

# Sequential change-point detection in high-dimensional Gaussian graphical models

**Hossein Keshavarz**

*Institute for Mathematics and its Applications  
University of Minnesota  
Minneapolis, MN 55455, USA*

HKESHAVA@UMN.EDU

**George Michailidis**

*Department of Statistics and UF Informatics Institute  
University of Florida  
Gainesville, FL 32611, USA*

GMICHAIL@UFL.EDU

**Yves Atchadé**

*Department of Mathematics and Statistics  
Boston University  
Boston, MA 02215, USA*

ATCHADE@BU.EDU

**Editor:** Animashree Anandkumar

## Abstract

High dimensional piecewise stationary graphical models represent a versatile class for modelling time varying networks arising in diverse application areas, including biology, economics, and social sciences. There has been recent work in offline detection and estimation of regime changes in the topology of sparse graphical models. However, the online setting remains largely unexplored, despite its high relevance to applications in sensor networks and other engineering monitoring systems, as well as financial markets. To that end, this work introduces a novel scalable online algorithm for detecting an unknown number of abrupt changes in the inverse covariance matrix of sparse Gaussian graphical models with small delay. The proposed algorithm is based upon monitoring the conditional log-likelihood of all nodes in the network and can be extended to a large class of continuous and discrete graphical models. We also investigate asymptotic properties of our procedure under certain mild regularity conditions on the graph size, sparsity level, number of samples, and pre- and post-changes in the topology of the network. Numerical works on both synthetic and real data illustrate the good performance of the proposed methodology both in terms of computational and statistical efficiency across numerous experimental settings.

**Keywords:** Sequential change-point detection, Gaussian graphical models, Pseudo-likelihood, Mini-batch update, Asymptotic analysis

## 1. Introduction

Recent technological advances in data mining have revolutionized the collection of complex and high-resolution financial, biological and social data (Wu et al., 2014). Characterizing and understanding the relationships amongst a large number of variables poses novel methodological and technical challenges. Probabilistic graphical models capture the conditional dependence structure between variables of interest (Wainwright and Jordan, 2008)

and thus have become a standard tool for the aforementioned task. Further, Gaussian processes provide powerful models in many applications, which when combined with the computational advantages of estimating Gaussian Graphical Models (GGM) has rendered them very popular in empirical work (see e.g., (Bessler and Yang, 2003; Jones and West, 2005)).

An undirected probabilistic graphical model comprises of  $p$  nodes (variables of interest  $\mathbf{X} = [X_1, \dots, X_p]^\top$ )  $\{V_1, \dots, V_p\}$ , whose edges represent conditional dependence relationships amongst them; specifically, there is an edge between  $V_i$  and  $V_j$  if they are conditionally dependent given all other variables (nodes). In GGMs,  $\mathbf{X}$  is a zero mean Gaussian random vector with covariance matrix  $\Sigma$  and in the precision matrix  $\Omega := \Sigma^{-1}$ , an edge connects  $V_i$  and  $V_j$  if and only if  $\Omega_{ij} \neq 0$ . In other words, the conditional dependency structure and topology of a GGM can be uniquely encoded by  $\Omega$ . Consequently, a rich literature has been developed around estimating  $\Omega$  in time invariant GGMs given  $n$  i.i.d zero mean observations  $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{R}^p$  from the network. When  $n > p$  the Maximum Likelihood estimator of  $\Omega$  corresponds to the inverse of the empirical covariance matrix of  $\Sigma$ , while in the case of  $n < p$  one needs to impose additional structural assumptions, such as sparsity, on the underlying  $\Omega$  (see (Bühlmann and Van De Geer, 2011) and references therein).

Despite voluminous research on estimating stationary graphical models, in various scenarios the underlying dependency structure dynamically evolves over time. Next, we discuss several real-world problems centering around temporally evolving high-dimensional data with underlying network structure.

- (a) Temporal fluctuations in functional connectivity (FC), which is referred to as dynamic-FC, has recently received a lot of attention in resting-state blood-oxygen level-dependent functional magnetic resonance imaging (rs-fMRI) (Hutchison et al., 2013; Hindriks et al., 2016). The dynamic-FC is commonly investigated using covariance matrix of FC over consecutive windowed segments.
- (b) The similarity pattern of streaming measurements in a large sensor network can be subject to abrupt changes due to anomalous behaviour of an unknown subset of nodes. For instance, Kalitsis et al. (Kallitsis et al., 2016) studied the online detection of false data injection attacks in wide-area smart grid networks.
- (c) Stock return time series occasionally exhibit radical changes associated with stochastic switching between high and low volatility regimes, financial crises or changes in government policy. Despite a rich literature on regime-switching low-dimensional time series models (Bai and Perron, 1998, 2003), there remains an incomplete understanding of this idea for high dimensional financial network data. Note that there are a few recent studies (Atchade and Bybee, 2017; Barnett and Onnela, 2016) on detecting multiple sudden changes in the dependency structure of S&P 500 stocks returns for the period of 1982 – 2000 and 2000 – 2016 as a result of the stock market crash of October 1987 (known as *Black Monday*), beginning of the *great recession* on January 2008, and the bankruptcy declaration day of *Lehman Brothers Holdings Inc.* on September 15, 2008.

Learning temporally evolving dependency structures across a large number of variables requires specification of the mechanism that drives the underlying dynamics. A simple,

analytically tractable and widely applicable mechanism is given by assuming piecewise stationary dynamics subject to unknown break points. This versatile model, which is known as the *change-point* model, has been extensively studied for low dimensional time series during the last few decades; see for example, (Basseville and Nikiforov, 1993; Horváth and Rice, 2014; Aue et al., 2009; Keshavarz et al., 2018) and references therein. However, there is significantly less work towards understanding the algorithmic and theoretical aspects of change-point estimation for high-dimensional time series data.

Change-point detection algorithms are classified into two groups: *offline* and *sequential (online)*. Given the entire data set beforehand, the objective of offline algorithms is to spot the abrupt changes by scanning through the available data. On the other hand, detection and collection of new samples run concurrently in the sequential framework and the goal is to find sudden changes with the smallest delay after they occur. Given independent observations in the high dimensional paradigm, (Li and Chen, 2012; Cai et al., 2013) designed offline two sample-tests for identifying a single change-points in the covariance matrix. In (Kolar et al., 2010; Kolar and Xing, 2012; Gibberd and Roy, 2020) maximizing a regularized *pseudo-likelihood* with *fused-lasso* penalty has been proposed for estimating multiple structural breaks in the inverse covariance matrix of high dimensional sparse GGMs. Roy et al. (Roy et al., 2016) introduced a two-step algorithm for estimating a single abrupt change in the parameters of high dimensional sparse Markov random fields. Given an estimate of sudden change, the parameters before and after the change-point are separately estimated by maximizing an  $\ell_1$  penalized pseudo-likelihood function. A brute force search on a coarse grid is necessary for updating the location of change-point in each step. Solely focusing on GGMs, Atchadé and Bybee (Atchade and Bybee, 2017) proposed an approximate majorize-minimize (MM) algorithm for reducing the computational cost of brute force search. Note that the  $\ell_1$  penalized loss maximization algorithms in (Roy et al., 2016; Atchade and Bybee, 2017) are capable of estimating a single change-point and extension to the case of multiple jumps requires binary segmentation. Note that despite its relatively low computational cost, the binary segmentation is a greedy procedure that is not guaranteed to maximize the log-likelihood function. Finally, Chen et al. (Chen and Zhang, 2015) proposed a non-parametric sequential algorithm for detecting sudden changes in the similarity graph of dependent random variables.

