 10 and Lemma 8 imply that $\mathrm{P}(\mathcal{A}')\geq 1-\alpha/(8|\mathcal{T}|)$ and $\mathrm{P}(\mathcal{D}')\geq 1-\alpha/(8|\mathcal{T}|)$ for sufficiently large such that$p\geq a_1\log(8|\mathcal{T}|/a_2)$ for some universal constants $a_1,a_2>0$. Similarly to the case (a), we have

$$\mathrm{P}_\Theta\{\mathcal{F}(t)>H_{\alpha,t}\} \leq \mathrm{P}_\Theta\Big\{\{\mathcal{F}(t)>H_{\alpha,t}\}\cap\mathcal{A}'\cap\mathcal{B}_1\cap\mathcal{B}_2\cap\mathcal{D}\Big\}+\mathrm{P}_\Theta((\mathcal{A}')^c)+\mathrm{P}_\Theta(\mathcal{B}_1^c)+\mathrm{P}_\Theta(\mathcal{B}_2^c)+\mathrm{P}_\Theta(\mathcal{D}^c).$$

We have to show that the first probability is bounded by $\alpha/(4|\mathcal{T}|)$. As in the previous case, we get using the Hölder inequality,

$$\left\langle \left(\widehat{\Theta}_2^t - \widehat{\Theta}_1^t\right), \mathfrak{Z}_{\geq t}\mathfrak{X}_{\geq t}^{\mathrm{T}} \right\rangle \leq \|\widehat{\Theta}_1^t - \Theta\|_* \|\mathfrak{Z}_{\geq t}\mathfrak{X}_{\geq t}^{\mathrm{T}}\|_{\mathrm{op}} + \|\mathbf{Pr}_\Theta\left(\Theta - \widehat{\Theta}_2^t\right)\|_* \|\mathfrak{Z}_{\geq t}\mathfrak{X}_{\geq t}^{\mathrm{T}}\|_{\mathrm{op}} + \|\mathbf{Pr}_\Theta^\perp\left(\widehat{\Theta}_2^t\right)\|_* \|\mathfrak{Z}_{\geq t}\mathfrak{X}_{\geq t}^{\mathrm{T}}\|_{\mathrm{op}}.$$

Remind that $\lambda_2^t = \dfrac{2c_2^*\|\Sigma\|_{\mathrm{op}}}{1-\gamma}\sqrt{\dfrac{p}{T-t}}$. Lemma 10 implies that for sufficiently large p, with the probability at least $1 - k_1 \exp(-k_2 p)$,

$$\frac{2}{T-t}\|\mathfrak{Z}_{\geq t}\mathfrak{X}_{\geq t}^{\mathrm{T}}\|_{\mathrm{op}} \leq \frac{2c_2^*\|\Sigma\|_{\mathrm{op}}}{1-\gamma}\sqrt{\frac{p}{T-t}} = \lambda_2^t.$$

Thus, given the event $\mathcal{A}'$, we get $B \leq \lambda_2^t\|\widehat{\Theta}_1^t - \Theta\|_* + \lambda_2^t\|\mathbf{Pr}_\Theta\left[\Theta - \widehat{\Theta}_2^t\right]\|_* + \lambda_2^t\|\mathbf{Pr}_\Theta^\perp\left[\widehat{\Theta}_2^t\right]\|_*$. Similarly, for the term C, using the triangle inequality and the inequality (41) implying that $\|\Theta\|_* - \|\widehat{\Theta}_2^t\|_* \leq \|\mathbf{Pr}_\Theta\left[\Theta - \widehat{\Theta}_2^t\right]\|_* - \|\mathbf{Pr}_\Theta^\perp\left[\Theta - \widehat{\Theta}_2^t\right]\|_*$, we obtain

$$C \leq \lambda_2^t\|\widehat{\Theta}_1^t - \Theta\|_* + \lambda_2^t\|\mathbf{Pr}_\Theta\left[\Theta - \widehat{\Theta}_2^t\right]\|_* - \lambda_2^t\|\mathbf{Pr}_\Theta^\perp\left[\widehat{\Theta}_2^t\right]\|_*.$$

Thus, given $\mathcal{A}'$, we have

$$\mathcal{F}(t) \leq \frac{1}{T-t}\|\left(\Theta - \widehat{\Theta}_1^t\right)\mathfrak{X}_{\geq t}\|_2^2 - \frac{1}{T-t}\|\left(\Theta - \widehat{\Theta}_2^t\right)\mathfrak{X}_{\geq t}\|_2^2 + 2\lambda_2^t\|\widehat{\Theta}_1^t - \Theta\|_* + 2\lambda_2^t\|\mathbf{Pr}_\Theta\left(\Theta - \widehat{\Theta}_2^t\right)\|_*.$$

As in the case (a), using Lemma 7 and $2ab \leq a^2 + b^2$, we can bound the last two terms of this inequality

$$2\lambda_2^t\|\widehat{\Theta}_1^t - \Theta\|_* \leq 12\lambda_2^t\sqrt{2R}\|\widehat{\Theta}_1^t - \Theta\|_2 \leq \frac{9R\left(\lambda_2^t\right)^2}{\sigma_{\min}(\Sigma)} + 8\sigma_{\max}(\Sigma)\|\Theta - \widehat{\Theta}_1^t\|_2^2$$

and

$$2\lambda_2^t\|\mathbf{Pr}_\Theta\left(\Theta - \widehat{\Theta}_2^t\right)\|_* \leq 2\lambda_2^t\sqrt{2R}\|\widehat{\Theta}_2^t - \Theta\|_2 \leq \frac{8R\left(\lambda_2^t\right)^2}{\sigma_{\min}(\Sigma)} + \frac{\sigma_{\min}(\Sigma)}{4}\|\widehat{\Theta}_2^t - \Theta\|_2^2.$$

By Proposition 5, with probability at least $1 - \alpha/(4|\mathcal{T}|)$,

$$\frac{1}{T-t}\|\left(\Theta - \widehat{\Theta}_1^t\right)\mathfrak{X}_{\geq t}\|_2^2 \leq \frac{c^*Rp}{t(T-t)}\frac{\kappa^2(\Sigma)}{(1-\gamma)^2}\|\mathfrak{X}_{\geq t}\|_{\mathrm{op}}^2.$$

This implies that given $\mathcal{A}' \cap \mathcal{B}_1$

$$\mathcal{F}(t) \leq \frac{c^*Rp}{t(T-t)}\frac{\kappa^2(\Sigma)}{(1-\gamma)^2}\|\mathfrak{X}_{\geq t}\|_{\mathrm{op}}^2 - \frac{1}{T-t}\sigma_{\min}(\mathfrak{X}_{\geq t}^{\mathrm{T}}\mathfrak{X}_{\geq t})\|\Theta - \widehat{\Theta}_2^t\|_2^2 + \frac{17R\left(\lambda_2^t\right)^2}{\sigma_{\min}(\Sigma)} + 8\sigma_{\max}(\Sigma)\|\Theta - \widehat{\Theta}_1^t\|_2^2 + \frac{\sigma_{\min}(\Sigma)}{4}\|\widehat{\Theta}_2^t - \Theta\|_2^2.$$

Using the definition of λ_2^t, Lemma 8 and Proposition 5, we finally obtain that given $\mathcal{A}' \cap \mathcal{B}_1 \cap \mathcal{B}_2 \cap \mathcal{D}$

$$\begin{aligned}
\mathcal{F}(t) &\le \frac{c^* Rp}{t(T-t)} \frac{\kappa^2(\Sigma)}{(1-\gamma)^2} \|\mathfrak{X}_{\ge t}\|_{\mathrm{op}}^2 + \frac{68(c_2^*)^2 \|\Sigma\|_{\mathrm{op}}^2}{(1-\gamma)^2 \sigma_{\min}(\Sigma)} \frac{Rp}{T-t} + 8c^* \sigma_{\max}(\Sigma) \frac{\kappa^2(\Sigma)}{(1-\gamma)^2} \frac{Rp}{t} \\
&\le \frac{c^* Rp}{t(T-t)} \frac{\kappa^2(\Sigma)}{(1-\gamma)^2} \|\mathfrak{X}_{\ge t}\|_{\mathrm{op}}^2 + \max(c^*, 68(c_2^*)^2) \frac{\kappa^2(\Sigma)\|\Sigma\|_{\mathrm{op}}}{(1-\gamma)^2} \frac{Rp}{Tq^2(t/T)} \\
&\le C \frac{Rp}{Tq^2(t/T)} \frac{\kappa^2(\Sigma)}{(1-\gamma)^2} \left(\frac{1}{T} \|\mathfrak{X}_{\ge t}\|_{\mathrm{op}}^2 + \|\Sigma\|_{\mathrm{op}} \right),
\end{aligned}$$

where C is an absolute constant depending on c^* and c_2^*. Note that since $T - t > p$ and $T - t \le T/2$, we have the following bound for a sufficiently large universal constant C,

$$H_{\alpha,t} \ge C \frac{\kappa^2(\Sigma)\|\Sigma\|_{\mathrm{op}}}{(1-\gamma)^2} \frac{Rp}{Tq^2(t/T)} \left(1 + \frac{T-t}{T} \left(1 + \frac{c_0}{1-\gamma} \left(1 + \frac{\log(8|\mathcal{T}|/\alpha)}{T-t} \right) \right) \right).$$

Using this fact and Lemma 8 with $\delta = \dfrac{\alpha}{8|\mathcal{T}|}$, we obtain that for any $\Theta \in \mathcal{M}_p(R, \gamma)$,

$$\begin{aligned}
\mathrm{P}_\Theta \Big\{ \{\mathcal{F}(t) > H_{\alpha,t}\} \cap \{\mathcal{A}' \cap \mathcal{B}_1 \cap \mathcal{B}_2\} \Big\} &\le \mathrm{P} \left\{ C \frac{Rp}{Tq^2(t/T)} \frac{\kappa^2(\Sigma)}{(1-\gamma)^2} \left(\frac{1}{T} \|\mathfrak{X}_{\ge t}\|_{\mathrm{op}}^2 + \|\Sigma\|_{\mathrm{op}} \right) > H_{\alpha,t} \right\} \\
&\le \mathrm{P} \left\{ \frac{1}{T-t} \|\mathfrak{X}_{\ge t}\|_{\mathrm{op}}^2 \ge 2\|\Sigma\|_{\mathrm{op}} \left(1 + \frac{c_0}{1-\gamma} \left(1 + \frac{\log(8|\mathcal{T}|/\alpha)}{T-t} \right) \right) \right\} \le \frac{\alpha}{4|\mathcal{T}|}
\end{aligned}$$

and the lemma follows. □

Lemma 3 (Type II error). *Let $\beta \in (0,1)$ be given significance level such that $p > \max\Big(k_2^{-1}\log(4k_1/\beta)$, $c_3^{-1}\log(4c_2/\beta), \log(8/\beta)\Big)$, where k_1, k_2, c_2, c_3 are the constants from Proposition 5 and Lemma 10. Assume that*

$$q^2\left(\frac{\tau}{T}\right) \|\Theta_1 - \Theta_2\|_2^2 \ge \Xi(\Sigma_1, \Sigma_2, \Sigma, \gamma) \frac{Rp}{T} \left(1 + \frac{\log(8|\mathcal{T}|/\alpha) + \log(8/\beta)}{T} \right), \tag{24}$$

where, for some absolute constant C^*,

$$\Xi(\Sigma_1, \Sigma_2, \Sigma, \gamma) = \frac{C^*}{(1-\gamma)^3} \frac{\mathfrak{M}_1 \vee \mathfrak{M}_2}{\mathfrak{m}}$$

with

$$\mathfrak{M}_1 = \max\Big(\frac{\|\Sigma_2\|_{\mathrm{op}}^2}{\sigma_{\min}(\Sigma_Z) \wedge \sigma_{\min}(\Sigma_1)}, \frac{\|\Sigma_1\|_{\mathrm{op}}^2}{\sigma_{\min}(\Sigma_Z) \wedge \sigma_{\min}(\Sigma_2)} \Big) \tag{25}$$

$$\mathfrak{M}_2 = \max\Big(\kappa^2(\Sigma_1)\|\Sigma_2\|_{\mathrm{op}}, \kappa^2(\Sigma_2)\|\Sigma_1\|_{\mathrm{op}}, \kappa^2(\Sigma)\|\Sigma\|_{\mathrm{op}}, \|\Sigma_Z\|_{\mathrm{op}} \Big) \tag{26}$$

and

$$\mathfrak{m} = \sigma_{\min}(\Sigma_Z) \wedge \sigma_{\min}(\Sigma_1) \wedge \sigma_{\min}(\Sigma_2). \tag{27}$$

Then, if we solve (11)–(12) with λ_1^t and λ_2^t given by (13) and (14), the type II error of the test ψ_α defined by (16) is bounded by β.

Proof. We consider the case of $\mathcal{T} = \mathcal{D}_T$. Let $\tau \in \mathcal{D}_T$ be the true change-point. The type II error of the test ψ_α is defined as

$$\begin{aligned}\beta(\psi_\alpha, \mathcal{R}_{\rho,\mathcal{D}_T}) &= \sup_{\tau \in \mathcal{D}_T} \sup_{(\Theta_1,\Theta_2) \in V_p(\tau, \mathcal{R}_{p,\mathcal{D}_T})} \mathrm{P}^{\tau}_{\Theta_1,\Theta_2}\Big\{\psi_\alpha = 0\Big\} \\ &= \sup_{\tau \in \mathcal{D}_T} \sup_{(\Theta_1,\Theta_2) \in V_p(\tau, \mathcal{R}_{p,\mathcal{D}_T})} \mathrm{P}^{\tau}_{\Theta_1,\Theta_2}\Big\{\max_{t \in \mathcal{T}} \frac{\mathcal{F}(t)}{H_{\alpha,t}} < 1\Big\} \\ &\leq \inf_{t \in \mathcal{T}} \sup_{\tau \in \mathcal{D}_T} \sup_{(\Theta_1,\Theta_2) \in V_p(\tau, \mathcal{R}_{p,\mathcal{D}_T})} \mathrm{P}^{\tau}_{\Theta_1,\Theta_2}\Big\{\mathcal{F}(t) < H_{\alpha,t}\Big\} \\ &\leq \sup_{\tau \in \mathcal{D}_T} \sup_{(\Theta_1,\Theta_2) \in V_p(\tau, \mathcal{R}_{p,\mathcal{D}_T})} \mathrm{P}^{\tau}_{\Theta_1,\Theta_2}\Big\{\mathcal{F}(\tau) < H_{\alpha,\tau}\Big\}.\end{aligned}$$

We have to show that if the minimal jump energy satisfies condition (24), then the type II error of the test is at most β, $\mathrm{P}^{\tau}_{\Theta_1,\Theta_2}\Big\{\mathcal{F}(\tau) < H_{\alpha,\tau}\Big\} \leq \beta$. As in the case of the type I error, we have to consider four cases of location of the change-point τ with respect to $T/2$ and the boundaries.