Motivated by applications in sensor networks and financial markets, we propose a novel *sequential algorithm* for detecting sudden changes in sparse inverse covariance structure of high-dimensional GGMs and investigate its theoretical and numerical properties. To the best of our knowledge, this constitutes the first attempt for addressing the problem of online detection of abrupt changes in high-dimensional and sparse graphical models. The sequential change-point detection in an inverse covariance (precision) matrix inevitably associates with two technical challenges. First, unlike the covariance matrix, the elements (or eigenvalues) of the precision matrix cannot be easily characterized by the data, particularly in the “*large p, small n*” framework. Another challenge, which is a distinctive feature of high dimensional sequential detection, arises when the main concern is spotting sudden shifts with reasonable delay. In contrast to offline high-dimensional change-point detection, the number of *post-change* observations is considerably smaller than  $p$ . Therefore unlike penalized likelihood-based abrupt-change estimators in (Atchade and Bybee, 2017; Gibberd and Roy, 2020; Kolar et al., 2010; Kolar and Xing, 2012; Roy et al., 2016), the detection

procedure should be decoupled from estimating the post-change dependence structure of GGM.

**Outline of online detection strategy.** We conclude this section by presenting a concise and high-level introduction to the proposed sequential detection algorithm. Let  $G_t = (V, E_t)$  denote a zero-mean GGM with  $p$  vertices and a time-varying precision matrix  $\Omega^{(t)}$ . For each  $t$ , we observe a single realization of  $G_t$ , which is represented by  $\mathbf{X}_t = [X_{t,1}, \dots, X_{t,p}]^\top$ . We conduct a statistical test to determine whether a jump occurred at time  $t$ , i.e.  $\Omega^{(t)} \neq \Omega^{(t+1)}$ . As detecting small changes can be challenging for complicated objects such as  $\Omega^{(t)}$  (especially for large  $p$ ), our study is hinged on two moderate and prevalent restrictive conditions.

- (a) (*Sparsity*) For regulating the amount of conditional dependence in  $G_t$ ,  $\Omega^{(t)}$  is assumed to be a sparse matrix, for each  $t$ . Some type of sparsity assumption often holds in the aforementioned applied problems.
- (b) (*Detection delay*) Many sequential decision rules rely on utilizing post-change features of the process. We adopt a similar approach by allowing to observe  $w$  samples ahead before making a decision. Namely, we raise an alarm at  $t$  by using two sources of information,  $w$  post-change observations  $(\mathbf{X}_{t+1}, \dots, \mathbf{X}_{t+w})$  and pre-change features such as an  $\ell_1$ -regularized estimate of  $\Omega^{(t)}$ . In previous work on offline change-point learning in sparse graphical models, the estimated jump is of order  $\log p$  distant from the true location of a change-point (e.g., Theorem 8 of (Atchade and Bybee, 2017) or Theorem 1 in (Roy et al., 2016)). So roughly speaking,  $w$  is an online variant of the corresponding abrupt-change estimation error in offline approach.

Comparing the pre- and post-change conditional log-likelihood of all nodes is at the heart of our proposed algorithm. We particularly show that  $-2 \log \mathbb{P}(X_{ts} | X_{ts'} : s' \neq s; \Omega^{(t)}) + \log \Omega_{ss}^{(t)} - \log(2\pi)$  has a  $\chi_1^2$  density. However if a change occurs at time  $t$ , and lasted for at least  $w$  time points, i.e.,  $\Omega^{(t+w)} = \dots = \Omega^{(t+1)} \neq \Omega^{(t)}$ , then

$$\Pi_s^{(t,w)} := \left\{ -2 \log \mathbb{P}(X_{t+r,s} | X_{t+r,s'} : s' \neq s; \Omega^{(t)}) + \log \Omega_{ss}^{(t)} - \log(2\pi) \right\}_{r=1}^w,$$

is a set of i.i.d. random variables that are not centered around one. Indeed there are some nodes for which

$$\beta_s^{(t,w)} := \mathbb{E} \left[ \frac{-2}{w} \sum_{r=1}^w \log \mathbb{P}(X_{t+r,s} | X_{t+r,s'} : s' \neq s; \Omega^{(t)}) + \log \left( \frac{\Omega_{ss}^{(t)}}{2\pi} \right) \mid \exists \text{ abrupt-change at } t \right] \neq 1.$$

Therefore for a non-negative convex function  $f$  for which  $f(x) > 0$  when  $x \neq 1$  (its exact formulation will be introduced in Section 2.1), after proper normalization  $\sum_{s=1}^p f(\beta_s^{(t,w)})$  concentrates around zero, if no jump occurs at time  $t$ . Indeed, we utilize a suitable convex barrier  $f$  for designing a decision rule in order to magnify any deviation from pre-change pseudo-likelihood function. The exact formulation of this idea is postponed to Section 2.2. It is also noteworthy to mention that since the core building blocks of our procedure are based upon comparing the pre- and post-change pseudo-likelihood of the data, we believe that it can be extended to more generic Markov random fields.

We study the performance of our algorithm in two settings: fully known and unknown  $\Omega^{(t)}$ . Although the case of a fully known  $\Omega^{(t)}$  is not realistic, it provides insight in understanding our contribution, as well as the intrinsic complexity of the underlying detection problem. When the pre-change precision matrix is unknown, we require the change-points to be far from each other and the boundary point  $t = 0$ . In this case, we first consistently estimate  $\Omega^{(1)}$  using any well-studied offline estimation algorithm such as QUIC (Hsieh et al., 2014) in the burn-in period ( $t = 1, \dots, n_0$  for some large enough  $n_0$ ). Note that the first abrupt-change is supposed to appear after  $n_0$ . After obtaining an adequate quality estimate of  $\Omega^{(1)}$  in hand, we concurrently run the detection test and update our pre-change precision matrix in a sequential fashion. Our asymptotic analysis reveals that when both  $p$  and sample size grow in such a way that we can consistently estimate the pre-change precision matrix of the GGM, the detection rate does not differ from the ideal case of a fully-known pre-change dependence structure. In other words, the detection power of our online algorithm is robust against small enough estimation errors, which is highly desirable for real applications.

The remainder of the paper is organized as follows: Section 2.1 rigorously formulates the online change-point problem as a hypothesis testing procedure and introduce the required statistical ingredients for understanding the subsequent sections. Section 2.2 is devoted to presenting the proposed detection algorithm for both fully known and unknown pre-change attributes of the GGM. Section 3 is reserved for investigating the behaviour of proposed decision rule under both null (no-change at  $t$ ) and alternative (sudden change at  $t$ ) hypotheses. In Section 4, we study asymptotic properties of our algorithm in the case of unknown pre-change precision matrix (combination of online detection and estimation). Section 5 assesses the performance of proposed algorithm by numerical experiments on synthetic and real data. Section 6 serves as the conclusion and discusses future directions. We prove the main results of the paper in Section 7. Lastly, Appendices A and B contain auxiliary technicalities which are essential for the results in Section 7.

**Notation.**  $\mathbb{1}(\cdot)$ ,  $\wedge$  and  $\vee$  successively refer to indicator function, minimum and maximum operators. We use  $I_m$ ,  $\mathbf{0}_m$  and  $\mathbf{1}_m$  respectively denote the  $m$ -by- $m$  identity matrix, all zeros column vector of length  $m$ , and all ones column vector of length  $m$ .  $S_{++}^{p \times p}$  stands for the set of strictly positive definite  $p \times p$  matrices. For two matrices of the same size  $M$  and  $M'$ ,  $\langle M, M' \rangle := \sum_{i,j} M_{ij} M'_{ij}$  denotes their usual inner product. For  $M \in \mathbb{R}^{p \times p}$  and  $B_1, B_2 \subset \{1, \dots, p\}$ ,  $M_{B_1, B_2} = [M_{ij} : i \in B_1, j \in B_2]$  denote the sub-matrix of  $M$  associated to  $(B_1, B_2)$ -block.  $\text{diag}(M)$  refers to the main diagonal entries of  $M$ . We use the following norms on  $M \in \mathbb{R}^{m \times n}$ .  $\|M\|_{2 \rightarrow 2}$  represents the usual operator norm (largest singular value of  $M$ ). For any  $1 \leq p \leq \infty$ ,  $\|M\|_{\ell_p}$  stands for element-wise  $\ell_p$ -norm of vectorized  $M$ . For any  $x > 0$ ,  $\Gamma(x)$  denotes the Gamma function at  $x$  and  $\psi^{(r)}$  stands for the *poly-gamma function* of order  $r$ , which is defined by

$$\psi^{(r)}(x) = \frac{d^{r+1} \log \Gamma(x)}{dx^{r+1}}, \quad \forall x > 0 \text{ and } r = 0, 1, \dots$$

For two non-negative sequences  $\{a_m\}_{m=1}^\infty$  and  $\{b_m\}_{m=1}^\infty$ , we write  $a_m \lesssim b_m$  if there exists a bounded positive scalar  $C_{\max}$  (depending on model parameters) such that  $\limsup_{m \rightarrow \infty} a_m/b_m \leq C_{\max}$ . Moreover,  $a_m \asymp b_m$  refers to the case that  $a_m \lesssim b_m$  and  $a_m \gtrsim b_m$ . For a non-negative deterministic  $\{a_m\}_{m=1}^\infty$  and random sequence  $\{b_m\}_{m=1}^\infty$ , we write  $b_n = \mathcal{O}_{\mathbb{P}}(a_n)$ ,

if  $\mathbb{P}(b_m \leq C_{\max} a_m) \rightarrow 1$ , as  $m \rightarrow \infty$ , for some bounded positive scalar  $C_{\max}$  (which may depend on model parameters).  $D_{KL}(\mathbb{P}_1 \parallel \mathbb{P}_2)$  represents the *Kullback–Leibler* (KL) divergence between two distributions  $\mathbb{P}_1$  and  $\mathbb{P}_2$ . Lastly for a binary test statistic  $\Xi$ , the false alarm and miss-detection probabilities are respectively defined by

$$\mathbf{P}_{\text{FA}}(\Xi) := \mathbb{P}(\Xi = 1 \mid \mathbb{H}_0), \quad \text{and} \quad \mathbf{P}_{\text{MD}}(\Xi) := \mathbb{P}(\Xi = 0 \mid \mathbb{H}_1).$$

## 2. Problem formulation and detection algorithm

We start by providing a rigorous formulation of the problem at hand and then introduce the detection strategy initially for the case of a known precision matrix before the change point, which provides insight into the technical aspects of the problem, followed by the real world setting of an unknown precision matrix that needs to be estimated in an online fashion from the available data.

### 2.1. Background and setup

Let  $G_t = (V, E_t)$ ,  $t \in \{1, \dots, T\}$  be a time-varying undirected zero mean GGM with (fixed) node set  $V$  of size  $p$ , where  $T$  denotes the number of observed independent samples. For any  $t \in \{1, \dots, T\}$ , we observe a single realization of  $G_t$  represented by  $\mathbf{X}_t = [X_{t,1}, \dots, X_{t,p}]^\top$ . In particular,  $\mathbf{X}_t$  is a  $p$ -variate Gaussian vector with density function

$$g(\mathbf{x}; \Omega^{(t)}) = (2\pi)^{-p/2} \sqrt{\det \Omega^{(t)}} \exp\left(-\frac{\mathbf{x}^\top \Omega^{(t)} \mathbf{x}}{2}\right), \quad \forall \mathbf{x} \in \mathbb{R}^p,$$

where  $\Omega^{(t)}$  denotes the symmetric positive definite precision matrix of  $\mathbf{X}_t$ . Note that  $E_t$  is fully identifiable from the non-zero off-diagonal elements of  $\Omega^{(t)}$ ; namely

$$E_t = \left\{ (\alpha, \alpha') \in V \times V : \alpha \neq \alpha', \Omega_{\alpha, \alpha'}^{(t)} \neq 0 \right\}.$$

Adopting a piecewise constant model for  $\Omega^{(t)}$  is a popular approach to modelling multiple abrupt changes in the dependency structure of  $G_t$ . Specifically, assume that there is a set  $\mathcal{D}^* \subset \{1, \dots, T\}$ , sorted in ascending order and with  $t_0^* = 1$ , such that

$$\Omega^{(t)} = \sum_{j=1}^T \Omega_j \mathbb{I}(t_j^* \leq t < t_{j+1}^*), \quad \forall t \in \{1, \dots, T\}. \quad (1)$$

In Eq. (1)  $\mathcal{D}^*$  stands for the collection of unknown change-points between 1 and  $T$ . As a consequence  $\{\mathbf{X}_t : t_j^* \leq t < t_{j+1}^*\}$  are independent and identically distributed samples drawn from  $g(\cdot; \Omega_j)$ .