Consider the case (a) of $\tau \geq T/2$ and $T - \tau \leq p$. Let us introduce the events

$$\mathcal{A} = \left\{\|\mathfrak{Z}_{\geq t}\|_{\mathrm{op}} \leq c_1^* \sqrt{\|\Sigma_Z\|_{\mathrm{op}}}\left(\sqrt{T-\tau} + \sqrt{p} + \sqrt{\log \frac{8}{\beta}}\right)\right\}, \quad \mathcal{B}_1 = \left\{\|\widehat{\Theta}_1^t - \Theta_1\|_2^2 \leq c^* \frac{\kappa^2(\Sigma_1)}{(1-\gamma)^2} \frac{Rp}{t}\right\},$$

$$\mathcal{C} = \left\{\frac{1}{T-\tau}\|(\Theta_1 - \Theta_2)\mathfrak{X}_{\geq \tau}\|_2^2 \geq \frac{\sigma_{\min}(\Sigma_Z) \vee \sigma_{\min}(\Sigma_2)}{2(\gamma^2+1)}\|\Theta_1 - \Theta_2\|_2^2\right\}.$$

Lemma 9 implies that $\mathrm{P}(\mathcal{A}) \geq 1 - \beta/4$. Proposition 5 implies $\mathrm{P}(\mathcal{B}_1) \geq 1 - c_2 \exp(-c_3 p) \geq 1 - \beta/4$ for sufficiently large p. Moreover, Lemma 11 implies that $\mathrm{P}(\mathcal{C}) \geq 1 - \beta/4$ under the condition

$$\|\Theta_2 - \Theta_1\|_2 > \frac{4\gamma(1+\gamma^2)\sigma_{\max}(\Sigma_2)}{\sigma_{\min}(\Sigma_Z) \wedge \sigma_{\min}(\Sigma_2)} \sqrt{\frac{1}{c}\frac{\log(8/\beta)}{T-\tau}}. \tag{28}$$

Note that (28) always holds under condition (24). We can write

$$\mathrm{P}^{\tau}_{\Theta_1,\Theta_2}\Big\{\mathcal{F}(\tau) < H_{\alpha,\tau}\Big\} \leq \mathrm{P}\Big\{\{\mathcal{F}(\tau) < H_{\alpha,\tau}\} \cap \mathcal{A} \cap \mathcal{B}_1 \cap \mathcal{C}\Big\} + \mathrm{P}(\mathcal{A}^c) + \mathrm{P}(\mathcal{B}_1^c) + \mathrm{P}(\mathcal{C}^c).$$

To bound the type II error by β, we need to show that the first probability is bounded by $\beta/4$.

By definition of $\widehat{\Theta}_2^\tau$, the test statistic can be bounded by

$$\mathcal{F}(\tau) = \varphi_2^\tau(\mathfrak{X}, \widehat{\Theta}_1^\tau) - \varphi_2^\tau(\mathfrak{X}, \widehat{\Theta}_2^\tau) \geq \varphi_2^\tau(\mathfrak{X}, \widehat{\Theta}_1^\tau) - \varphi_2^\tau(\mathfrak{X}, \Theta_2)$$

Taking into account that $\varphi_2^\tau(\mathfrak{X}, \Theta_2) = \frac{1}{T-\tau}\sum_{i=\tau}^{T-1} \|Z_i\|_2^2 + \lambda_2^\tau \|\Theta_2 \mathfrak{X}_{\geq \tau}\|_*$ and

$$\begin{aligned}\varphi_2^\tau(\mathfrak{X}, \widehat{\Theta}_1^\tau) &= \frac{1}{T-\tau}\sum_{i=\tau}^{T-1} \|\left(\Theta_2 - \widehat{\Theta}_1^\tau\right) X_i\|_2^2 + \frac{1}{T-\tau}\sum_{i=\tau}^{T-1}\|Z_i\|_2^2 \\ &\quad + \frac{2}{T-t}\left\langle \left(\Theta_2 - \widehat{\Theta}_1^\tau\right)\mathfrak{X}_{\geq\tau}, \mathfrak{Z}_{-\tau}\right\rangle + \lambda_2^\tau \|\widehat{\Theta}_1^\tau \mathfrak{X}_{\geq \tau}\|_*\end{aligned}$$

we obtain

$$\mathcal{F}(\tau) \geq \frac{1}{T-\tau}\sum_{i=\tau}^{T-1} \| \left(\Theta_2 - \widehat{\Theta}_1^{\tau}\right) X_i\|_2^2 + \frac{2}{T-t}\left\langle \left(\Theta_2 - \widehat{\Theta}_1^{\tau}\right) \mathfrak{X}_{\geq \tau}, \mathfrak{Z}_{-\tau}\right\rangle$$
$$+ \lambda_2^{\tau}\left(\|\widehat{\Theta}_1^{\tau}\mathfrak{X}_{\geq\tau}\|_* - \|\Theta_2\mathfrak{X}_{\geq\tau}\|_*\right) := A + B + C.$$

Note that for any i we have $\frac{1}{2}\| \left(\Theta_2 - \Theta_1\right) X_i\|_2^2 \leq \| \left(\Theta_2 - \widehat{\Theta}_1^{\tau}\right) X_i\|_2^2 + \| \left(\Theta_1 - \widehat{\Theta}_1^{\tau}\right) X_i\|_2^2$. Thus

$$A \geq \frac{1}{2(T-\tau)}\sum_{i=\tau}^{T-1} \| \left(\Theta_2 - \Theta_1\right) X_i\|_2^2 - \frac{1}{T-\tau}\sum_{i=\tau}^{T-1} \| \left(\Theta_1 - \widehat{\Theta}_1^{\tau}\right) X_i\|_2^2.$$

Next, by the Hölder inequality,

$$B \geq -\frac{2}{T-\tau}\| \left(\widehat{\Theta}_1^{\tau} - \Theta_1\right) \mathfrak{X}_{\geq\tau}\|_* \|\mathfrak{Z}_{-\tau}\|_{\mathrm{op}} - \frac{2}{T-\tau}\| \left(\Theta_2 - \Theta_1\right) \mathfrak{X}_{\geq\tau}\|_* \|\mathfrak{Z}_{-\tau}\|_{\mathrm{op}}.$$

Since $\lambda_2^{\tau} = \dfrac{6c_1^*\sqrt{\|\Sigma_Z\|_{\mathrm{op}}}\sqrt{p}}{T-\tau}$, using Lemma 9 we can show using $p \geq \log(8/\beta)$ and $T-\tau \leq p$ that $\lambda_2^{\tau} \geq \dfrac{2\|\mathfrak{Z}_{-\tau}\|_{\mathrm{op}}}{T-\tau}$ holds under the event $\mathcal{A}$. Consequently, given $\mathcal{A}$,

$$\begin{aligned}
B &\geq -\lambda_2^{\tau}\| \left(\widehat{\Theta}_1^{\tau} - \Theta_1\right) \mathfrak{X}_{\geq\tau}\|_* - \lambda_2^{\tau}\| \left(\Theta_2 - \Theta_1\right) \mathfrak{X}_{\geq\tau}\|_* \\
&= -\lambda_2^{\tau}\left(\|\mathbf{Pr}^{\perp}_{\Theta_1\mathfrak{X}_{\geq\tau}}\left[\left(\Theta_1 - \widehat{\Theta}_1\right)\mathfrak{X}_{\geq\tau}\right]\|_* + \|\mathbf{Pr}_{\Theta_1\mathfrak{X}_{\geq\tau}}\left[\left(\Theta_1 - \widehat{\Theta}_1\right)\mathfrak{X}_{\geq\tau}\right]\|_* + \| \left(\Theta_2 - \Theta_1\right) \mathfrak{X}_{\geq\tau}\|_*\right)
\end{aligned}$$

Similarly to Lemma 2 we can bound the term C using the triangle inequality and (41):

$$C \geq \lambda_2^{\tau}\left(\|\mathbf{Pr}^{\perp}_{\Theta_1\mathfrak{X}_{\geq\tau}}\left[\left(\Theta_1 - \widehat{\Theta}_1^{\tau}\right)\mathfrak{X}_{\geq\tau}\right]\|_* - \| \left(\Theta_2 - \Theta_1\right) \mathfrak{X}_{\geq\tau}\|_* - \|\mathbf{Pr}_{\Theta_1\mathfrak{X}_{\geq\tau}}\left[\left(\Theta_1 - \widehat{\Theta}_1\right)\mathfrak{X}_{\geq\tau}\right]\|_*\right).$$

Gathering all these bounds, we obtain that given $\mathcal{A}\cap\mathcal{B}_1\cap\mathcal{C}$,

$$\mathcal{F}(t) \geq \frac{1}{2(T-\tau)}\sum_{i=\tau}^{T-1} \| \left(\Theta_2 - \Theta_1\right) X_i\|_2^2 - \frac{1}{T-\tau}\sum_{i=\tau}^{T-1} \| \left(\Theta_1 - \widehat{\Theta}_1^{\tau}\right) X_i\|_2^2$$
$$- 2\lambda_2^{\tau}\| \left(\Theta_2 - \Theta_1\right) \mathfrak{X}_{\geq\tau}\|_* - 2\lambda_2^{\tau}\|\mathbf{Pr}_{\Theta_1\mathfrak{X}_{\geq\tau}}\left(\Theta_1 - \widehat{\Theta}_1^{\tau}\right)\mathfrak{X}_{\geq\tau}\|_*.$$

Using the trivial inequality $a^2 + b^2 \geq 2ab$ and the fact that $\mathrm{rank}(\Theta_2 - \Theta_1) \leq 2R$, we can show that

$$\begin{aligned}
2\lambda_2^{\tau}\| \left(\Theta_2 - \Theta_1\right) \mathfrak{X}_{\geq\tau}\|_* &\leq 2\lambda_2^{\tau}\sqrt{2\mathrm{rank}((\Theta_2 - \Theta_1)\mathfrak{X}_{\geq\tau})}\| \left(\Theta_2 - \Theta_1\right) \mathfrak{X}_{\geq\tau}\|_2 \\
&\leq 4\lambda_2^{\tau}\sqrt{R}\| \left(\Theta_2 - \Theta_1\right) \mathfrak{X}_{\geq\tau}\|_2 \\
&\leq 8\left(\lambda_2^{\tau}\right)^2 R(T-\tau) + \frac{\| \left(\Theta_2 - \Theta_1\right) \mathfrak{X}_{\geq\tau}\|_2^2}{4(T-\tau)}
\end{aligned}$$

and

$$
\begin{aligned}
2\lambda_2^\tau \|\mathbf{Pr}_{\Theta_1 \mathfrak{X}_{\geq\tau}} \left(\Theta_1 - \widehat{\Theta}_1^\tau\right) \mathfrak{X}_{\geq\tau}\|_* &\leq 2\lambda_2^\tau \sqrt{2\mathrm{rank}(\Theta_1 \mathfrak{X}_{\geq\tau})} \| \left(\Theta_1 - \widehat{\Theta}_1^\tau\right) \mathfrak{X}_{\geq\tau}\|_2 \\
&\leq 2\lambda_2^\tau \sqrt{2R} \| \left(\Theta_1 - \widehat{\Theta}_1^\tau\right) \mathfrak{X}_{\geq\tau}\|_2 \\
&\leq 2\left(\lambda_2^\tau\right)^2 R(T-\tau) + \frac{1}{T-\tau} \| \left(\Theta_1 - \widehat{\Theta}_1^\tau\right) \mathfrak{X}_{\geq\tau}\|_2^2
\end{aligned}
$$

Consequently, gathering the bounds, we get that given $\mathcal{A} \cap \mathcal{B}_1 \cap \mathcal{C}$,

$$
\begin{aligned}
\mathcal{F}(\tau) &\geq -\frac{2}{T-\tau} \| \left(\Theta_1 - \widehat{\Theta}_1^\tau\right) \mathfrak{X}_{\geq\tau}\|_2^2 + \frac{1}{4(T-\tau)} \| \left(\Theta_1 - \Theta_2\right) \mathfrak{X}_{\geq\tau}\|_2^2 - 10\left(\lambda_2^\tau\right)^2 R(T-\tau) \\
&\geq -\frac{2}{T-\tau} \|\mathfrak{X}_{\geq\tau}\|_{\mathrm{op}}^2 \|\Theta - \widehat{\Theta}_1^\tau\|_2^2 + \frac{1}{4(T-\tau)} \| \left(\Theta_1 - \Theta_2\right) \mathfrak{X}_{\geq\tau}\|_2^2 - 10\left(\lambda_2^\tau\right)^2 R(T-\tau)
\end{aligned}
$$

By Proposition 5 and Lemma 11, we have, given $\mathcal{B}_1$ and $\mathcal{C}$, $\|\widehat{\Theta}_1^\tau - \Theta_1\|_2^2 \leq c^* \dfrac{\kappa^2(\Sigma_1)}{(1-\gamma)^2} \dfrac{Rp}{\tau}$ and

$$
\frac{1}{4(T-\tau)} \| \left(\Theta_1 - \Theta_2\right) \mathfrak{X}_{\geq\tau}\|_2^2 \geq \frac{\sigma_{\min}(\Sigma_Z) \wedge \sigma_{\min}(\Sigma_2)}{8(\gamma^2+1)} \|\Theta_1 - \Theta_2\|_2^2
$$

which implies, given the event $\mathcal{A} \cap \mathcal{B}_1 \cap \mathcal{C}$,

$$
\mathcal{F}(\tau) \geq -\frac{2c^* Rp}{\tau(T-\tau)} \frac{\kappa^2(\Sigma_1)}{(1-\gamma)^2} \|\mathfrak{X}_{\geq\tau}\|_{\mathrm{op}}^2 + \frac{\sigma_{\min}(\Sigma_Z) \wedge \sigma_{\min}(\Sigma_2)}{8(\gamma^2+1)} \|\Theta_1 - \Theta_2\|_2^2 - 360 \frac{Rp\|\Sigma_Z\|_{\mathrm{op}}}{T-\tau}.
$$

Thus, to show that $\mathrm{P}\left\{\mathcal{F}(\tau) \leq H_{\alpha,\tau}\right\} \leq \beta$ we need to prove that under assumption (24)

$$
\mathrm{P}\left\{-\frac{2c^* Rp}{\tau(T-\tau)} \frac{\kappa^2(\Sigma_1)}{(1-\gamma)^2} \|\mathfrak{X}_{\geq\tau}\|_{\mathrm{op}}^2 + \frac{\sigma_{\min}(\Sigma_Z) \wedge \sigma_{\min}(\Sigma_2)}{2(\gamma^2+1)} \|\Theta_1 - \Theta_2\|_2^2 - 360 \frac{Rp\|\Sigma_Z\|_{\mathrm{op}}}{T-\tau} \leq H_{\alpha,\tau}\right\} \leq \beta/4. \tag{29}
$$

Using Lemma 8 we can show that (29) holds under the condition

$$
\begin{aligned}
\frac{\sigma_{\min}(\Sigma_Z) \wedge \sigma_{\min}(\Sigma_2)}{2(\gamma^2+1)} \|\Theta_1 - \Theta_2\|_2^2 \geq H_{\alpha,\tau} &+ 360 \frac{Rp\|\Sigma_Z\|_{\mathrm{op}}}{T-\tau} \\
&+ \frac{4c^* Rp}{\tau} \frac{\kappa^2(\Sigma_1)}{(1-\gamma)^2} \|\Sigma_2\|_{\mathrm{op}} \left(1 + \frac{c_0}{1-\gamma} \frac{p + \log(8/\beta)}{T-\tau}\right).
\end{aligned} \tag{30}
$$

Using the definition of $H_{\alpha,\tau}$ and the inequality

$$
360 \frac{Rp\|\Sigma_Z\|_{\mathrm{op}}}{T-\tau} + \frac{2c^* Rp}{\tau} \frac{\kappa^2(\Sigma_1)\|\Sigma_2\|_{\mathrm{op}}}{(1-\gamma)^2} \leq \max(360, 4c^*) \max\left(\|\Sigma_Z\|_{\mathrm{op}}, \frac{\kappa^2(\Sigma_1)\|\Sigma_2\|_{\mathrm{op}}}{(1-\gamma)^2}\right) \frac{Rp}{Tq^2(\tau/T)}
$$

we can bound the right-hand side of inequality (30) from above by

$$
C \frac{Rp}{Tq^2(\tau/T)} \max\left(\|\Sigma_Z\|_{\mathrm{op}}, \frac{\kappa^2(\Sigma)\|\Sigma\|_{\mathrm{op}}}{(1-\gamma)^2}, \frac{\kappa^2(\Sigma_1)\|\Sigma_2\|_{\mathrm{op}}}{(1-\gamma)^2}\right) \left(1 + \frac{c_0}{1-\gamma} \frac{2p + \log(8|\mathcal{T}|/\alpha) + \log(8/\beta)}{T}\right),
$$

where C is a constant depending only on c^*, c_1^* and c_2^*. Applying the condition $2p < T$ we obtain that (30) is guaranteed if assumption (24) on the norm of the matrix jump is satisfied. This implies (29).