Detecting jumps in  $\Omega^{(t)}$  is equivalent to solving  $T$  separate hypothesis testing problems formulated by

$$\mathbb{H}_{0,t} : t \notin \mathcal{D}^* \quad \text{vs.} \quad \mathbb{H}_{1,t} : t \in \mathcal{D}^*, \quad \forall t \in \{1, \dots, T\}. \quad (2)$$

In an offline setting, we observe the entire set  $\{\mathbf{X}_t\}_{t=1}^T$  prior to test any hypothesis in (2). Simply put, a binary decision function  $\Xi_t(\mathbf{X}_1, \dots, \mathbf{X}_T) \in \{0, 1\}$  is designed for disambiguating  $\mathbb{H}_{0,t}$  from  $\mathbb{H}_{1,t}$  with false alarm rate below some  $\pi_0 \in (0, 1)$ , for any  $t \in \{1, \dots, T\}$ . On

the other hand, in the online regime,  $\Xi_t$  solely depends on  $\{\mathbf{X}_i\}_{i=1}^{t+w}$ , for some pre-specified delay  $w$ . Henceforth, we focus exclusively on the online setting.

For better understanding the rationale behind our proposed detection algorithm, we make a succinct overview of two key statistical concepts: KL-divergence and conditional log-likelihood of multivariate Gaussian vectors. We first introduce an alternative formulation of the KL-divergence between underlying distributions under  $\mathbb{H}_{0,t}$  and  $\mathbb{H}_{1,t}$ . From a statistical viewpoint, the false alarm of  $\Xi_t$ , regardless of its complexity, is not negligible when the KL-divergence between  $\mathbb{P}(\mathbf{X}_1, \dots, \mathbf{X}_{t+w} \mid \mathbb{H}_{0,t})$  and  $\mathbb{P}(\mathbf{X}_1, \dots, \mathbf{X}_{t+w} \mid \mathbb{H}_{1,t})$  is small. For ease of presentation, we consider a simple scenario. Suppose that the inverse covariance matrix of  $G_t$  switches from  $\Omega_1$  to  $\Omega_2$  at time  $t$  and it remains in the new regime until  $t+w$ , i.e.  $\mathcal{D}^* \cap \{t, \dots, t+w\} = \{t\}$ . For notational convenience let  $\Psi := \Omega_2^{-1/2} \Omega_1 \Omega_2^{-1/2}$  and define  $f : (0, \infty) \mapsto [0, \infty)$  by

$$f(x) := x - 1 - \log x, \quad \forall x > 0, \quad (3)$$

Since  $\{\mathbf{X}_r\}_{r=1}^{t+w}$  independent and zero-mean random vectors under both  $\mathbb{H}_{0,t}$  and  $\mathbb{H}_{1,t}$ , the KL-divergence between  $\mathbb{P}(\mathbf{X}_1, \dots, \mathbf{X}_{t+w} \mid \mathbb{H}_{1,t})$  and  $\mathbb{P}(\mathbf{X}_1, \dots, \mathbf{X}_{t+w} \mid \mathbb{H}_{0,t})$ , which is denoted by  $\mathbb{D}$ , can be written as

$$\begin{aligned} \mathbb{D} &= \sum_{1 \leq r \leq w} D_{KL} \left( \mathbb{P}(\mathbf{X}_{t+r} \mid \mathbb{H}_{1,t}) \parallel \mathbb{P}(\mathbf{X}_{t+r} \mid \mathbb{H}_{0,t}) \right) = w D_{KL} \left( \mathbb{P}(\mathbf{X}_{t+1} \mid \mathbb{H}_{1,t}) \parallel \mathbb{P}(\mathbf{X}_{t+1} \mid \mathbb{H}_{0,t}) \right) \\ &= w \left[ \text{tr} \left( \Omega_2^{-1} \Omega_1 \right) - \log \left( \frac{\det \Omega_1}{\det \Omega_2} \right) - p \right] = w \left( \text{tr}(\Psi) - \log \det \Psi - p \right) = w \sum_{j=1}^p f(\lambda_j(\Psi)). \end{aligned} \quad (4)$$

The conditional log-likelihood of a GGM is another key component in the proposed detection algorithm. It is known that for a Gaussian random vector  $\mathbf{X} = [X_1, \dots, X_p]^\top \sim \mathcal{N}(\mathbf{0}_p, \Sigma = \Omega^{-1})$ , the log-likelihood function of  $X_s$  given  $X_{-s}$  satisfies the following identity for any  $s \in \{1, \dots, p\}$ .

$$\begin{aligned} Z_s &:= -2 \log \mathbb{P}(X_s \mid X_{-s}) - \log \left( \frac{2\pi}{\Omega_{ss}} \right) = \Omega_{ss} \left( X_s + \sum_{t \neq s} \frac{X_t \Omega_{st}}{\Omega_{ss}} \right)^2 = \Omega_{ss} \left( \frac{\sum_{t=1}^p X_t \Omega_{st}}{\Omega_{ss}} \right)^2 \\ &= \frac{\langle \mathbf{X}, \Omega_{s,:} \rangle^2}{\Omega_{ss}}. \end{aligned} \quad (5)$$

Note that  $Z_s$  represents the data dependent part of the conditional negative log-likelihood. Moreover,

$$\text{var}(\langle \mathbf{X}, \Omega_{:,s} \rangle) = \Omega_{s,:} \text{cov}(\mathbf{X}) \Omega_{:,s} = \Omega_{s,:} \Omega^{-1} \Omega_{:,s} = (\Omega \Omega^{-1} \Omega)_{ss} = \Omega_{ss}.$$

That is,  $\{Z_s : s = 1 \dots, p\}$  forms a class of dependent  $\chi_1^2$  random variables.

## 2.2. Detection algorithm for a known pre-change precision matrix

Next, we introduce a novel online algorithm for detecting abrupt changes in  $\Omega^{(t)}$  with delay  $w$ ; namely, determine whether  $(t+1) \in \mathcal{D}^*$ , given  $\{\mathbf{X}_1, \dots, \mathbf{X}_t, \dots, \mathbf{X}_{t+w}\}$ . To gain insights into the nature of the problem, we first consider the oracle framework in which the pre-change precision matrix  $\Omega^{(t)}$  is fully known, that allows us to solely focus on the

detection procedure without having to estimate the pre-change parameters. The setting of an unknown  $\Omega^{(t)}$  is addressed in Section 2.3.