Now we have to bound the type II error in case of $\dfrac{T}{2} \leq \tau < T - p$. Set

$$\mathcal{A}' = \left\{ \frac{1}{T-t} \|\mathfrak{Z}_{\geq t} \mathfrak{X}_{\geq \tau}^{\mathrm{T}}\|_{\mathrm{op}} \leq \frac{c_2^* \|\Sigma\|_{\mathrm{op}}}{1-\gamma} \sqrt{\frac{p}{T-\tau}} \right\}.$$

Lemma 10 implies that $\mathrm{P}(\mathcal{A}') \geq 1 - k_1 \exp(-k_2 p) \geq 1 - \beta/4$ if $p \geq k_2^{-1} \log(4k_1/\beta)$. We have to show that $\mathrm{P}\{\{\mathcal{F}(\tau) < H_{\alpha,\tau}\} \cap \mathcal{A}' \cap \mathcal{B}_1 \cap \mathcal{C}\} \leq \beta/4$.

Using the same reasoning as above we obtain

$$\begin{aligned}
\mathcal{F}(\tau) &= \varphi_2^\tau(\mathfrak{X}, \widehat{\Theta}_1^\tau) - \varphi_2^\tau(\mathfrak{X}, \widehat{\Theta}_2^\tau) \geq \varphi_2^\tau(\mathfrak{X}, \widehat{\Theta}_1^\tau) - \varphi_2^\tau(\mathfrak{X}, \Theta_2^\tau) \\
&= \frac{1}{T-\tau} \sum_{i=\tau}^{T-1} \| \left(\Theta_2 - \widehat{\Theta}_1^\tau\right) X_i \|_2^2 - \frac{2}{T-\tau} \left\langle \widehat{\Theta}_1^\tau - \Theta_2, \mathfrak{Z}_{-\tau} \mathfrak{X}_{\geq \tau}^{\mathrm{T}} \right\rangle + \lambda_2^\tau \left(\|\widehat{\Theta}_1(\tau)\|_* - \|\Theta_2\|_* \right) \\
&\geq \frac{1}{2(T-\tau)} \sum_{i=\tau}^{T-1} \| \left(\Theta_2 - \Theta_1\right) X_i \|_2^2 - \frac{1}{T-\tau} \sum_{i=\tau}^{T-1} \| \left(\Theta_1 - \widehat{\Theta}_1^\tau\right) X_i \|_2^2 + \\
&\quad - \frac{2}{T-\tau} \|\widehat{\Theta}_1^\tau - \Theta_1\|_* \|\mathfrak{Z}_{-\tau} \mathfrak{X}_{\geq \tau}^{\mathrm{T}}\|_{\mathrm{op}} - \frac{2}{T-\tau} \|\Theta_2 - \Theta_1\|_* \|\mathfrak{Z}_{-\tau} \mathfrak{X}_{\geq \tau}^{\mathrm{T}}\|_{\mathrm{op}} \\
&\quad - \lambda_2^\tau \left(\|\Theta_2 - \Theta_1\|_* + \|\mathbf{Pr}_{\Theta_1} \left[\Theta_1 - \widehat{\Theta}_1\right] \|_* - \|\mathbf{Pr}_{\Theta_1}^\perp \left[\Theta_1 - \widehat{\Theta}_1\right] \|_* \right).
\end{aligned}$$

Remind that $\lambda_2^\tau = 2c_2^* \dfrac{\|\Sigma_2\|_{\mathrm{op}}}{1-\gamma} \sqrt{\dfrac{p}{T-\tau}}$. Using Lemma 10, we obtain that given $\mathcal{A}'$,

$$\frac{2}{T-\tau} \|\mathfrak{Z}_{-\tau} \mathfrak{X}_{\geq t}^{\mathrm{T}}\|_{\mathrm{op}} \leq \frac{2c_2^* \|\Sigma_2\|_{\mathrm{op}}}{1-\gamma} \sqrt{\frac{p}{T-\tau}} = \lambda_2^\tau.$$

By Lemma 11 provided the condition (28), we have that given the event $\mathcal{C}$

$$\frac{1}{2(T-\tau)} \sum_{i=\tau}^{T-1} \| \left(\Theta_2 - \Theta_1\right) X_i \|_2^2 \geq \frac{\sigma_{\min}(\Sigma_Z) \wedge \sigma_{\min}(\Sigma_2)}{4(1+\gamma^2)} \|\Theta_2 - \Theta_1\|_2^2.$$

Consequently, given the event $\mathcal{A}' \cap \mathcal{B}_1 \cap \mathcal{C}$,

$$\begin{aligned}
\mathcal{F}(\tau) &\geq \frac{1}{2(T-\tau)} \sum_{i=\tau}^{T-1} \| \left(\Theta_2 - \Theta_1\right) X_i \|_2^2 - \frac{1}{T-\tau} \sigma_{\max}(\mathfrak{X}_{\geq t}^{\mathrm{T}} \mathfrak{X}_{\geq t}^{\mathrm{T}}) \|\Theta_1 - \widehat{\Theta}_1^\tau\|_2^2 \\
&\quad - 2\lambda_2^\tau \left(\|\Theta_2 - \Theta_1\|_* + \|\mathbf{Pr}_{\Theta_1} \left(\Theta_1 - \widehat{\Theta}_1\right) \|_* \right) \\
&\geq \frac{\sigma_{\min}(\Sigma_Z) \wedge \sigma_{\min}(\Sigma_2)}{4(1+\gamma^2)} \|\Theta_2 - \Theta_1\|_2^2 - \frac{1}{T-\tau} \|\mathfrak{X}_{\geq \tau}\|_{\mathrm{op}}^2 c^* \frac{Rp}{\tau} \frac{\kappa^2(\Sigma_1)}{(1-\gamma)^2} \\
&\quad - 2\lambda_2^\tau \left(\|\Theta_2 - \Theta_1\|_* + \|\mathbf{Pr}_{\Theta_1} \left(\Theta_1 - \widehat{\Theta}_1^\tau\right) \|_* \right).
\end{aligned}$$

Next, as in the proof of the previous case we can obtain the following bounds given $\mathcal{B}_1$,

$$\begin{aligned}2\lambda_2^\tau\|\Theta_2-\Theta_1\|_* &\le 2\lambda_2^\tau\sqrt{2R}\|\Theta_2-\Theta_1\|_2\\ &\le \frac{16R(1+\gamma^2)(\lambda_2^\tau)^2}{\sigma_{\min}(\Sigma_Z)\wedge\sigma_{\min}(\Sigma_2)}+\frac{\sigma_{\min}(\Sigma_Z)\wedge\sigma_{\min}(\Sigma_2)}{8(1+\gamma^2)}\|\Theta_2-\Theta_1\|_2^2\\ &= \frac{64(c_2^*)^2(1+\gamma^2)\|\Sigma_2\|_{\mathrm{op}}^2}{(1-\gamma)^2(\sigma_{\min}(\Sigma_Z)\wedge\sigma_{\min}(\Sigma_2))}\frac{Rp}{T-\tau}+\frac{\sigma_{\min}(\Sigma_Z)\wedge\sigma_{\min}(\Sigma_2)}{8(1+\gamma^2)}\|\Theta_2-\Theta_1\|_2^2\end{aligned}$$

and

$$\begin{aligned}2\lambda_2^\tau\|\mathbf{Pr}_{\Theta_1}\left(\Theta_1-\widehat{\Theta}_1^\tau\right)\|_* &\le 2\lambda_2^\tau\sqrt{2R}\|\Theta_1-\widehat{\Theta}_1^\tau\|_2\\ &\le 2c_2^*\sqrt{2c^*}\frac{\|\Sigma_2\|_{\mathrm{op}}\kappa(\Sigma_1)}{(1-\gamma)^2}\frac{Rp}{Tq(\tau/T)}\le\frac{\sqrt{2}}{2}c_2^*\sqrt{c^*}\frac{\|\Sigma_2\|_{\mathrm{op}}\kappa(\Sigma_1)}{(1-\gamma)^2}\frac{Rp}{Tq^2(\tau/T)}.\end{aligned}$$

Consequently, gathering all the above inequalities, we obtain

$$\begin{aligned}\Big\{\{\mathcal{F}(\tau)<H_{\alpha,\tau}\}\cap\mathcal{A}'\cap\mathcal{B}_1\cap\mathcal{C}\Big\}\subseteq\Big\{H_{\alpha,\tau}>\mathcal{F}(\tau)\ge\frac{\sigma_{\min}(\Sigma_Z)\wedge\sigma_{\min}(\Sigma_2)}{8(1+\gamma^2)}\|\Theta_2-\Theta_1\|_2^2\\ -\frac{2c^*Rp}{\tau(T-\tau)}\frac{\kappa^2(\Sigma_1)}{(1-\gamma)^2}\|\mathfrak{X}_{\ge\tau}\|_{\mathrm{op}}^2-\frac{c_2^*\sqrt{2c^*}}{2}\frac{\|\Sigma_2\|_{\mathrm{op}}\kappa(\Sigma_1)}{(1-\gamma)^2}\frac{Rp}{Tq^2(\tau/T)}\\ -\frac{64(c_2^*)^2(1+\gamma^2)\|\Sigma_2\|_{\mathrm{op}}^2}{(1-\gamma)^2(\sigma_{\min}(\Sigma_Z)\wedge\sigma_{\min}(\Sigma_2))}\frac{Rp}{T-\tau}\Big\}.\end{aligned}$$

Thus, to show that $\mathrm{P}_{\Theta_1,\Theta_2}^\tau\Big\{\mathcal{F}(\tau)\le H_{\alpha,\tau}\Big\}\le\beta$ we need to prove that under assumption (24)

$$\begin{aligned}\mathrm{P}\Big\{\frac{2c^*Rp}{\tau(T-\tau)}\frac{\kappa^2(\Sigma_1)}{(1-\gamma)^2}\|\mathfrak{X}_{\ge\tau}\|_{\mathrm{op}}^2\ge\frac{\sigma_{\min}(\Sigma_Z)\wedge\sigma_{\min}(\Sigma_2)}{2(\gamma^2+1)}\|\Theta_1-\Theta_2\|_2^2\\ -\frac{c_2^*\sqrt{2c^*}}{2}\frac{\|\Sigma_2\|_{\mathrm{op}}\kappa(\Sigma_1)}{(1-\gamma)^2}\frac{Rp}{Tq^2(\tau/T)}-\frac{64c_1^2(1+\gamma^2)\|\Sigma_2\|_{\mathrm{op}}^2}{(1-\gamma)^2(\sigma_{\min}(\Sigma_Z)\wedge\sigma_{\min}(\Sigma_2))}\frac{Rp}{T-\tau}-H_{\alpha,\tau}\Big\}\le\beta/4.\end{aligned}\tag{31}$$

By Lemma 8 and by the definition of $H_{\alpha,\tau}$, this bound holds if

$$\begin{aligned}\frac{\sigma_{\min}(\Sigma_Z)\wedge\sigma_{\min}(\Sigma_2)}{2(\gamma^2+1)}\|\Theta_1-\Theta_2\|_2^2\ge\frac{c_2^*\sqrt{2c^*}}{2}\frac{\|\Sigma_2\|_{\mathrm{op}}\kappa(\Sigma_1)}{(1-\gamma)^2}\frac{Rp}{Tq^2(\tau/T)}+\frac{64c_1^2(1+\gamma^2)\|\Sigma_2\|_{\mathrm{op}}^2}{(1-\gamma)^2(\sigma_{\min}(\Sigma_Z)\wedge\sigma_{\min}(\Sigma_2))}\frac{Rp}{T-\tau}\\ +C\|\Sigma\|_{\mathrm{op}}\frac{\kappa^2(\Sigma)}{(1-\gamma)^2}\frac{Rp}{Tq^2(\tau/T)}\left(1+\frac{2}{1-\gamma}\left(1+\frac{\log(8|\mathcal{T}|/\alpha)}{T}\right)\right)\\ +\frac{2c^*\kappa^2(\Sigma_1)}{(1-\gamma)^2}\frac{Rp}{\tau}\|\Sigma_2\|_{\mathrm{op}}\left(1+\frac{2}{1-\gamma}\left(1+\frac{\log(8/\beta)}{T-\tau}\right)\right).\end{aligned}$$

In what follows C denotes an absolute constant. Using the identity $\tau^{-1}(T-\tau)^{-1}=1/(T^2q^2(\tau/T))$ and the bound

$$\frac{64c_1^2(1+\gamma^2)\|\Sigma_2\|_{\mathrm{op}}^2}{(1-\gamma)^2(\sigma_{\min}(\Sigma_Z)\wedge\sigma_{\min}(\Sigma_2))}\frac{Rp}{T-\tau}+\frac{1+\gamma}{1-\gamma}\frac{2c^*\kappa^2(\Sigma_1)\|\Sigma_2\|_{\mathrm{op}}}{(1-\gamma)^2}\frac{Rp}{\tau}\le C\frac{\mathfrak{M}_1\vee\mathfrak{M}_2}{(1-\gamma)^3}\frac{Rp}{Tq^2(\tau/T)}.$$

We can show that (31) holds if

$$\frac{\sigma_{\min}(\Sigma_Z)\wedge\sigma_{\min}(\Sigma_2)}{2(\gamma^2+1)}\|\Theta_1-\Theta_2\|_2^2\ge C\frac{\mathfrak{M}_1\vee\mathfrak{M}_2}{(1-\gamma)^3}\frac{Rp}{Tq^2(\tau/T)}\left(1+\frac{\log(8|\mathcal{T}|/\alpha)+\log(8/\beta)}{T}\right).$$

which holds if assumption (24) on the norm of the matrix jump is satisfied. $\square$

C Proof of the lower bound

Our goal is to show that there exists a sequence $\rho_{p,T}$ such that if $\mathcal{R}_{p,T} = o(\rho_{p,T})$ as $p, T \to \infty$, then

$$\liminf_{p,T\to\infty} \inf_{\psi\in\Psi_\alpha} \beta(\psi, \mathcal{R}_{p,T}) \geq 1 - \alpha.$$

Proof of Theorem 2.

Step 1. Reduction to Bayesian testing.

Let $\mathfrak{X} = (X_0, \dots, X_T)$, where the observations X_t are defined on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ and follow the $\mathrm{VAR}_p(\Theta_1, \Theta_2, \Sigma_Z)$ model (3). Let $\pi^T_{p,0}$ and $\pi^T_{p,1}$ be two families of priors on the parameter $(\Theta^t)_{0\leq t<T}$ under H_0 and H_1, respectively such that $\pi^T_{p,i}(\mathcal{M}_p(R,\gamma)) \to 1$, $i = 0, 1$, $p, T \to \infty$ and $\pi^T_{p,1}(\Theta_1, \Theta_2) \in \mathcal{V}_{p,\tau}(\rho)) \to 1$, $p, T \to \infty$, where $\rho = \mathcal{R}_{p,T} > 0$ is the minimal detectable energy we are interested in.

Define the mixtures $\mathrm{P}_{\pi^T_{p,i}} = \mathrm{E}_{\pi^T_{p,i}} \mathrm{P}(A)$, $A \in \mathcal{A}$. The standard technique to prove the lower bound (see, for example, (Ingster and Suslina 2003)) consists in showing that the testing risk is bounded from below as follows,

$$\begin{aligned}
\inf_{\psi_{p,T}\in\Psi_\alpha} \beta(\psi_{p,T}, \mathcal{R}_{p,T}) &\geq \beta(\psi^*_{p,T}, \mathrm{P}_{\pi^T_{p,0}}, \mathrm{P}_{\pi^T_{p,1}}) + o(1) \\
&\geq 1 - \frac{1}{2}\|\mathrm{P}_{\pi^T_{p,1}} - \mathrm{P}_{\pi^T_{p,0}}\|_{\mathrm{TV}} - \alpha + o(1) \\
&\geq 1 - \frac{1}{2}\left(\mathrm{E}_{\pi^T_{p,0}}\left[\frac{d\mathrm{P}_{\pi^T_{p,1}}}{d\mathrm{P}_{\pi^T_{p,0}}}(\mathfrak{X})\right]^2 - 1\right)^{1/2} - \alpha + o(1). \qquad (32)
\end{aligned}$$

Here $\psi^*_{p,T}$ is the optimal Bayesian test for the hypotheses $\mathrm{H}_0 : \ \mathrm{P} = \mathrm{P}_{\pi^T_{p,0}}$ against $\mathrm{H}_1 : \ \mathrm{P} = \mathrm{P}_{\pi^T_{p,1}}$ and $\beta(\psi^*_{p,T}, \mathrm{P}_{\pi^T_{p,0}}, \mathrm{P}_{\pi^T_{p,1}})$ its type II error. If, for the appropriately chosen priors $\pi^T_{p,1}$ and $\pi^T_{p,0}$, we can show that

$$\mathrm{E}_{\pi^T_{p,0}}\left[\frac{d\mathrm{P}_{\pi^T_{p,1}}}{d\mathrm{P}_{\pi^T_{p,0}}}(\mathfrak{X})\right]^2 \leq 1 + o(1), \quad p, T \to \infty$$

then (32) will imply $\inf_{\psi_{p,T}\in\Psi_\alpha} \beta(\psi_{p,T}, \mathcal{R}_{p,T}) \geq 1 - \alpha + o(1)$ and the lower bound will follow.

Step 2. Computation of the second moment of likelihood ratio.

Suppose that under H_0 the transition matrix parameter is zero, $\Theta^t = 0$ $\forall 0 \leq t < T$ and under the alternative $\Theta^t = \Theta_1 = -(1 - \tau/T)\Delta\Theta$ for $0 \leq t \leq \tau$ and $\Theta^t = \Theta_2 = (\tau/T)\Delta\Theta$ for $\tau + 1 \leq t < T$. Assume that the change matrix $\Delta\Theta = \Theta_2 - \Theta_1$ is distributed according to some prior distribution $\mu = \mu_{p,T}$ defined on $\mathcal{M}_p(R,\gamma)$ that will be chosen later. Define the corresponding likelihood ratio of mixtures under H_1 and H_0,

$$L = \frac{d\mathrm{P}_{\pi^T_{p,1}}}{d\mathrm{P}_{\pi^T_{p,0}}}(\mathfrak{X}) := \mathrm{E}_{\Delta\Theta\sim\mu}\Big(\frac{d\mathrm{P}_{\Delta\Theta}}{d\mathrm{P}_0}(\mathfrak{X})\Big),$$

where P_0 stands for the measure of T i.i.d. $\mathcal{N}_p(0,\Sigma_Z)$ Gaussian vectors and P_Δ stands for the measure of T consecutive VAR_p observations with the change-point at τ and the transition matrices Θ_1 and Θ_2 defined above. We have

$$\begin{aligned}\frac{d\mathrm{P}_{\Delta\Theta}}{d\mathrm{P}_0}(\mathfrak{X}) = \exp\Bigg\{-\frac{1}{2}\Bigg(\Big(1-\frac{\tau}{T}\Big)^2\sum_{t=0}^{\tau-1} X_t^{\mathrm{T}}\Delta\Theta^{\mathrm{T}}\Sigma_Z^{-1}\Delta\Theta X_t + \Big(\frac{\tau}{T}\Big)^2\sum_{t=\tau}^{T-1} X_t^{\mathrm{T}}\Delta\Theta^{\mathrm{T}}\Sigma_Z^{-1}\Delta\Theta X_t \\ + \Big(1-\frac{\tau}{T}\Big)\sum_{t=0}^{\tau-1}(X_t^{\mathrm{T}}\Delta\Theta^{\mathrm{T}}\Sigma_Z^{-1}X_{t+1} + X_{t+1}^{\mathrm{T}}\Sigma_Z^{-1}\Delta\Theta X_t) \\ -\Big(\frac{\tau}{T}\Big)\sum_{t=\tau}^{T-1}(X_t^{\mathrm{T}}\Delta\Theta^{\mathrm{T}}\Sigma_Z^{-1}X_{t+1} + X_{t+1}^{\mathrm{T}}\Sigma_Z^{-1}\Delta\Theta X_t)\Bigg) - \frac{1}{2}X_0^{\mathrm{T}}(\Sigma_1^{-1}-\Sigma_Z^{-1})X_0\Bigg\}.\end{aligned}$$

Here Σ_1 is the covariance matrix of the $\mathrm{VAR}_p(\Theta_1)$ process with $\Theta_1 = -(1-\frac{\tau}{T})\Delta\Theta$. Recall that Θ_1 and Σ_1 satisfy the Lyapunov equation $\Sigma_1 = \Theta_1\Sigma_1\Theta_1^{\mathrm{T}} + \Sigma_Z$ and, consequently, $\Sigma_1 = (1-\frac{\tau}{T})^2\Delta\Theta\Sigma_1\Delta\Theta^{\mathrm{T}} + \Sigma_Z$.

Let $\Delta\tilde{\Theta}$ be an independent copy of $\Delta\Theta$ following the law μ. Then the second moment of the likelihood ratio can be written as

$$\begin{aligned}\mathrm{E}_0[L^2] &= \mathrm{E}_0\Big[\mathrm{E}_{\Delta\Theta\sim\mu}\Big(\frac{d\mathrm{P}_{\Delta\Theta}}{d\mathrm{P}_0}(\mathfrak{X})\Big)^2\Big] \\ &= \mathrm{E}_0\Big[\mathrm{E}_{\mu^2}\Big(\frac{d\mathrm{P}_{\Delta\Theta}}{d\mathrm{P}_0}(\mathfrak{X})\frac{d\mathrm{P}_{\Delta\tilde{\Theta}}}{d\mathrm{P}_0}(\mathfrak{X})\Big)\Big] = \mathrm{E}_{\mu^2}\mathrm{E}_0\Big(\frac{d\mathrm{P}_{\Delta\Theta}}{d\mathrm{P}_0}(\mathfrak{X})\frac{d\mathrm{P}_{\Delta\tilde{\Theta}}}{d\mathrm{P}_0}(\mathfrak{X})\Big)\end{aligned}$$

Thus, to bound from above $\mathrm{E}_0[L^2]$, we will have to bound $\mathrm{E}_0\Big(\frac{d\mathrm{P}_{\Delta\Theta}}{d\mathrm{P}_0}\frac{d\mathrm{P}_{\Delta\tilde{\Theta}}}{d\mathrm{P}_0}\Big)$. We have

$$\begin{aligned}\frac{dP_{\Delta\Theta}}{dP_0}\frac{dP_{\Delta\tilde{\Theta}}}{dP_0}(\mathfrak{X}) = \exp\Bigg\{-\frac{1}{2}\Bigg(\Big(1-\frac{\tau}{T}\Big)^2\sum_{t=0}^{\tau-1} X_t^{\mathrm{T}}(\Delta\Theta^{\mathrm{T}}\Sigma_Z^{-1}\Delta\Theta + \Delta\tilde{\Theta}^{\mathrm{T}}\Sigma_Z^{-1}\Delta\tilde{\Theta})X_t \\ + \Big(\frac{\tau}{T}\Big)^2\sum_{t=\tau}^{T-1} X_t^{\mathrm{T}}(\Delta\Theta^{\mathrm{T}}\Sigma_Z^{-1}\Delta\Theta + \Delta\tilde{\Theta}^{\mathrm{T}}\Sigma_Z^{-1}\Delta\tilde{\Theta})X_t \\ + \Big(1-\frac{\tau}{T}\Big)\sum_{t=0}^{\tau-1}(X_t^{\mathrm{T}}(\Delta\Theta+\Delta\tilde{\Theta})^{\mathrm{T}}\Sigma_Z^{-1}X_{t+1} + X_{t+1}^{\mathrm{T}}\Sigma_Z^{-1}(\Delta\Theta+\Delta\tilde{\Theta})X_t) \\ -\Big(\frac{\tau}{T}\Big)\sum_{t=\tau}^{T-1}(X_t^{\mathrm{T}}(\Delta\Theta+\Delta\tilde{\Theta})^{\mathrm{T}}\Sigma_Z^{-1}X_{t+1} + X_{t+1}^{\mathrm{T}}\Sigma_Z^{-1}(\Delta\Theta+\Delta\tilde{\Theta})X_t)\Bigg) \\ -\frac{1}{2}X_0^{\mathrm{T}}(\Sigma_1^{-1}+\tilde{\Sigma}_1^{-1}-2\Sigma_Z^{-1})X_0\Bigg\}.\end{aligned}$$

Since under H_0 the observations are i.i.d. $X_t\sim\mathcal{N}_p(0,\Sigma_Z)$, we can set $X_t = \Sigma_Z^{1/2}U_t$, where $U_t\sim\mathcal{N}_p(0,I_p)$, $t=0,\dots,T$ are i.i.d. Denote for simplicity $\Omega = \Sigma_Z^{-1/2}\Delta\Theta\Sigma_Z^{1/2}$,

$\tilde{\Omega} = \Sigma_Z^{-1/2}\Delta\tilde{\Theta}\Sigma_Z^{1/2}$. Then

$$\begin{aligned}
\mathrm{E}_0\Big(\frac{dP_{\Delta\Theta}}{dP_0}\frac{dP_{\Delta\tilde{\Theta}}}{dP_0}(\mathfrak{X})\Big) = \mathrm{E}\exp\Bigg\{-\frac{1}{2}\Bigg(&\Big(1-\frac{\tau}{T}\Big)^2\sum_{t=0}^{\tau-1}U_t^{\mathrm{T}}(\Omega^{\mathrm{T}}\Omega+\tilde{\Omega}^{\mathrm{T}}\tilde{\Omega})U_t \\
&+\Big(\frac{\tau}{T}\Big)^2\sum_{t=\tau}^{T-1}U_t^{\mathrm{T}}(\Omega^{\mathrm{T}}\Omega+\tilde{\Omega}^{\mathrm{T}}\tilde{\Omega})U_t \\
&+\Big(1-\frac{\tau}{T}\Big)\sum_{t=0}^{\tau-1}(U_t^{\mathrm{T}}(\Omega+\tilde{\Omega})^{\mathrm{T}}U_{t+1}+U_{t+1}^{\mathrm{T}}(\Omega+\tilde{\Omega})U_t) \\
&-\Big(\frac{\tau}{T}\Big)\sum_{t=\tau}^{T-1}(U_t^{\mathrm{T}}(\Omega+\tilde{\Omega})^{\mathrm{T}}U_{t+1}+U_{t+1}^{\mathrm{T}}(\Omega+\tilde{\Omega})U_t)\Bigg) \\
&-\frac{1}{2}U_0^{\mathrm{T}}(\Sigma_Z^{1/2}(\Sigma_1^{-1}+\tilde{\Sigma}_1^{-1})\Sigma_Z^{1/2}-2I_p)U_0\Bigg\}.
\end{aligned}$$

Denote by $\mathbf{u} = \mathrm{vec}(U_0,\ldots,U_T) \in \mathbb{R}^{p(T+1)}$ the vector obtained by concatenation of $T+1$ vectors U_t of dimension p. Then

$$\mathrm{E}_0\Big(\frac{dP_{\Delta\Theta}}{dP_0}\frac{dP_{\Delta\tilde{\Theta}}}{dP_0}(\mathfrak{X})\Big) = \mathrm{E}\exp\Big\{-\frac{1}{2}\mathbf{u}^{\mathrm{T}}M\mathbf{u}\Big\} = \Big(\det(I_{T+1}\otimes I_p+M)\Big)^{-1/2},$$

where $\mathbf{u}\sim\mathcal{N}_{p(T+1)}(0, I_{T+1}\otimes I_p)$ and $M = M(\Omega,\tilde{\Omega})$ is the following $p(T+1)\times p(T+1)$ block-tridiagonal symmetric matrix of blocks of size p:

$$M = \begin{pmatrix}
D_0 & L^{\mathrm{T}} & O_{p\times p} & \ldots & \ldots & \ldots & O_{p\times p} & O_{p\times p} \\
L & D_1 & L^{\mathrm{T}} & O_{p\times p} & \ddots & \ddots & \ddots & O_{p\times p} \\
O_{p\times p} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
O_{p\times p} & \ddots & L & D_{\tau-1} & L^{\mathrm{T}} & O_{p\times p} & \ddots & \vdots \\
\vdots & \ddots & \ddots & K & D_\tau & K^{\mathrm{T}} & \ddots & O_{p\times p} \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & O_{p\times p} \\
O_{p\times p} & \ddots & \ddots & \ddots & O_{p\times p} & K & D_{T-1} & K^{\mathrm{T}} \\
O_{p\times p} & O_{p\times p} & \ldots & \ldots & \ldots & O_{p\times p} & K & D_T
\end{pmatrix}$$

with $L = \Big(1-\frac{\tau}{T}\Big)(\Omega+\tilde{\Omega})$, $K = -\Big(\frac{\tau}{T}\Big)(\Omega+\tilde{\Omega})$, $D_0 = \Big(1-\frac{\tau}{T}\Big)^2 W + \Sigma_Z^{1/2}(\Sigma_1^{-1} + \tilde{\Sigma}_1^{-1})\Sigma_Z^{1/2} - 2I_p$, $D_T = O_{p\times p}$ and

$$D_t = \Big(1-\frac{\tau}{T}\Big)^2 W\mathbf{1}_{\{1\le t<\tau\}} + \Big(\frac{\tau}{T}\Big)^2 W\mathbf{1}_{\{\tau\le t<T\}},$$

where $W = \Omega^{\mathrm{T}}\Omega + \tilde{\Omega}^{\mathrm{T}}\tilde{\Omega}$.

Step 3. Bounds on the determinants.

To calculate $\det(I_{T+1}\otimes I_p+M)$, we will use the determinant formula for block matrices: suppose that $F = \begin{pmatrix} A & B \\ C & D\end{pmatrix}$ and D is invertible, then $\det(F) = \det(D)\det(A -$

$BD^{-1}C$). Applying T times the block matrix determinant formula starting from the lower-right block matrix $I_p + D_T$, we obtain

$$\det(I_{T+1} \otimes I_p + M) = \prod_{k=0}^{T} \det(I_p + A_k),$$

where $A_T = O_{p\times p}$ and A_k are defined recursively as follows,

$$\begin{aligned}
A_{T-k} &= \Big(\frac{\tau}{T}\Big)^2 W - K^{\mathrm{T}}(I_p + A_{T-k+1})^{-1}K, \quad k = 1, \dots, T-\tau-1\\
A_\tau &= \Big(\frac{\tau}{T}\Big)^2 W - L^{\mathrm{T}}(I_p + A_{\tau+1})^{-1}K\\
A_{\tau-k} &= \Big(1-\frac{\tau}{T}\Big)^2 W - L^{\mathrm{T}}(I_p + A_{\tau-k+1})^{-1} : L \quad k = 1, \dots, \tau-1\\
A_0 &= \Sigma_Z^{1/2}(\Sigma_1^{-1} + \tilde{\Sigma}_1^{-1})\Sigma_Z^{1/2} - 2I_p + \Big(1-\frac{\tau}{T}\Big)^2 W - L^{\mathrm{T}}(I_p + A_1)^{-1}L.
\end{aligned} \tag{33}$$

To bound the likelihood ratio from above, we have to calculate the lower bounds on the determinants $\det(I_p + A_k)$ that are provided by Lemma 4. Let Ω and $\tilde{\Omega}$ be chosen in a way that $\|\Omega\|_2 = \|\tilde{\Omega}\|_2 = \delta$ for some $\delta > 0$ (see Step 4). Using the identity

$$(T-\tau)\Big(\frac{\tau}{T}\Big)^2 + \tau\Big(1-\frac{\tau}{T}\Big)^2 = Tq^2\Big(\frac{\tau}{T}\Big).$$

we obtain from Lemma 4 that for sufficiently small $\delta \in (0,1)$ satisfying the conditions of the lemma,

$$\begin{aligned}
\prod_{k=0}^{T} \det(I + A_k)^{-1/2} \le \exp\Big\{&Tq^2\Big(\frac{\tau}{T}\Big)\mathrm{tr}(\Omega\tilde{\Omega}^T) + 3\kappa^2(\Sigma_Z)\Big(1-\frac{\tau}{T}\Big)^2\delta^2\\
&+16Tq^2\Big(\frac{\tau}{T}\Big)\delta^4 + 12\Big(1-\frac{\tau}{T}\Big)^2(1+\kappa^2(\Sigma_Z))^2\delta^4\Big\}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathrm{E}_0[L^2] \le \exp&\Big(C_1\Big(1-\frac{\tau}{T}\Big)^2\delta^2 + C_2\Big(1-\frac{\tau}{T}\Big)^4\delta^4 + 16Tq^2\Big(\frac{\tau}{T}\Big)\delta^4\Big)\\
&\times \mathrm{E}_{(\Omega,\tilde{\Omega})}\exp\Big\{Tq^2\Big(\frac{\tau}{T}\Big)\mathrm{tr}(\Omega\tilde{\Omega}^{\mathrm{T}})\Big\},
\end{aligned} \tag{34}$$

where the constants C_1 and C_2 depend only Σ_Z.