As we discussed in Section ??, the proposed test statistic is motivated by employing a network based pseudo-likelihood function. For any  $s \in \{1, \dots, p\}$ , define

$$Y_s^{(t,w)} := \frac{1}{w\Omega_{ss}^{(t)}} \sum_{r=1}^w \langle \mathbf{X}_{t+r}, \Omega_{:,s}^{(t)} \rangle^2. \quad (6)$$

Based on Eq. (5),  $Y_s^{(t,w)}$  is indeed the empirical average of the (shifted) negative conditional log-likelihood of node  $V_s$  given all the other nodes in  $G_t$ . When dealing with GGMs,  $Y_s^{(t,w)}$  is a linear combination of  $w$  independent quadratic forms of Gaussian random variables. Next, we investigate the distribution of  $Y_s^{(t,w)}$  under  $\mathbb{H}_{0,t}$ . In this case  $\Omega^{(t)} = \Omega^{(t+1)} = \dots = \Omega^{(t+w)}$ . Thus, for any  $r \in \{1, \dots, w\}$ ,  $\langle \mathbf{X}_{t+r}, \Omega_{:,s}^{(t)} \rangle$  is a centered Gaussian random variable and its variance is given by

$$\begin{aligned} \text{var} \left( \langle \mathbf{X}_{t+r}, \Omega_{:,s}^{(t)} \rangle \right) &= \Omega_{s,:}^{(t)} \text{cov}(\mathbf{X}_{t+r}) \Omega_{:,s}^{(t)} = \Omega_{s,:}^{(t)} \left[ \Omega^{(t)} \right]^{-1} \Omega_{:,s}^{(t)} = \left( \Omega^{(t)} \left[ \Omega^{(t)} \right]^{-1} \Omega^{(t)} \right)_{ss} \\ &= \Omega_{ss}^{(t)}. \end{aligned} \quad (7)$$

Identity (7) reveals that when  $\mathbb{H}_{0,t}$  holds, then

$$\frac{\langle \mathbf{X}_{t+r}, \Omega_{:,s}^{(t)} \rangle}{\sqrt{\Omega_{ss}^{(t)}}} \sim \mathcal{N}(0, 1) \implies wY_s^{(t,w)} \sim \chi_w^2, \quad \forall s \in \{1, \dots, p\}.$$

According to the strong law of large numbers,  $Y_s^{(t,w)}$  concentrates around one with high probability for all  $s = 1, \dots, p$ , as  $w$  increases. In contrast, under  $\mathbb{H}_{1,t}$ , whenever  $\Omega^{(t)} \neq \Omega^{(t+1)} = \dots = \Omega^{(t+w)}$ , the expected value of  $Y_s^{(t,w)}$  is given by

$$\mathbb{E} \left( Y_s^{(t,w)} \mid \mathbb{H}_{1,t} \right) = \frac{1}{w} \sum_{r=1}^w \frac{\left( \Omega^{(t)} \left[ \Omega^{(t+1)} \right]^{-1} \Omega^{(t)} \right)_{ss}}{\Omega_{ss}^{(t)}} = \frac{\left( \Omega^{(t)} \left[ \Omega^{(t+1)} \right]^{-1} \Omega^{(t)} \right)_{ss}}{\Omega_{ss}^{(t)}}. \quad (8)$$

**Remark 1** Define  $\Psi^{(t)} \in S_{++}^{p \times p}$  by  $\Psi^{(t)} := \left[ \Omega^{(t)} \right]^{1/2} \left[ \Omega^{(t+1)} \right]^{-1} \left[ \Omega^{(t)} \right]^{1/2}$ . The formulation of KL-divergence between multi-variate Gaussian densities in Eq. (4) shows that distinguishing  $\mathbb{H}_{0,t}$  and  $\mathbb{H}_{1,t}$  is not possible when  $\Psi^{(t)} = I_p$ . On the other hand, a careful comparison between Eq. (7) and (8) reveals a slightly stronger condition for distinguishing  $\mathbb{H}_{0,t}$  from  $\mathbb{H}_{1,t}$ . Strictly speaking, conditional log-likelihood terms can not differentiate null and alternative hypotheses, when  $\text{diag}(\Psi^{(t)}) = \mathbf{1}_p$ . In words, a test based on the pseudo-likelihood function rules out the possibility of detecting abrupt changes at  $t+1$  when

$$\Psi^{(t)} \neq I_p, \quad \text{and} \quad \Psi_{ss}^{(t)} = 1, \quad \forall s = 1, \dots, p.$$

Although this may seem a statistical drawback of using the pseudo-likelihood function for detection purposes, we believe that such situations do not arise in many practical scenarios involving high-dimensional streaming network data. Therefore for scalability purposes, it is still of great interest to design online change-point detection algorithms based on the pseudo-likelihood function. Further, such an algorithm (unlike likelihood-based detection algorithms) can be also applicable to more general Markov random field models.



Next, we introduce our online detection algorithm. Recall function  $f$  from Eq. (3) and consider the following test statistic.

$$T_t = \frac{\sum_{s=1}^p f\left(Y_s^{(t,w)}\right) - \mathbb{E}\left[\sum_{s=1}^p f\left(Y_s^{(t,w)}\right) \middle| \mathbb{H}_{0,t}\right]}{\text{std}\left[\sum_{s=1}^p f\left(Y_s^{(t,w)}\right) \middle| \mathbb{H}_{0,t}\right]}.$$

For a pre-specified  $\pi_0 \in (0, 1)$  (denoting the desirable false alarm rate) and a pre-determined critical value  $\zeta_{\pi_0}$ , we design the following binary decision function

$$\Xi_t = \mathbf{1}(T_t \geq \zeta_{\pi_0}), \quad (9)$$

for selecting between  $\mathbb{H}_{0,t}$  and  $\mathbb{H}_{1,t}$ . Notice that we raise an abrupt change flag, whenever  $\Xi_t = 1$ . Namely,  $\Omega^{(t)} = \Omega^{(t+1)} = \dots = \Omega^{(t+w)}$ , as long as  $T_t$  is strictly less than  $\zeta_{\pi_0}$ .

Before looking closely into technical aspects of the proposed algorithm, we concisely explain the rationale behind our approach. We argued that if the dependence structure of  $G_t$  does not vary at  $t+1$ , all random variables  $Y_s^{(t,w)}$ ,  $s = 1, \dots, p$  lie in a neighborhood of one. We also selected a non-negative strongly convex barrier function  $f$ , whose single root is at  $x = 1$ . Hence,  $f(Y_s^{(t,w)})$  lies around zero, for any  $s$ . In contrast, when the network undergoes an abrupt change,  $\mathbb{E}f(Y_s^{(t,w)})$  is strictly positive for some nodes  $s \in \{1, \dots, p\}$ . As a result,  $T_t$  exhibits relatively larger values under the alternative hypothesis. Finally, due to the convexity of  $f$ , the deviation between  $\mathbb{H}_{0,t}$  and  $\mathbb{H}_{1,t}$  is more pronounced in  $T_t$  for stronger changes in  $\Omega^{(t)}$ .

The first question that needs to be addressed is how to simply standardize  $\sum_{s=1}^p f(Y_s^{(t,w)})$  under  $\mathbb{H}_{0,t}$ .