Step 4. Prior on the jump matrix.

We define a low-rank prior μ on the matrix Ω as in (Carpentier and Nickl 2015). Assume that p is a multiple of the rank R. Let v_k, $k = 1, \dots, R$ be random vectors of size p consisting of p independent Rademacher random entries taking values ± 1 with probability $1/2$. Let the matrix H be composed of R blocks H_k, $k = 1, \dots, R$ of size $p \times (p/R)$ defined as $H_k = v_k \otimes \xi_k^{\mathrm{T}}$, where $\xi_k = (\xi_{1,k}, \dots, \xi_{p/R,k})^{\mathrm{T}}$ is a random vector of size p/R with the Rademacher independent entries $\xi_{1,k}$:

$$H = \begin{pmatrix} \underbrace{v_1 \otimes \xi_1^{\mathrm{T}}}_{p\times p/R} & \underbrace{v_2 \otimes \xi_2^{\mathrm{T}}}_{p\times p/R} & \cdots & \underbrace{v_R \otimes \xi_R^{\mathrm{T}}}_{p\times p/R} \end{pmatrix}.$$

Set $\Omega = \dfrac{\delta_{p,T}}{p} H$ so that $\|\Omega\|_{\mathrm{op}} \leq \|\Omega\|_2 = \delta_{p,T} = \delta$. Note that

$$\|\Omega\|_2 = \|\Sigma_Z^{-1/2} \Delta\Theta \Sigma_Z^{1/2}\|_2 \leq \kappa^{1/2}(\Sigma_Z)\|\Delta\Theta\|_2$$

and

$$\|\Delta\Theta\|_2 = \|\Sigma_Z^{1/2} \Omega \Sigma_Z^{-1/2}\|_2 \leq \kappa^{1/2}(\Sigma_Z)\|\Omega\|_2.$$

The corresponding matrix $\Delta\Theta$ belongs to $\mathcal{M}_p(R, \gamma)$ if $\delta < \gamma\kappa^{-1/2}(\Sigma_Z)$ since in this case $\|\Delta\Theta\|_{\mathrm{op}} \leq \|\Delta\Theta\|_2 \leq \kappa^{1/2}(\Sigma_Z)\delta < \gamma$.

Step 5. Bound on the second moment of likelihood ratio.

Using exactly the same reasoning as in the proof of Theorem 1 in (Carpentier and Nickl 2015) (see p. 2686) we can show that

$$\mathrm{E}_{(\Omega,\tilde{\Omega})} \exp\left\{Tq^2(\tau/T)\mathrm{tr}(\Omega\tilde{\Omega}^{\mathrm{T}})\right\} \leq 1 + \frac{2}{\dfrac{p^2}{2T^2q^4(\tau/T)\delta^4} - 1}.$$

Thus (34) implies that

$$\mathrm{E}_0[L^2] \leq e^{C_1\delta^2 + C_2\delta^4 + 16Tq^2\left(\frac{\tau}{T}\right)\delta^4} \left[1 + 2\Big(\frac{p^2}{2T^2q^4(\tau/T)\delta^4} - 1\Big)^{-1}\right].$$

In order to analyze the decay of each term in this upper bound we consider two asymptotic regimes, $p \lesssim \sqrt{T}$ and $p \gtrsim \sqrt{T}$.

1. In the case of $p \lesssim \sqrt{T}$ and $q^2(\tau/T)\delta^2 = o(p/T)$ we have $\dfrac{p^2}{T^2q^4(\tau/T)\delta^4} \to \infty$, as $p, T \to \infty$. Moreover, as $\min(\tau/T, 1-\tau/T) \geq h_* > 0$, we have that

$$Tq^2\Big(\frac{\tau}{T}\Big)\delta^4 = \frac{T}{q^2\Big(\dfrac{\tau}{T}\Big)}\left(q\Big(\frac{\tau}{T}\Big)\delta\right)^4 = \frac{o(p^2/T)}{q^2\Big(\dfrac{\tau}{T}\Big)} = o(1) \quad \text{as } p, T \to \infty$$

   Since $\min(\tau/T, 1-\tau/T) \geq h_* > 0$, and $p \lesssim T$, $\delta^2 = o(p/T) = o_T(T^{-1/2})$ as $p, T \to \infty$.

2. In the case of $p \gtrsim \sqrt{T}$ and $q^2(\tau/T)\delta^2 = o(T^{-1/2})$ we have $\dfrac{p^2}{T^2q^4(\tau/T)\delta^4} \to \infty$, and

$$Tq^2\Big(\frac{\tau}{T}\Big)\delta^4 = \frac{T}{q^2\Big(\frac{\tau}{T}\Big)}\left(q\Big(\frac{\tau}{T}\Big)\delta\right)^4 = \frac{o(T^{-1})}{q^2\Big(\frac{\tau}{T}\Big)} = o_T(1) \quad \text{as } p, T \to \infty$$

   Similary, using $\min(\tau/T, 1-\tau/T) \geq h_* > 0$, we obtain $\delta^2 = o_T(T^{-1/2})$.

In both cases we have $\mathrm{E}_0[L^2] \leq 1 + o(1)$ as $p, T \to \infty$ and

$$\kappa(\Sigma_Z)\mathcal{E}(\tau, \Theta_1, \Theta_2) = \kappa(\Sigma_Z)q^2\Big(\frac{\tau}{T}\Big)\|\Delta\Theta\|_2^2 \geq q^2\Big(\frac{\tau}{T}\Big)\delta^2 = o\Big(\frac{p}{T} \wedge T^{-1/2}\Big).$$

Thus the rate satisfies

$$\mathcal{R}_{p,T} = \kappa^{-1/2}(\Sigma_Z) o\Big(\sqrt{\frac{p}{T}} \wedge T^{-1/4}\Big)$$

and the theorem follows. $\square$

The following lemma provides the upper bound on the determinant of the matrix $I+M$.

Lemma 4. *Let Ω, $\tilde{\Omega} \in \mathbb{R}^{p\times p}$ be the matrices of rank R with the same Frobenius norm, $\|\Omega\|_2 = \|\tilde{\Omega}\|_2 = \delta$, $0 < \delta < 1$ and with the operator norm bounded by γ. Let $L = \Big(1-\frac{\tau}{T}\Big)(\Omega+\tilde{\Omega})$, $K = -\Big(\frac{\tau}{T}\Big)(\Omega+\tilde{\Omega})$, $W = \Omega^{\mathrm{T}}\Omega+\tilde{\Omega}^{\mathrm{T}}\tilde{\Omega}$ and the matrices A_k, $k = 0, \dots, T-1$ be defined in (33). Then*

1. *For any $0 < \delta < 1/4$ and any $1 \le k \le T-1$, we have $\|A_k\|_{\mathrm{op}} \le 4\delta^2$, and*

$$(\det(I_p + A_k))^{-1/2} \le \begin{cases} \exp\Big\{\Big(1-\frac{\tau}{T}\Big)^2 \mathrm{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) + 16\Big(1-\frac{\tau}{T}\Big)^2\delta^4\Big\}, & k = 1, \dots, \tau-1 \\ \exp\Big\{\Big(\frac{\tau}{T}\Big)^2 \mathrm{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) + 16\Big(\frac{\tau}{T}\Big)^2\delta^4\Big\}, & k = \tau, \dots, T-1. \end{cases} \tag{35}$$

2. *For any $0 < \delta < 1/4$ and $\delta^2 < \frac{1}{2}\left(4+3\kappa^2(\Sigma_Z)\right)^{-1}$ we have*

$$\begin{aligned}(\det(I_p + A_0))^{-1/2} \le \exp\Big\{&\Big(1-\frac{\tau}{T}\Big)^2 \mathrm{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) \\ &+3\Big(1-\frac{\tau}{T}\Big)^2\kappa^2(\Sigma_Z)\delta^2 + 12\Big(1-\frac{\tau}{T}\Big)^2(1+\kappa^2(\Sigma_Z))^2\delta^4\Big\}.\end{aligned} \tag{36}$$

Proof. To bound from below the determinants $\det(I_p + A_k)$, we use the result of Lemma 12 from (Rump 2018):

$$\det(I_p + A_k) \ge \exp\Big\{\mathrm{tr}(A_k) - \frac{\frac{1}{2}\|A_k\|_2^2}{1-\rho(A_k)}\Big\} \ge \exp\Big\{\mathrm{tr}(A_k) - \frac{\frac{1}{2}\|A_k\|_2^2}{1-\|A_k\|_{\mathrm{op}}}\Big\},$$

under the condition that $\|A_k\|_{\mathrm{op}} < 1$. Let us start with A_{T-1}. We have

$$A_{T-1} = \Big(\frac{\tau}{T}\Big)^2(\Omega^{\mathrm{T}}\Omega + \tilde{\Omega}^{\mathrm{T}}\tilde{\Omega}) - \Big(\frac{\tau}{T}\Big)^2(\Omega+\tilde{\Omega})^{\mathrm{T}}(\Omega+\tilde{\Omega}) = -\Big(\frac{\tau}{T}\Big)^2(\Omega^{\mathrm{T}}\tilde{\Omega} + \tilde{\Omega}^{\mathrm{T}}\Omega)$$

We have $\mathrm{tr}(\Omega^{\mathrm{T}}\tilde{\Omega} + \tilde{\Omega}^{\mathrm{T}}\Omega) = 2\mathrm{tr}(\Omega^{\mathrm{T}}\tilde{\Omega})$ and

$$\|\Omega^{\mathrm{T}}\tilde{\Omega} + \tilde{\Omega}^{\mathrm{T}}\Omega\|_{\mathrm{op}} \le \|\Omega^{\mathrm{T}}\tilde{\Omega} + \tilde{\Omega}^{\mathrm{T}}\Omega\|_2 \le 2\|\Omega\|_2\|\tilde{\Omega}\|_2 = 2\delta^2. \tag{37}$$

Consequently, for any $0 < \delta < 1/\sqrt{2}$,

$$\begin{aligned}\det(I_p + A_{T-1}) &= \det\Big(I_p - \Big(\frac{\tau}{T}\Big)^2(\Omega^{\mathrm{T}}\tilde{\Omega} + \tilde{\Omega}^{\mathrm{T}}\Omega)\Big) \\ &\ge \exp\Big\{-\Big(\frac{\tau}{T}\Big)^2 \mathrm{tr}(\Omega^{\mathrm{T}}\tilde{\Omega} + \tilde{\Omega}^{\mathrm{T}}\Omega) - \frac{\frac{1}{2}\Big(\frac{\tau}{T}\Big)^4\|(\Omega^{\mathrm{T}}\tilde{\Omega} + \tilde{\Omega}^{\mathrm{T}}\Omega\|_2^2}{1-\Big(\frac{\tau}{T}\Big)^2\|\Omega^{\mathrm{T}}\tilde{\Omega} + \tilde{\Omega}^{\mathrm{T}}\Omega\|_{\mathrm{op}}}\Big\} \\ &\ge \exp\Big\{-2\Big(\frac{\tau}{T}\Big)^2 \mathrm{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) - \frac{2\Big(\frac{\tau}{T}\Big)^4\delta^4}{1-2\delta^2}\Big\},\end{aligned}$$

where we used the fact that $\tau/T \le 1$.

Let us turn to $\det(I_p + A_{T-k})$, where $k = 1, \ldots, T-\tau-1$. We will show by induction that $\|A_{T-k}\|_{\text{op}} \le 4\delta^2$ if $0 < \delta < 1/4$. Note that $\|A_{T-1}\|_{\text{op}} \le 2\delta^2$. We have

$$\begin{aligned} A_{T-k} &= \Big(\frac{\tau}{T}\Big)^2(\Omega^{\mathrm{T}}\Omega + \tilde{\Omega}^{\mathrm{T}}\tilde{\Omega}) - \Big(\frac{\tau}{T}\Big)^2(\Omega+\tilde{\Omega})^{\mathrm{T}}(I_p + A_{T-k+1})^{-1}(\Omega+\tilde{\Omega}) \\ &= -\Big(\frac{\tau}{T}\Big)^2(\Omega^{\mathrm{T}}\tilde{\Omega} + \tilde{\Omega}^{\mathrm{T}}\Omega) - \Big(\frac{\tau}{T}\Big)^2(\Omega+\tilde{\Omega})^{\mathrm{T}}\left(\sum_{j=1}^{\infty}(-1)^j A_{T-k+1}^j\right)(\Omega+\tilde{\Omega}). \end{aligned}$$

Suppose that $\|A_{T-k+1}\|_{\text{op}} < 4\delta^2 < 1/4$, where $0 < \delta < 1/4$. Then using (37), we get

$$\begin{aligned} \|A_{T-k}\|_2 &\le 2\Big(\frac{\tau}{T}\Big)^2\delta^2 + \Big(\frac{\tau}{T}\Big)^2\|\Omega+\tilde{\Omega}\|_2^2\sum_{j=1}^{\infty}\|A_{T-k+1}\|_{\text{op}}^j \\ &\le \Big(\frac{\tau}{T}\Big)^2\left(2\delta^2 + 4\delta^2\frac{\|A_{T-k+1}\|_{\text{op}}}{1-\|A_{T-k+1}\|_{\text{op}}}\right) \le 2\delta^2\Big(\frac{\tau}{T}\Big)^2\frac{1+\|A_{T-k+1}\|_{\text{op}}}{1-\|A_{T-k+1}\|_{\text{op}}} \le 4\Big(\frac{\tau}{T}\Big)^2\delta^2. \end{aligned}$$

Consequently, $\|A_{T-k}\|_{\text{op}} \le \|A_{T-k}\|_2 \le 4\Big(\dfrac{\tau}{T}\Big)^2\delta^2$.