**Remark 2** *One can easily justify that under  $\mathbb{H}_{0,t}$ ,  $\langle \mathbf{X}_{t+r}, \Omega_{:,s}^{(t)} \rangle / \sqrt{\Omega_{ss}^{(t)}}$ ,  $r = 1, \dots, w$ , form a set of i.i.d. standard Gaussian random variables, for any  $s \in \{1, \dots, p\}$ . Thus,  $wY_s^{(t,w)}$  has a chi-square density with  $w$  degrees of freedom. Lemma 18 states that*

$$\begin{aligned} g_1(w) &:= \mathbb{E}\left[f\left(Y_s^{(t,w)}\right) \middle| \mathbb{H}_{0,t}\right] = \log\left(\frac{w}{2}\right) - \psi^{(0)}\left(\frac{w}{2}\right), \\ g_2(w) &:= \text{std}\left[f\left(Y_s^{(t,w)}\right) \middle| \mathbb{H}_{0,t}\right] = \sqrt{\psi^{(1)}\left(\frac{w}{2}\right) - \frac{2}{w}}, \quad \forall s = 1, \dots, p. \end{aligned} \quad (10)$$

So  $T_t$  can be rewritten in the following form.

$$T_t = \frac{\sum_{s=1}^p f\left(Y_s^{(t,w)}\right) - pg_1(w)}{\text{std}\left[\sum_{s=1}^p f\left(Y_s^{(t,w)}\right) \middle| \mathbb{H}_{0,t}\right]}.$$

$g_1(\cdot)$  and  $g_2(\cdot)$  are respectively exhibited in the left and right panels of Figure 1. It is obvious from Figure 1 that both  $g_1(w)$  and  $g_2(w)$  converge to zero, as  $w \rightarrow \infty$ . In particular, it is known that<sup>1</sup>

$$\lim_{w \rightarrow \infty} wg_1(w) = 1, \quad \text{and} \quad \lim_{w \rightarrow \infty} wg_2(w) = \sqrt{2}.$$

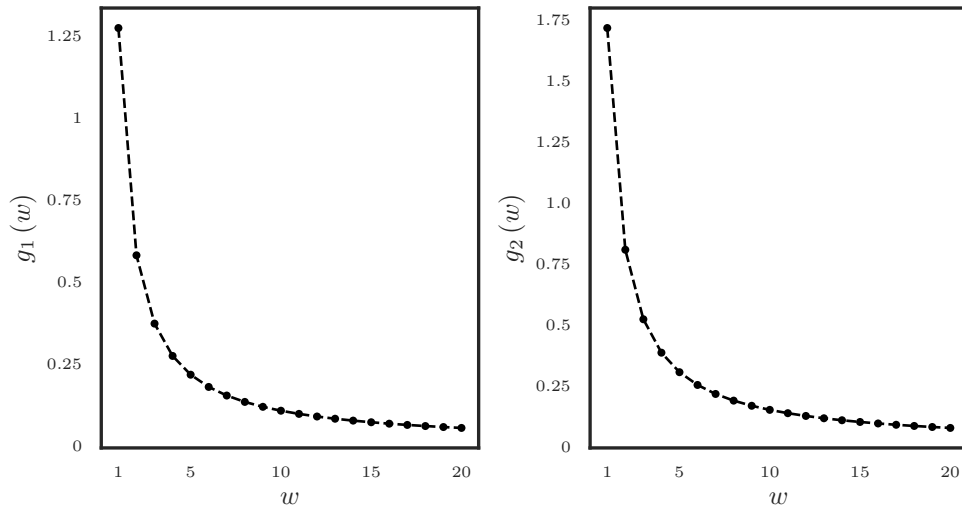


Figure 1: black dots in the left and right panels respectively exhibit  $g_1(w)$  and  $g_2(w)$  for  $w \in \{1, \dots, 20\}$ .

Next, we evaluate the denominator of  $T_t$ . To do so, we first define the partial correlation matrix of  $G_t$ , denoted by  $R^{(t)}$ .

$$R^{(t)} := \left[ \frac{\Omega_{s_1, s_2}^{(t)}}{\sqrt{\Omega_{s_1, s_1}^{(t)} \Omega_{s_2, s_2}^{(t)}}} \right]_{s_1, s_2=1}^p \in \mathbb{R}^{p \times p}.$$

Straightforward calculations show that for standard Gaussian random variables  $Z_1$  and  $Z_2$  with correlation  $r$ ,  $\text{cov}(Z_1^2, Z_2^2) = 2r^2$ . This fact together with the linearity property of the covariance function imply that

$$\text{cov}\left(Y_{s_1}^{(t,w)}, Y_{s_2}^{(t,w)} \mid \mathbb{H}_{0,t}\right) = \frac{2}{w} \left(R_{s_1, s_2}^{(t)}\right)^2, \quad \forall s_1, s_2 \in \{1, \dots, p\}. \quad (11)$$

Simply put,  $R^{(t)}$  encodes the dependence between conditional log-likelihood functions in  $G_t$ . Now employing the second identity in Eq. (10) yields

$$\text{var} \left[ \sum_{s=1}^p f\left(Y_s^{(t,w)}\right) \mid \mathbb{H}_{0,t} \right] = g_2^2(w) \sum_{s_1, s_2=1}^p \text{corr} \left[ f\left(Y_{s_1}^{(t,w)}\right), f\left(Y_{s_2}^{(t,w)}\right) \mid \mathbb{H}_{0,t} \right].$$

Finding a closed form expression for the correlation between  $f(Y_{s_1}^{(t,w)})$  and  $f(Y_{s_2}^{(t,w)})$ , under  $\mathbb{H}_{0,t}$ , demands algebraically cumbersome manipulations. However, it is not too difficult to see that it only depends on  $R_{s_1, s_2}^{(t)}$  and  $w$ . Namely, there exists  $h_w : [-1, 1] \mapsto [-1, 1]$  such that

$$\text{corr} \left[ f\left(Y_{s_1}^{(t,w)}\right), f\left(Y_{s_2}^{(t,w)}\right) \mid \mathbb{H}_{0,t} \right] = h_w \left(R_{s_1, s_2}^{(t)}\right). \quad (12)$$

---

1. We refer the reader to 5.11.2 of <https://dlmf.nist.gov/5.11> and 5.15.8 of <https://dlmf.nist.gov/5.15>

**Remark 3** We employ numerical techniques for approximating  $h_w$ . Figure 2 displays  $h_w$  for different values of  $w$ . The plot in the right panel in Figure 2 depicts that  $h_w(r) \approx r^4$  for  $w \geq 10$ . In Lemma 17, we rigorously substantiate this observation by introducing a uniform upper bound on  $|h_w(r) - r^4|$ . Specifically, we show that

- $h_w$  passes the origin and  $h_w(1) = h_w(-1) = 1$ .
- As  $w \rightarrow \infty$ ,  $\max_{r \in [-1,1]} |h_w(r) - r^4| = \mathcal{O}(w^{-1})$ .