To bound the trace of A_{T-k} we use the fact that for any two square matrices A and B, $\operatorname{tr}(AB) \le \|A\|_2\|B\|_2$. Then, for $0 < \delta < 1/4$, since $\|A_{T-k+1}\|_{\text{op}} \le 4\delta^2$,

$$\begin{aligned} \operatorname{tr}(A_{T-k}) &= -2\Big(\frac{\tau}{T}\Big)^2\operatorname{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) - \operatorname{tr}\left(\Big(\frac{\tau}{T}\Big)^2(\Omega+\tilde{\Omega})^{\mathrm{T}}\Big(\sum_{j=1}^{\infty}(-1)^j A_{T-k+1}^j\Big)(\Omega+\tilde{\Omega})\right) \\ &\ge -2\Big(\frac{\tau}{T}\Big)^2\operatorname{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) - \Big(\frac{\tau}{T}\Big)^2\|\Omega+\tilde{\Omega}\|_2^2\Big\|\sum_{j=1}^{\infty}(-1)^j A_{T-k+1}^j\Big\|_{\text{op}} \\ &\ge -2\Big(\frac{\tau}{T}\Big)^2\operatorname{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) - \Big(\frac{\tau}{T}\Big)^2\|\Omega+\tilde{\Omega}\|_2^2\frac{\|A_{T-k+1}\|_{\text{op}}}{1-\|A_{T-k+1}\|_{\text{op}}} \\ &\ge -2\Big(\frac{\tau}{T}\Big)^2\operatorname{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) - 4\Big(\frac{\tau}{T}\Big)^2\delta^2\frac{4\delta^2}{1-4\delta^2} \\ &\ge -2\Big(\frac{\tau}{T}\Big)^2\operatorname{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) - \frac{64}{3}\Big(\frac{\tau}{T}\Big)^2\delta^4. \end{aligned}$$

Thus, using Lemma 12 we obtain for $0 < \delta < 1/4$ that

$$\begin{aligned} \det(I + A_{T-k}) &\ge \exp\left\{\operatorname{tr}(A_{T-k}) - \frac{\frac{1}{2}\|A_{T-k}\|_2^2}{1-\|A_{T-k}\|_{\text{op}}}\right\} \\ &\ge \exp\left\{-2\Big(\frac{\tau}{T}\Big)^2\operatorname{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) - \frac{64}{3}\Big(\frac{\tau}{T}\Big)^2\delta^4 - \frac{8\Big(\frac{\tau}{T}\Big)^2\delta^4}{1-4\delta^2}\right\} \\ &\ge \exp\left\{-2\Big(\frac{\tau}{T}\Big)^2\operatorname{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) - 32\Big(\frac{\tau}{T}\Big)^2\delta^4\right\}. \end{aligned}$$

Following the same approach, we obtain the same bound for $\det(I_p + A_\tau)$. For $\det(I_p + A_{\tau-k})$, $k = 1, \ldots, \tau-1$ we have

$$\det(I + A_{T-k}) \ge \exp\left\{-2\Big(1-\frac{\tau}{T}\Big)^2\operatorname{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) - 32\Big(1-\frac{\tau}{T}\Big)^2\delta^4\right\}.$$

Let us bound $\det(I + A_0)$. We have

$$\begin{aligned}\|A_0\|_2 &\le \|\Big(1-\frac{\tau}{T}\Big)^2(\Omega^{\mathrm{T}}\Omega + \tilde{\Omega}^{\mathrm{T}}\tilde{\Omega}) - L^{\mathrm{T}}(I_p + A_1)^{-1}L\|_2 \\ &\quad + \|\Sigma_Z^{1/2}(\Sigma_1^{-1} - \Sigma_Z^{-1})\Sigma_Z^{1/2}\|_2 + \|\Sigma_Z^{1/2}(\tilde{\Sigma}_1^{-1} - \Sigma_Z^{-1})\Sigma_Z^{1/2}\|_2 \\ &\le 4\Big(1-\frac{\tau}{T}\Big)^2\delta^2 + \|\Sigma_Z^{1/2}(\Sigma_1^{-1} - \Sigma_Z^{-1})\Sigma_Z^{1/2}\|_2 + \|\Sigma_Z^{1/2}(\tilde{\Sigma}_1^{-1} - \Sigma_Z^{-1})\Sigma_Z^{1/2}\|_2\end{aligned}$$

From the Lyapunov equation $\Sigma_1 = \Big(1-\frac{\tau}{T}\Big)^2\Delta\Theta\Sigma_1\Delta\Theta^{\mathrm{T}} + \Sigma_Z$ it follows that

$$\Sigma_Z^{1/2}(\Sigma_1^{-1} - \Sigma_Z^{-1})\Sigma_Z^{1/2} = \Sigma_Z^{-1/2}(\Sigma_Z - \Sigma_1)\Sigma_1^{-1}\Sigma_Z^{1/2} = -\Big(1-\frac{\tau}{T}\Big)^2\Sigma_Z^{1/2}\Delta\Theta\Sigma_1\Delta\Theta^{\mathrm{T}}\Sigma_1^{-1}\Sigma_Z^{1/2}$$

and

$$\begin{aligned}\|\Sigma_Z^{1/2}(\Sigma_1^{-1} - \Sigma_Z^{-1})\Sigma_Z^{1/2}\|_2 &= \Big(1-\frac{\tau}{T}\Big)^2\|\Sigma_Z^{-1/2}\Delta\Theta\Sigma_1\Delta\Theta^{\mathrm{T}}\Sigma_1^{-1}\Sigma_Z^{1/2}\|_2 \\ &= \Big(1-\frac{\tau}{T}\Big)^2\|\Omega\Sigma_Z^{-1/2}\Sigma_1\Sigma_Z^{-1/2}\Omega^{\mathrm{T}}\Sigma_Z^{1/2}\Sigma_1^{-1}\Sigma_Z^{1/2}\|_2 \\ &\le \Big(1-\frac{\tau}{T}\Big)^2\|\Omega\|_2^2\kappa(\Sigma_1)\kappa(\Sigma_Z).\end{aligned}$$

Lemma 13 and $\|\Delta\Theta\|_{\mathrm{op}} \le \kappa^{1/2}(\Sigma_Z)\|\Omega\|_{\mathrm{op}}$ imply that for $\delta < \frac{1}{2}\kappa^{-1/2}(\Sigma_Z)$,

$$\kappa(\Sigma_1) \le \kappa(\Sigma_Z)\frac{1 + \Big(1-\frac{\tau}{T}\Big)^2\|\Delta\Theta\|_{\mathrm{op}}^2}{1 - \Big(1-\frac{\tau}{T}\Big)^2\|\Delta\Theta\|_{\mathrm{op}}^2} \le \kappa(\Sigma_Z)\frac{1 + \Big(1-\frac{\tau}{T}\Big)^2\kappa(\Sigma_Z)\delta^2}{1 - \Big(1-\frac{\tau}{T}\Big)^2\kappa(\Sigma_Z)\delta^2} \le 3\kappa(\Sigma_Z)$$

Therefore, for any $\delta < \frac{1}{2}\kappa^{-1/2}(\Sigma_Z)$,

$$\|\Sigma_Z^{1/2}(\Sigma^{-1} - \Sigma_Z^{-1})\Sigma_Z^{1/2}\|_2 \le 3\Big(1-\frac{\tau}{T}\Big)^2\kappa^2(\Sigma_Z)\delta^2.$$

Thus

$$\|A_0\|_{\mathrm{op}} \le \|A_0\|_2 \le \Big(1-\frac{\tau}{T}\Big)^2(4 + 3\kappa^2(\Sigma_Z))\delta^2. \tag{38}$$

Next,

$$\begin{aligned}\mathrm{tr}(A_0) &= \mathrm{tr}\Big(\Big(1-\frac{\tau}{T}\Big)^2(\Omega^{\mathrm{T}}\Omega + \tilde{\Omega}^{\mathrm{T}}\tilde{\Omega}) - L^{\mathrm{T}}(I_p + A_1)^{-1}L\Big) \\ &\qquad + \mathrm{tr}(\Sigma_Z^{1/2}(\Sigma_1^{-1} - \Sigma_Z^{-1})\Sigma_Z^{1/2}) + \mathrm{tr}(\Sigma_Z^{1/2}(\tilde{\Sigma}_1^{-1} - \Sigma_Z^{-1})\Sigma_Z^{1/2}) \\ &= \mathrm{tr}\Big(\Big(1-\frac{\tau}{T}\Big)^2(\Omega^{\mathrm{T}}\Omega + \tilde{\Omega}^{\mathrm{T}}\tilde{\Omega}) - L^{\mathrm{T}}(I_p + A_1)^{-1}L\Big) \\ &\qquad - \Big(1-\frac{\tau}{T}\Big)^2\mathrm{tr}(\Sigma_Z^{-1/2}\Delta\Theta\Sigma_1\Delta\Theta^{\mathrm{T}}\Sigma_1^{-1}\Sigma_Z^{1/2}) - \Big(1-\frac{\tau}{T}\Big)^2\mathrm{tr}(\Sigma_Z^{-1/2}\Delta\tilde{\Theta}\tilde{\Sigma}_1\Delta\tilde{\Theta}^{\mathrm{T}}\tilde{\Sigma}_1^{-1}\Sigma_Z^{1/2}).\end{aligned}$$

We have for $\delta < \frac{1}{2}\kappa^{-1/2}(\Sigma_Z)$,

$$\begin{aligned}\mathrm{tr}(\Sigma_Z^{-1/2}\Delta\Theta\Sigma_1\Delta\Theta^{\mathrm{T}}\Sigma_1^{-1}\Sigma_Z^{1/2}) &= \mathrm{tr}(\Delta\Theta\Sigma_1\Delta\Theta^{\mathrm{T}}\Sigma_1^{-1}) \le \|\Delta\Theta\Sigma_1\|_2\|\Delta\Theta^{\mathrm{T}}\Sigma_1^{-1}\|_2 \\ &\le \|\Delta\Theta\|_2^2\|\Sigma_1\|_{\mathrm{op}}\|\Sigma_1^{-1}\|_{\mathrm{op}} = \|\Sigma_Z^{1/2}\Omega\Sigma_Z^{-1/2}\|_2^2\kappa(\Sigma_1) \\ &\le \delta^2\kappa(\Sigma_Z)\kappa(\Sigma_1) \le 3\delta^2\kappa^2(\Sigma_Z).\end{aligned}$$

Thus, using the same reasoning as above to bound the first term of $\mathrm{tr}(A_0)$, we get

$$\mathrm{tr}(A_0) \geq -2\Big(1-\frac{\tau}{T}\Big)^2 \mathrm{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) - \frac{64}{3}\Big(1-\frac{\tau}{T}\Big)^2\delta^4 - 6\Big(1-\frac{\tau}{T}\Big)^2\delta^2\kappa^2(\Sigma_Z).$$

If $\delta^2 < \frac{1}{2}\Big(4+3\kappa^2(\Sigma_Z)\Big)^{-1}$, then using (38) we obtain the bound

$$\frac{\frac{1}{2}\|A_0\|_2^2}{1-\|A_0\|_{\mathrm{op}}} \leq \Big(1-\frac{\tau}{T}\Big)^2\delta^4\Big(4+3\kappa^2(\Sigma_Z)\Big)^2.$$

Taking into account that fact that $\frac{1}{4}\kappa^{-1}(\Sigma_Z) > \frac{1}{2}\Big(4+3\kappa^2(\Sigma_Z)\Big)^{-1}$ and that $64/3+\Big(4+3\kappa^2(\Sigma_Z) < 24(1+\kappa^2(\Sigma_Z))^2$, we obtain that for $\delta^2 < \frac{1}{2}\Big(4+3\kappa^2(\Sigma_Z)\Big)^{-1}$,

$$\begin{aligned}
\det(I+A_0) &\geq \exp\Big\{\mathrm{tr}(A_0) - \frac{\frac{1}{2}\|A_0\|_2^2}{1-\|A_0\|_{\mathrm{op}}}\Big\} \\
&\geq \exp\Big\{-2\Big(1-\frac{\tau}{T}\Big)^2\mathrm{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) - \frac{64}{3}\Big(1-\frac{\tau}{T}\Big)^2\delta^4 \\
&\qquad -6\Big(1-\frac{\tau}{T}\Big)^2\delta^2\kappa^2(\Sigma_Z) - \Big(1-\frac{\tau}{T}\Big)^2\delta^4\Big(4+3\kappa^2(\Sigma_Z)\Big)^2\Big\} \\
&\geq \exp\Big\{-2\Big(1-\frac{\tau}{T}\Big)^2\mathrm{tr}(\Omega^{\mathrm{T}}\tilde{\Omega}) - 6\Big(1-\frac{\tau}{T}\Big)^2\delta^2\kappa^2(\Sigma_Z) - 24\Big(1-\frac{\tau}{T}\Big)^2\delta^4(1+\kappa^2(\Sigma_Z))^2\Big\}.
\end{aligned}$$

and the lemma follows. □

D Results on the estimation of the transition matrix

The following result is a reformulation of Corollary 4 in (Negahban and Wainwright 2011).

Proposition 5. *Let $(X_0,\dots,X_n)$ be n consecutive realizations of a $\mathrm{VAR}_p(\Theta,\Sigma_Z)$ stationary process (1) with the transition matrix Θ of rank at most R and satisfying $\|\Theta\|_{\mathrm{op}} = \gamma < 1$. Assume that $n \geq p$.*

Let $\widehat{\Theta}$ be a solution of the SDP (10) with the regularization parameter λ_n given by (9). Then there exist a universal constants c^*, c_2, c_3 such that

$$\|\widehat{\Theta}-\Theta\|_2^2 \leq c^*\frac{\kappa^2(\Sigma)}{(1-\gamma)^2}\frac{Rp}{n}. \tag{39}$$

with probability greater than $1-c_2\exp(-c_3 p)$.

The next proposition shows the consistency of the transition matrix estimator in the operator norm.

Proposition 6 (Operator norm consistency)**.** *Let $(X_0,\dots,X_n)$ be consecutive realizations of a $\mathrm{VAR}_p(\Theta,\Sigma_Z)$ stationary process (1) with the transition matrix Θ of rank at most R satisfying $\|\Theta\|_{\mathrm{op}} = \gamma < 1$. Assume that $n \geq p$. Let $\widehat{\Theta}$ be a solution of SDP (10) with the regularization parameter defined in (9). Then with probability at least $1-2/p$*

$$\|\widehat{\Theta}-\Theta\|_{\mathrm{op}} \leq 8c_1\frac{\kappa(\Sigma)}{1-\gamma}\sqrt{\frac{p}{n}},$$

where c_1 is an absolute constant.

Proof. We have
$$\frac{1}{n}\sum_{i=0}^{n-1}\|X_{i+1}-MX_i\|_2^2=\frac{1}{n}\|\mathfrak{Y}_n-M\mathfrak{X}_n\|_2^2.$$
Its the derivative with respect to M is equal to $-\dfrac{2}{n}(\mathfrak{Y}-M\mathfrak{X}_n)\mathfrak{X}_n^{\mathrm{T}}$. On the other hand, for any subgradient matrix $W\in\partial\|\widehat{\Theta}\|_*$, we are guaranteed $\|W\|_{\mathrm{op}}\le 1$. By the Karush-Kuhn-Tucker condition, any solution $\widehat{\Theta}$ of (10) must satisfy
$$-\frac{2}{n}(\mathfrak{Y}_n-\widehat{\Theta}\mathfrak{X}_n)\mathfrak{X}_n^{\mathrm{T}}+\lambda_1 W=0.$$
Note that $\mathfrak{Y}_n=\Theta\mathfrak{X}_n+\mathfrak{Z}_n$, then we can rewrite the KKT condition as
$$\frac{2}{n}((\widehat{\Theta}-\Theta)\mathfrak{X}_n-\mathfrak{Z}_n)\mathfrak{X}_n^{\mathrm{T}}+\lambda_1 W=0$$
It implies that
$$\widehat{\Theta}-\Theta=\left(\frac{2}{n}\mathfrak{Z}_n\mathfrak{X}_n^{\mathrm{T}}-\lambda_1 W\right)\left(\frac{2}{n}\mathfrak{X}_n\mathfrak{X}_n^{\mathrm{T}}\right)^{-1}.$$
Thus, from Lemma 8 and Lemma 10 it follows that
$$\begin{aligned}\|\widehat{\Theta}-\Theta\|_{\mathrm{op}}&=\left\|\left(\frac{2}{n}\mathfrak{Z}_n\mathfrak{X}_n^{\mathrm{T}}-\lambda_1 W\right)\left(\frac{2}{n}\mathfrak{X}_n\mathfrak{X}_n^{\mathrm{T}}\right)^{-1}\right\|_{\mathrm{op}}\le\left(2\left\|\frac{1}{n}\mathfrak{Z}_n\mathfrak{X}_n^{\mathrm{T}}\right\|_{\mathrm{op}}+\lambda_1\right)\left\|\left(\frac{2}{n}\mathfrak{X}_n\mathfrak{X}_n^{\mathrm{T}}\right)^{-1}\right\|_{\mathrm{op}}\\&\le 2c_2^*\frac{\|\Sigma\|_{\mathrm{op}}}{1-\gamma}\sqrt{\frac{p}{n}}\sigma_{\min}^{-1}\left(\frac{1}{n}\mathfrak{X}_n\mathfrak{X}_n^{\mathrm{T}}\right)\le 8c_2^*\frac{\kappa(\Sigma)}{1-\gamma}\sqrt{\frac{p}{n}}\end{aligned}$$
with probability at least $1-2/p$. ◻

E Auxiliary results

Throughout this section $\mathfrak{X}_n=(X_1,\dots,X_n)$ denotes a $p\times n$ matrix of n consecutive realizations X_t of a $\mathrm{VAR}_p(\Theta,\Sigma_Z)$ process defined in (1). The corresponding noise matrix $\mathfrak{Z}_n=(Z_1,\dots,Z_n)$ is a $p\times n$ matrix with i.i.d. Gaussian centered columns Z_t with the covariance matrix Σ_Z.

Lemma 7. *Assume that we solve* (11) - (15) *with* $\lambda_1^t\ge 3\|\mathfrak{Z}_t\mathfrak{X}_t^{\mathrm{T}}\|_{\mathrm{op}}$ *and* $\lambda_2^t\ge 3\|\mathfrak{Z}_{\ge t}\mathfrak{X}_{\ge t}^{\mathrm{T}}\|_{\mathrm{op}}$. *Then, for* $i=1,2$, *we have*
$$\|\mathbf{Pr}_{\Theta_i}^{\perp}[\Theta-\widehat{\Theta}_i(t)]\|_*\le 5\|\mathbf{Pr}_{\Theta_i}[\Theta-\widehat{\Theta}_i(t)]\|_*.\tag{40}$$
Proof. We will prove (40) for $i=1$. The proof for $i=2$ is completely analogous. Using the triangle inequality, for any two matrices A and B we easily get
$$\|A\|_*-\|B\|_*\le\|\mathbf{Pr}_A[A-B]\|_*-\|\mathbf{Pr}_A^{\perp}[A-B]\|_*.\tag{41}$$
Indeed,
$$\begin{aligned}\|B\|_*=\|B+A-A\|_*&=\|A+\mathbf{Pr}_A[B-A]+\mathbf{Pr}_A^{\perp}[B-A]\|_*\\&\ge\|A+\mathbf{Pr}_A^{\perp}[B-A]\|_*-\|\mathbf{Pr}_A[B-A]\|_*\\&=\|A\|_*+\|\mathbf{Pr}_A^{\perp}[B-A]\|_*-\|\mathbf{Pr}_A[B-A]\|_*\end{aligned}$$

which implies (41).

Now, by convexity of $\|\mathfrak{Y}_t - \Theta\mathfrak{X}_t\|_2^2$ we have

$$\begin{aligned}
\|\mathfrak{Y}_t - \widehat{\Theta}_1^t\mathfrak{X}_t\|_2^2 - \|\mathfrak{Y}_t - \Theta\mathfrak{X}_t\|_2^2 &\geq -2\left\langle \mathfrak{Y}_t - \Theta\mathfrak{X}_t, \left(\widehat{\Theta}_1^t - \Theta\right)\mathfrak{X}_t \right\rangle \\
&= -2\left\langle \mathfrak{Z}_t\mathfrak{X}_t^{\mathrm{T}}, \widehat{\Theta}_1^t - \Theta \right\rangle \\
&\geq -2\|\widehat{\Theta}_1^t - \Theta\|_*\|\mathfrak{Z}_t\mathfrak{X}_t^{\mathrm{T}}\|_{\mathrm{op}} \geq -\frac{2}{3}\lambda_1^t\|\widehat{\Theta}_1^t - \Theta\|_*.
\end{aligned}$$

On the other hand, using the definition of $\widehat{\Theta}_1$ we compute

$$\lambda_1^t\|\widehat{\Theta}_1(t)\|_* - \lambda_1^t\|\Theta\|_* \leq \|\mathfrak{Y}_t - \Theta\mathfrak{X}_t\|_2^2 - \|\mathfrak{Y}_t - \widehat{\Theta}_1^t\mathfrak{X}_t\|_2^2 \leq \frac{2}{3}\lambda_1^t\|\widehat{\Theta}_1^t - \Theta\|_*,$$

which combined with (41) proves (40). □

The following result about the largest and smallest eigenvalue of empirical covariance matrix of the process (3) follows from Lemma 4 of (Negahban and Wainwright 2011).

Lemma 8. *Let $\delta \in (0,1)$ and $c_0 > 0$ be a universal constant. The operator norm of the matrix $\frac{1}{n}\mathfrak{X}_n\mathfrak{X}_n^{\mathrm{T}}$ is well controlled in terms of the covariance matrix Σ:*

$$\mathrm{P}\left\{\sigma_{\max}\left(\frac{1}{n}\mathfrak{X}_n\mathfrak{X}_n^{\mathrm{T}}\right) \leq 2\|\Sigma\|_{\mathrm{op}}\left[1 + \frac{c_0}{1-\gamma}\left(\frac{p}{n} \vee 1 - \frac{\log(\delta)}{n}\right)\right]\right\} \geq 1 - 2\delta \tag{42}$$

and, if $n > p$, for some constants $c_1, c_2 > 0$,

$$\mathrm{P}\left\{\sigma_{\min}\left(\frac{1}{n}\mathfrak{X}_n\mathfrak{X}_n^{\mathrm{T}}\right) \geq \frac{\sigma_{\min}(\Sigma)}{4}\right\} \geq 1 - c_1e^{-c_2p}. \tag{43}$$

Proof. We follow the proof of Lemma 4 from (Negahban and Wainwright 2011). The only difference is that we allow the dimension p to be greater than the number of observations, $p > n$.

Let $u \in \mathbb{R}^p$ with $\|u\|_2 = 1$ and $Y \in \mathbb{R}^n$ be a random vector with elements $Y_i = \langle u, X_i\rangle$. By the covering argument from the proof of Lemma 4 from (Negahban and Wainwright 2011), it can be shown that

$$\mathrm{P}\left\{\frac{1}{n}\sigma_{\max}(\mathfrak{X}_n\mathfrak{X}_n^T) > t\right\} \leq 4^p \max_{u\in\mathcal{A}} \mathrm{P}\left\{\frac{1}{n}\sum_{i=1}^n\langle u, X_i\rangle^2 > t/2\right\},$$

where $\mathcal{A}$ is a 1/2-cover of S^{p-1}. Thus, we need to bound $\dfrac{1}{n}\|Y\|_2^2$. Note that $Y \sim \mathcal{N}_n(0, R)$ with the covariance matrix R with the elements $R_{ij} = u^{\mathrm{T}}\Sigma(\Theta^{\mathrm{T}})^{|i-j|}u$. By the Hanson-Wright inequality, we have

$$\mathrm{P}\left(|\|Y\|_2^2 - \mathrm{tr}(R)\,| > x\right) \leq 2\exp\left[-c\min\left(\frac{x^2}{\|R\|_2^2}, \frac{x}{\|R\|_{\mathrm{op}}}\right)\right].$$

Let $x = \dfrac{2\|R\|_{\mathrm{op}}}{c}\left((p \vee n) - \frac{1}{2}\log\delta\right)$. Then

$$\mathrm{P}\left(\frac{1}{n}\|Y\|_2^2 > \frac{1}{n}\mathrm{tr}(R) + \frac{2\|R\|_{\mathrm{op}}}{cn}\left(p \vee n - \frac{1}{2}\log\delta\right)\right) \leq 2\delta e^{-2(p\vee n)}.$$

Using the bounds $\|R\|_2^2 \le n\|R\|_{\text{op}}^2$, $\|R\|_{\text{op}} \le \dfrac{2\|\Sigma\|_{\text{op}}}{1-\gamma}$, and $\text{tr}(R)/n \le \|\Sigma\|_{\text{op}}$, we obtain (42). $\square$

Lemma 9. *Let $\mathfrak{Z}_n$ be a $p \times n$ matrix with n i.i.d. Gaussian columns with zero mean and covariance matrix Σ_Z. For any $\delta \in (0,1)$, there exists a universal constant c_* independent of p, n and Σ_Z such that*

$$\mathrm{P}\left\{\|\mathfrak{Z}_n\|_{\text{op}} \le c_1^* \sqrt{\|\Sigma_Z\|_{\text{op}}}\left(\sqrt{n} + \sqrt{p} + \sqrt{\log(2/\delta)}\right)\right\} \ge 1 - \delta.$$

Proof follows from Theorem 4.4.5 in Vershynin (2018).

The following lemma reformulates Lemma 5 in (Negahban and Wainwright 2011).

Lemma 10. *Let $\mathfrak{X}_n = (X_1, \ldots, X_n) \in \mathbb{R}^{p \times n}$ be a realization of n observations of a $\text{VAR}_p(\Theta, \Sigma)$ process (1) and $\mathfrak{Z}_n$ be the corresponding $p \times n$ matrix with i.i.d. columns of Gaussian $\mathcal{N}_p(0, \Sigma_Z)$ noise. There exist universal constants c_2^*, k_1, k_2 (independent of p, n, Σ and Σ_Z) such that*

$$\mathrm{P}\left\{\frac{1}{n}\|\mathfrak{Z}_n \mathfrak{X}_n^{\mathrm{T}}\|_{\text{op}} \ge \frac{c_2^*\|\Sigma\|_{\text{op}}}{1-\gamma}\sqrt{\frac{p}{n}}\right\} \le k_1 \exp\left(-k_2 p\right).$$

The proof of the following lemma is based on the Hanson–Wright inequality.

Lemma 11. *Let $\delta \in (0,1)$ and $A \in \mathbb{R}^{p \times p}$. Let $X = (X_1, \ldots, X_n)$ be a realization of a $\text{VAR}_p(\Theta, \Sigma)$ process defined in (1). Assume that for some universal constant $c > 0$,*

$$\frac{\|A\|_2}{\|A\|_{\text{op}}} \ge \frac{2(1+\gamma^2)\sigma_{\max}(\Sigma)}{\sigma_{\min}(\Sigma_Z) \wedge \sigma_{\min}(\Sigma)} \sqrt{\frac{1}{c}\frac{\log(2/\delta)}{n}}. \tag{44}$$

Then for any $\delta > 0$

$$\mathrm{P}\left[\frac{1}{n}\sum_{t=1}^{n} \|AX_t\|_2^2 \ge \frac{\sigma_{\min}(\Sigma_Z) \wedge \sigma_{\min}(\Sigma)}{2(1+\gamma^2)}\|A\|_2^2\right] \ge 1 - \delta.$$

Proof. Recall that $X_{t+1} = \Theta X_t + Z_t$, where Z_t are i.i.d. Gaussian vectors with the covariance matrix Σ_Z. We can write

$$\sum_{t=1}^{n} \|AX_t\|_2^2 = \|B\mathbb{X}\|_2^2,$$

where B is a $pn \times pn$ block-diagonal matrix with the identical diagonal blocks A and $\mathbb{X}$ is the pn-dimensional Gaussian vector obtained by concatenation of the vectors X_t, $t = 1, \ldots, n$. Denote the covariance matrix of $\mathbb{X}$ by $\mathbb{\Sigma}$, thus $\mathbb{X} \sim \mathcal{N}_{pn}(0, \mathbb{\Sigma})$.

It can be easily seen that the covariance matrix of $\mathbb{X}$

$$\mathbb{\Sigma} = \mathbb{E}(\mathbb{X}\mathbb{X}^{\mathrm{T}}) = \left(\mathbb{E}(X_i X_j^{\mathrm{T}})\right)_{1 \le i,j \le n}$$

is a symmetric block-Toeplitz matrix with the blocks

$$\mathbb{E}(X_i X_j^{\mathrm{T}}) = \Sigma \mathbf{1}_{\{i=j\}} + \Theta^{i-j}\Sigma \mathbf{1}_{\{i>j\}} + \Sigma\big(\Theta^{j-i}\big)^{\mathrm{T}} \mathbf{1}_{\{i<j\}}.$$

Using the representation $\mathbb{X} = \mathbb{\Sigma}^{1/2} W$, where $W \sim \mathcal{N}_{pn}(0, I_{pn})$ is a pn-dimensional Gaussian vectors with i.i.d. entries, we see that $B\mathbb{X} = (B\sqrt{\mathbb{\Sigma}})W \sim \mathcal{N}_{pn}(0, B\mathbb{\Sigma}B^{\mathrm{T}})$. Note that $\mathbb{E}(\|B\mathbb{X}\|_2^2) = \mathbb{E}\|B\sqrt{\mathbb{\Sigma}}W\|_2^2) = \mathrm{tr}(B\mathbb{\Sigma}B^{\mathrm{T}})$. Then, using the Hanson-Wright inequality we get

$$\mathrm{P}\left(\left|\|B\mathbb{X}_T\|_2^2 - \mathrm{tr}(B\mathbb{\Sigma}B^{\mathrm{T}})\right| > x\right) \le 2\exp\left[-c\min\left(\frac{x^2}{\|B\mathbb{\Sigma}B^{\mathrm{T}}\|_2^2}, \frac{x}{\|B\mathbb{\Sigma}B^{\mathrm{T}}\|_{\mathrm{op}}}\right)\right], \tag{45}$$

where $c > 0$ is a universal constant.

Note that $\mathrm{tr}(\mathbb{\Sigma}B^{\mathrm{T}}B) \ge \sigma_{\min}(\mathbb{\Sigma})\mathrm{tr}(B^{\mathrm{T}}B)$, since $\mathrm{tr}((\mathbb{\Sigma} - \sigma_{\min}(\mathbb{\Sigma})I_d)B^{\mathrm{T}}B) \ge 0$. Thus,

$$\mathrm{tr}(B\mathbb{\Sigma}B^{\mathrm{T}}) \ge \sigma_{\min}(\mathbb{\Sigma})n\|A\|_2^2.$$

Set $x = \|\mathbb{\Sigma}\|_{\mathrm{op}}\|A\|_{\mathrm{op}}\|A\|_2\sqrt{\log(2/\delta)n/c}$. We have $\|B\mathbb{\Sigma}B^{\mathrm{T}}\|_{\mathrm{op}} \le \|\mathbb{\Sigma}\|_{\mathrm{op}}\|A\|_{\mathrm{op}}^2$. Consequently, the condition

$$\frac{\|A\|_2}{\|A\|_{\mathrm{op}}} \ge \sqrt{\frac{\log(2/\delta)}{nc}} \tag{46}$$

implies $\dfrac{cx}{\|B\mathbb{\Sigma}B^{\mathrm{T}}\|_{\mathrm{op}}} > \log(2/\delta)$. On the other hand, we have

$$\|B\mathbb{\Sigma}B^{\mathrm{T}}\|_2^2 \le \|\mathbb{\Sigma}B^{\mathrm{T}}B\|_2^2\| \le \|\mathbb{\Sigma}\|_{\mathrm{op}}^2\|B\|_{\mathrm{op}}^2\|B\|_2^2 \le \|\mathbb{\Sigma}\|_{\mathrm{op}}^2\|A\|_{\mathrm{op}}^2 n\|A\|_2^2$$

which implies $\dfrac{cx^2}{\|B\mathbb{\Sigma}B^{\mathrm{T}}\|_2^2} \ge \log(2/\delta)$. In addition,

$$\frac{\|A\|_2}{\|A\|_{\mathrm{op}}} \ge 2\kappa(\mathbb{\Sigma})\sqrt{\frac{\log(2/\delta)}{nc}} \tag{47}$$

immediately implies that $\mathrm{tr}(B\mathbb{\Sigma}B^{\mathrm{T}}) - x \ge \dfrac{\mathrm{tr}(B\mathbb{\Sigma}B^{\mathrm{T}})}{2}$. Thus, (45) implies

$$\mathrm{P}\left(\|B\mathbb{X}_T\|_2^2 \le \frac{\sigma_{\min}(\mathbb{\Sigma})}{2}\|A\|_2^2\right) \le \mathrm{P}\left(\|B\mathbb{X}_T\|_2^2 \le \frac{1}{2}\mathrm{tr}(B\mathbb{\Sigma}B^{\mathrm{T}})\right) \le \delta$$

We can bound the maximal eigenvalue of $\mathbb{\Sigma}$ using the decomposition $\mathbb{\Sigma} = \Lambda + \Lambda^{\mathrm{T}}$ into the sum of a lower and upper triangular parts of $\mathbb{\Sigma}$ where the matrices Λ and Λ^{T} have the same diagonal blocks $\frac{1}{2}\Sigma$. Then $\sigma_{\max}(\mathbb{\Sigma}) = \sigma_{\max}(\Lambda + \Lambda^{\mathrm{T}}) \le 2\sigma_{\max}(\Lambda) \le \sigma_{\max}(\Sigma)$.