Approximating  $h_w(r)$  with  $r^4$  offers the following numerically convenient proxy for  $T_t$  for large enough  $w$ .

$$T_t = \frac{\sum_{s=1}^p [f(Y_s^{(t,w)}) - g_1(w)]}{g_2(w) \sqrt{\sum_{s_1, s_2=1}^p h_w(R_{s_1, s_2}^{(t)})}} \approx \frac{\sum_{s=1}^p [f(Y_s^{(t,w)}) - g_1(w)]}{g_2(w) \|R^{(t)}\|_{\ell_4}^2}. \quad (13)$$

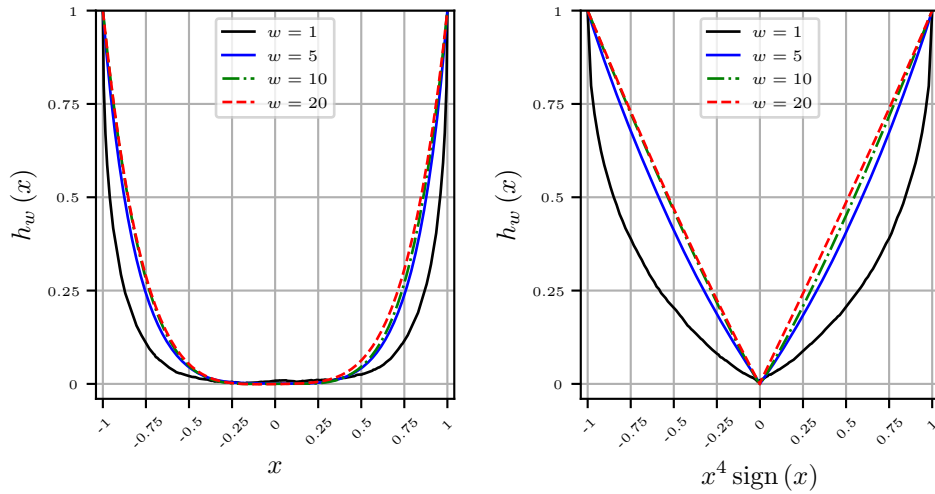


Figure 2:  $h_w(x)$  versus  $x$  and  $x^4 \text{sign}(x)$  for  $w = 1, 5, 10$  and  $20$ .

In summary, when  $\Omega^{(t)}$  is fully known, we propose the following sequential test for detecting a sudden change at time  $t + 1$ .

$$\Xi_t = \mathbb{1} \left( \frac{\sum_{s=1}^p [f(Y_s^{(t,w)}) - g_1(w)]}{g_2(w) \sqrt{\sum_{s_1, s_2=1}^p h_w(R_{s_1, s_2}^{(t)})}} \geq \zeta_{\pi_0} \right) \quad (14)$$

### 2.3. Detection algorithm: Unknown pre-change precision matrix

The precision matrix is usually unknown before the change-point and needs to be estimated from the data. In this case,  $T_t$  is approximated by plugging into Eq. (13) a positive

definite estimate of  $\Omega^{(t)}$ . Specifically, let  $\hat{\Omega}^{(t)}$  be an estimate of  $\Omega^{(t)}$ . Then, the partial correlation matrix  $R^{(t)}$  can be estimated by

$$\hat{R}^{(t)} = \left[ \frac{\hat{\Omega}_{ij}^{(t)}}{\sqrt{\hat{\Omega}_{ii}^{(t)} \hat{\Omega}_{jj}^{(t)}}} \right]_{i,j=1}^p.$$

Moreover for any  $s \in \{1, \dots, p\}$ , we can estimate  $Y_s^{(t,w)}$  (recall it from Eq. (6)) in the following way.

$$\hat{Y}_s^{(t,w)} := \frac{1}{w \hat{\Omega}_{ss}^{(t)}} \sum_{r=1}^w \langle \mathbf{X}_{t+r}, \hat{\Omega}_{:,s}^{(t)} \rangle^2.$$

When we have access to  $\hat{R}^{(t)}$  and  $\left\{ \hat{Y}_s^{(t,w)} \right\}_{s=1}^p$ , we propose to estimate  $T_t$  and  $\Xi_t$  by

$$\hat{T}_t = \frac{\sum_{s=1}^p \left[ f \left( \hat{Y}_s^{(t,w)} \right) - g_1(w) \right]}{g_2(w) \sqrt{\sum_{s_1, s_2=1}^p h_w \left( \hat{R}_{s_1, s_2}^{(t)} \right)}}, \quad \hat{\Xi}_t = \mathbf{1} \left( \hat{T}_t \geq \zeta_{\pi_0} \right). \quad (15)$$

As seen in Eq. (15), the key to approximating  $T_t$  is the availability of a good estimate of  $\Omega^{(t)}$ , which is a challenging task, especially in settings where consecutive change-points are close to each other. Thus, adequate separation in time between two consecutive change-points should be present for obtaining a good approximation of the proposed test statistic. The following assumption formalizes this notion. Recall that  $t_j^*$  denotes the location of  $j$ -th change-point and for notational consistency, we choose  $t_0^* = 1$ .

**Assumption 2.1** *For each  $j \in \mathbb{N}$ , there exists a large enough  $n_j \in \mathbb{N}$  (depending on  $p, w$  and the sparsity pattern of network between  $j$ -th and  $(j-1)$ -th abrupt changes) such that*

$$|t_j^* - t_{j-1}^*| > n_j, \quad \forall j \geq 1.$$

Note that Assumption 2.1 generalizes the boundary condition for the offline single change-point estimation problem (see Assumption 2 in (Roy et al., 2016)). For the time being, selection of  $n_j$  is postponed for later sections. We refer to the first  $n_j$  samples after  $t_j^*$  as the *burn-in period*. We also assume that there exists a fixed, bounded window size  $n_0$  such that  $n_j = n_0$  for any  $j \geq 1$ .

We first provide the intuition behind the algorithm. Detecting each change-point goes through two phases: *warm-up* and *detection-estimation cycle*. For simplicity, we only focus on detecting the first abrupt change (located at  $t_1^*$ ). Note that for each  $t \leq t_1^*$ ,  $\Omega^{(t)} = \Omega^{(1)}$ .

- (a) (warm-up) We estimate  $\Omega^{(1)}$ , which is denoted by  $\hat{\Omega}^{(1)}$ , using  $\mathbf{X}_1, \dots, \mathbf{X}_{n_0}$ .
- (b) For any  $t > n_0$ , as long as  $\hat{\Xi}_t = 0$ , we use  $\mathbf{X}_t$  for updating  $\hat{\Omega}^{(1)}$ . In contrast, if  $\hat{\Xi}_t = 1$  (an abrupt change at  $t$ ), then we wait for  $\mathbf{X}_{t+1}, \dots, \mathbf{X}_{t+n_0}$  for estimating the post-change inverse covariance matrix.