To bound $\sigma_{\min}(\mathbb{\Sigma})$, note that $\sigma_{\min}(\mathbb{\Sigma}) = (\sigma_{\max}(\mathbb{\Sigma}^{-1}))^{-1}$ where the inverse matrix Σ has a simple block-tridiagonal with constant upper and lower diagonal blocks $-\Theta^{\mathrm{T}}\Sigma_Z^{-1}$ and $-\Theta^{\mathrm{T}}\Sigma_Z^{-1}$, respectively. On the diagonal of this matrix we have the same first $T-1$ blocks $\Theta^{\mathrm{T}}\Sigma_Z^{-1}\Theta + \Sigma^{-1}$ and the last block is equal to Σ_Z^{-1}. Using the same reasoning as above, we see that

$$\begin{aligned}\sigma_{\max}(\mathbb{\Sigma}^{-1}) &\le \max\Big(\|\Theta^{\mathrm{T}}\Sigma_Z^{-1}\Theta + \Sigma^{-1}\|_{\mathrm{op}}, \|\Theta^{\mathrm{T}}\Sigma_Z^{-1}\Theta + \Sigma_Z^{-1}\|_{\mathrm{op}}, \|\Sigma_Z^{-1}\|_{\mathrm{op}}\Big) \\ &\le \gamma^2\|\Sigma_Z^{-1}\|_{\mathrm{op}} + \max(\|\Sigma_Z^{-1}\|_{\mathrm{op}}, \|\Sigma^{-1}\|_{\mathrm{op}}).\end{aligned}$$

Therefore $\sigma_{\min}(\Sigma) \geq \dfrac{\sigma_{\min}(\Sigma_Z) \wedge \sigma_{\min}(\Sigma)}{1+\gamma^2}$ and

$$\kappa(\Sigma) \leq \frac{(1+\gamma^2)\sigma_{\max}(\Sigma)}{\sigma_{\min}(\Sigma_Z) \wedge \sigma_{\min}(\Sigma)}.$$

Finally, these estimates together with condition (44) imply (46) and (47) and the statement of the lemma follows. □

The following lemma is from (Rump 2018):

Lemma 12. *Let A be real or complex $n \times n$ matrix. Then*

1. $|\det(I+A)| \leq \exp\Big\{\mathrm{Re}(\mathrm{tr}(A)) + \frac{1}{2}\|A\|_2^2\Big\}$
2. *Suppose the eigenvalues λ_k of A satisfy $\mathrm{Re}(\lambda_k) > -1$, denote $\mu_k = \min(0, \mathrm{Re}(\lambda_k))$. Then*
$$|\det(I+A)| \geq \exp\Big\{\mathrm{Re}(\mathrm{tr}(A)) - \frac{\frac{1}{2}\|A\|_2^2}{1+\min_k \mu_k}\Big\}$$
3. *Let $\rho(A)$ be spectral radius of A and $\rho(A) < 1$. Then*
$$|\det(I+A)| \geq \exp\Big\{\mathrm{Re}(\mathrm{tr}(A)) - \frac{\frac{1}{2}\|A\|_2^2}{1-\rho(A)}\Big\}.$$

The following lemma is the result of Theorem 1.1 of Tippett et al. (2000) reformulated in the notation of our paper.

Lemma 13. *Let Θ be a stable matrix, $\|\Theta\|_{\mathrm{op}} = \gamma < 1$. Let $\Sigma = \Sigma^{\mathrm{T}} > 0$ be the solution of the Lyapunov equation $\Sigma = \Theta\Sigma\Theta^{\mathrm{T}} + \Sigma_Z$, where Σ_Z is symmetric positive definite. Then*

$$\kappa(\Sigma) \leq \kappa(\Sigma_Z)\frac{1+\gamma^2}{1-\gamma^2}. \tag{48}$$

F Additional simulation results

We consider the situation of uncorrelated Gaussian noise, $\Sigma_Z = \mathrm{id}_p$. We report the results of 100 simulations. The transition matrices Θ_1 and Θ_2 are defined as follows:

$$\Theta_1 = UD_1U^{\mathrm{T}}, \quad \Theta_2 = UD_2U^{\mathrm{T}},$$

where U is a unitary matrix, D_1 and D_2 are diagonal matrices with the largest absolute eigenvalues γ_1 and γ_2, respectively. Thus the Frobenius norm of the change is expressed in terms of the eigenvalues of D_1 and D_2: $\|\Delta\Theta\|_2 = \|D_1 - D_2\|_2$.

F.1 Quantile of the test statistic under the null

To simulate the quantile, we need to sample the test statistic $\mathcal{G}(t)$ under the null with the transition matrices estimated from the data. Since we do not know whether our data contains a change-point or not, we will sample the quantiles of $\mathcal{G}(t)$ based on the distribution $\mathrm{VAR}_p(\widehat{\Theta}_1, \Sigma_Z)$ and on the distribution $\mathrm{VAR}_p(\widehat{\Theta}_2, \Sigma_Z)$, where $\widehat{\Theta}_1$ and $\widehat{\Theta}_2$ are estimations of Θ_1 and Θ_2 obtained from the first $\lfloor Th \rfloor$ and the last $\lfloor Th \rfloor$ observations of the process. To guarantee the stability of simulated processes, these two estimators are adjusted to have the operator norm to be equal to γ_1 and γ_2, respectively. Thus, we will have two quantiles of level $1-\alpha$, the first one is based on the hypothesis that the true parameter matrix is the one of the first $\lfloor Th \rfloor$ observations and the other one is based on the hypothesis that the true parameter matrix equals to the one of the last $\lfloor Th \rfloor$ observations. The quantile simulation from a given data subsample $\breve{\mathfrak{X}}$ is presented in Algorithm 1.

Algorithm 1: Simulation of $1-\alpha$ quantile of $\mathcal{G}(t)$

Data: $\breve{\mathfrak{X}}$, Θ, Σ_Z, α, t
Result: Quantile $q_{\alpha,t}$ of the distrbution of $\mathcal{G}(t)$
S=1500;
s=1;
while $s < S$ **do**
  Sample $\tilde{X}_0$ from $\breve{\mathfrak{X}}$ with replacement;
  Sample observations $\tilde{\mathfrak{X}} = (\tilde{X}_0, \tilde{X}_1, \ldots, \tilde{X}_T)$ from $\mathrm{VAR}_p(\Theta, \Sigma_Z)$;
  Fing the SDP solutions $\widetilde{\Theta}_1^t = \arg\min_{M\in\mathbb{R}^{p\times p}} \varphi_1^t(\tilde{\mathfrak{X}}, M)$ and
  $\widetilde{\Theta}_2^t = \arg\min_{M\in\mathbb{R}^{p\times p}} \varphi_2^t(\tilde{\mathfrak{X}}, M)$;
  Calculate

$$\mathcal{G}_s(t) = \frac{t}{T}\Big(\varphi_1^t(\tilde{\mathfrak{X}}, \widetilde{\Theta}_2^t) - \varphi_1^t(\tilde{\mathfrak{X}}, \widetilde{\Theta}_1^t)\Big) + \frac{T-t}{T}\Big(\varphi_2^t(\tilde{\mathfrak{X}}, \widetilde{\Theta}_1^t) - \varphi_2^t(\tilde{\mathfrak{X}}, \widetilde{\Theta}_2^t)\Big)$$

  s=s+1;
end
Calculate the $1-\alpha$ level empirical quantile $q_{\alpha,t}$ from the sample $\mathcal{G}_1(t), \ldots, \mathcal{G}_S(t)$

F.2 Algorithm of testing for a change-point

We propose the following practical testing procedure. For each value of a possible change-point $t \in \mathcal{T}$, we calculate the value of the test statistic $\mathcal{G}(t)$. The details about estimation of the transition matrices using the SDP programs are given in Section 3. We simulate the empirical quantiles $q_{\alpha,t}^{(1)}$ and $q_{\alpha,t}^{(2)}$ of the test statistic distribution under the null based on the first and last portions of observations. We detect a change if for some $t \in \mathcal{T}$ the test statistic $\mathcal{G}(t)$ is greater that $q_{\alpha,t} = \max(q_{\alpha,t}^{(1)}, q_{\alpha,t}^{(2)})$. The details are presented in Algorithm 2.

Algorithm 2: VAR change-point detection

Data: $\mathfrak{X} = (X_1, \ldots, X_T)$, γ_1, γ_2, $\mathcal{T}$, h

Result: $\psi = 1$ if there is a change-point, otherwise $\psi = 0$

for $t \in \mathcal{T}$ **do**

Estimate $\breve{\Theta}_1$ using $X_1, \ldots, X_{\lfloor Th \rfloor}$;

Estimate $\breve{\Theta}_2$ using $X_{T-\lfloor Th \rfloor+1}, \ldots, X_T$;

Adjustement of estimators:

$$\widehat{\Theta}_1 = \gamma_1 \frac{\breve{\Theta}_1}{\|\breve{\Theta}_1\|_{\mathrm{op}}}, \quad \widehat{\Theta}_2 = \gamma_2 \frac{\breve{\Theta}_2}{\|\breve{\Theta}_2\|_{\mathrm{op}}}$$

Generate quantile $q^{(1)}_{\alpha,t}$ using $\mathfrak{X}^{(1)} = (X_0, \ldots, X_{\lfloor Th \rfloor})$ and $\widehat{\Theta}_1$;

Generate quantile $q^{(2)}_{\alpha,t}$ based on $\mathfrak{X}^{(2)} = (X_{T-\lfloor Th \rfloor}, \ldots, X_T)$ and $\widehat{\Theta}_2$;

Calculate $\widehat{\Theta}^t_1 = \arg\min_{M \in \mathbb{R}^{p \times p}} \varphi^t_1(\mathfrak{X}, M)$ and $\widehat{\Theta}^t_2 = \arg\min_{M \in \mathbb{R}^{p \times p}} \varphi^t_2(\mathfrak{X}, M)$;

Calculate the test statstic

$$\mathcal{G}(t) = \frac{t}{T}\Big(\varphi^t_1(\mathfrak{X}, \widehat{\Theta}^t_2) - \varphi^t_1(\mathfrak{X}, \widehat{\Theta}^t_1)\Big) + \frac{T-t}{T}\Big(\varphi^t_2(\mathfrak{X}, \widehat{\Theta}^t_1) - \varphi^t_2(\mathfrak{X}, \widehat{\Theta}^t_2)\Big)$$

$\psi_t = 0$;

if $\mathcal{G}(t) > q^{(1)}_{\alpha,t} \vee q^{(2)}_{\alpha,t}$ **then**

$\psi_t = 1$

end

t=t+1;

end

$\psi = \max_{t=1,\ldots,T-1} \psi_t$

F.3 Simulation results

We provide the power of the test obtained for different simulation scenarios with the quantiles simulated at the level $\alpha = 0.05$. We consider the VAR processes of dimension $p = 100$.

First, we consider the case of the change-point location in the middle, $\tau = T/2$ for the number of observations T varying from 1500 to 5000 with $\gamma_1 = \gamma_2 = 0.9$ and the ranks $R_1 = R_2 = 4$. The results in Fig. 3 demonstrate that the detection becomes easier for a larger number of observations T. The red curve corresponds to the case of $T = 1500$, the black one stands for $T = 6000$.

Next, we consider the case of $T = 5000$ observations of the VAR process with $\gamma_1 = \gamma_2 = 0.9$ and the ranks $R_1 = R_2 = 5$. The change-point location τ/T varies from 0.1 to 0.9. The results in Fig. 4 show that the detection of the change-point is easier if it is located in the middle and harder when the change-point is close to the interval boundaries. We can also see that the power curves for $\tau/T = 0.1$ and $\tau = 0.9$ are almost identical and the same holds for $\tau/T = 0.3$ and $\tau/T = 0.7$. This simulation result is in line with the definition of the jump energy $q(\tau/T)\|\Delta\Theta\|_2$ that is symmetric with respect to $\tau = T/2$.

Finally, we consider $T = 5000$ observations with fixed values $\gamma_1 = \gamma_2 = 0.9$ and varying

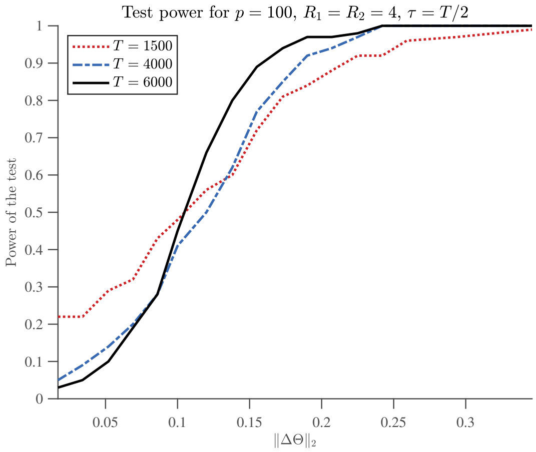

Figure 3: The test power for known location of the change-point and different number of observations T

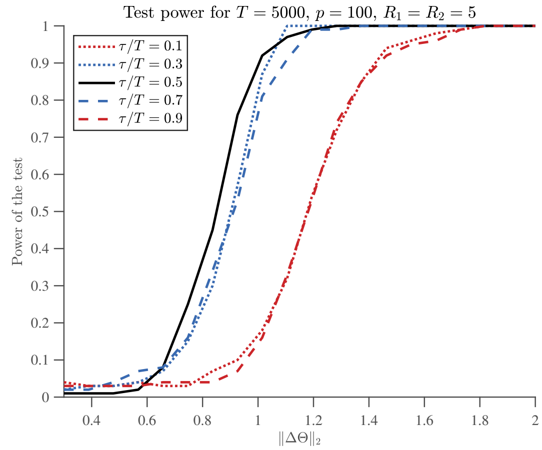

Figure 4: Dependence of the test power on different locations of the change-point

matrix ranks $R_1 = R_2 = R \in \{5, 9, 17, 25\}$. The results in Fig. 5 demonstrate that the detection is easier for smaller rank of the matrices which is in line with the obtained detection rate of order Rp/T.

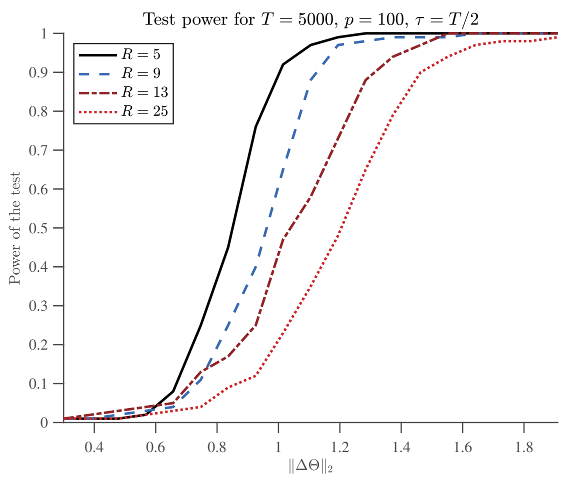

Figure 5: The test power depending on rank of the transition matrices