Next, we describe our approach to updating the estimated pre-change precision matrix in phase (b). Again for simplicity, we only focus on updating  $\Omega^{(1)}$ .

**Updating  $\Omega^{(1)}$ .** A mini-batch procedure is used for updating  $\Omega^{(1)}$ , wherein we first obtain  $B$  (a predetermined block size) new samples and subsequently a new estimate at time  $t = n + kB$  ( $k \in \mathbb{N}$ ) by employing  $\mathbf{X}_1, \dots, \mathbf{X}_{n_0+kB}$ ; the parameter  $k$  tracks the number of size- $B$  batches before the first abrupt change. Throughout this paper, we employ the CLIME algorithm (Cai et al., 2011) or alternatively the QUIC estimator (Hsieh et al., 2014) for estimating  $\Omega^{(1)}$ , since both enjoy desirable theoretical and numerical properties. The detailed pseudocode of this proposed procedure is presented in *Algorithm 1*.

---

**Algorithm 1** Online detection with batch update of pre-change precision matrix

---

**Input:**  $n_0, w, B, \zeta_{\pi_0}$  and tuning parameter  $\tau$

**Initialization** Set  $\hat{\mathcal{D}} = \emptyset$  and  $b = 0$ . Given  $\mathbf{X}_1, \dots, \mathbf{X}_{n_0}$ , obtain  $\Omega^{(1)}$  by CLIME or QUIC with tuning parameter  $\tau$ . Also set  $\hat{t}_{last} = 0$ , where  $\hat{t}_{last}$  denotes the estimated location of the last change-point.

**Iterate** For  $t = (n_0 + 1), \dots, T$

Set  $\Xi_t = \mathbb{1}(\hat{T}_t \geq \zeta_{\pi_0})$ .

If  $\Xi_t = 0$  (no change-point)

$b \leftarrow b + 1$  and  $t \leftarrow t + 1$ .

If  $b = B$  (Update pre-change precision matrix)

Obtain  $\hat{\Omega}^{(t)}$  using the CLIME or QUIC methods to  $\mathbf{X}_{1+\hat{t}_{last}}, \dots, \mathbf{X}_{t-1}, \mathbf{X}_t$ .

$b \leftarrow 0$ .

Else

$\hat{\Omega}^{(t)} = \hat{\Omega}^{(t-1)}$

Else

$\hat{t}_{last} = t$  and  $\hat{\mathcal{D}} \leftarrow \hat{\mathcal{D}} \cup \{\hat{t}_{last}\}$ .

Given  $\mathbf{X}_t, \dots, \mathbf{X}_{t+n_0-1}$ , estimate post-change precision matrix using CLIME or QUIC methods.

$t \leftarrow t + n_0$  and  $b \leftarrow 0$ .

**Output:**  $\hat{\mathcal{D}}$

---

### 3. Asymptotic analysis of $\Xi_t$ for a fully known $\Omega^{(t)}$

Next, we establish large-sample properties of the proposed test  $\Xi_t$  introduced in Eq. (14), under both the null and alternative hypotheses. The section addresses the following three issues:

- Choosing the critical value of the test,  $\zeta_{\pi_0}$ , for a fixed false alarm probability  $\pi_0 \in (0, 1)$ .
- Investigating the false alarm and detection power of  $\Xi_t$  for large sparse graphs.
- Introducing the key concepts for the asymptotic analysis of the algorithm in the realistic case of an unknown pre-change precision matrix.

We present the obtained results in two sub-sections. Section 3.1 studies the distribution of  $T_t$  (recall its formulation from Eq. (13)) under the null hypothesis  $\mathbb{H}_{0,t}$  and introduces a

simple way of choosing the critical value  $\zeta_{\pi_0}$ . In Section 3.2, we investigate the statistical power of  $T_t$  for detecting small changes in the structure of  $\Omega^{(t)}$ . We start by defining the set of well-behaved precision matrices.

**Definition 4** *Let  $M$  and  $\alpha_{\min}$  be bounded and strictly positive scalars. Define  $\mathcal{C}_{++}^{p \times p}(\alpha_{\min}, M)$  by*

$$\mathcal{C}_{++}^{p \times p}(\alpha_{\min}, M) = \left\{ A \in \mathcal{S}_{++}^{p \times p} : \|A\|_{1 \rightarrow 1} \leq M, \lambda_{\min}(A) \geq \alpha_{\min} \right\}.$$

For theoretical purposes, throughout this section both pre- and post-change precision matrices are assumed to belong to  $\mathcal{C}_{++}^{p \times p}(\alpha_{\min}, M)$  for some bounded scalars  $M$  and  $\alpha_{\min}$ . Namely, although an abrupt change may affect the topology of the network, the post-change precision matrix is still well-behaved, exhibiting a bounded condition number and a sparse edge set.

### 3.1. Distribution of $T_t$ under $\mathbb{H}_{0,t}$

Precise evaluation of null distribution is extremely challenging, due to the complex nature of  $T_t$ . Therefore, for approximating  $\zeta_{\pi_0}$ , especially for large networks, we inevitably focus on finding the asymptotic null distribution of  $T_t$ . Let  $d_{\max}$  and  $\bar{d}$  respectively denote the maximum and average degree of  $G_t$ . Namely

$$d_{\max} := \max_{s_1=1, \dots, p} \left| \left\{ s_2 : \Omega_{s_1, s_2}^{(t)} \neq 0 \right\} \right|, \quad \text{and} \quad \bar{d} := \frac{1}{p} \sum_{s_1=1}^p |\{s_2 : \Omega_{s_1, s_2} \neq 0\}|.$$

For brevity, the dependence of  $p$  and  $t$  in  $d_{\max}$  and  $\bar{d}$  are dropped. Throughout the remainder of the paper, *asymptotic regime* refers to the scenario that  $p$ , and possibly  $d_{\max}$  and  $\bar{d}$ , tend to infinity.

**Assumption 3.1** *The following conditions hold in the asymptotic regime.*

- (a)  $\Omega^{(t)} \in \mathcal{C}_{++}^{p \times p}(\alpha_{\min}, M)$  for some fixed, bounded and strictly positive scalars  $M$  and  $\alpha_{\min}$ .
- (b)  $\bar{d}d_{\max}$  grows slower than  $\sqrt{p}$ , i.e.,  $\limsup_{p \rightarrow \infty} \frac{d_{\max}\bar{d}}{\sqrt{p}} = 0$ .

Next, we present the first main result of this section.

**Theorem 5** *Suppose that there is no change-point between  $t$  and  $t+w$ , i.e.  $\Omega^{(t)} = \Omega^{(t+1)} = \dots = \Omega^{(t+w)}$ . If  $\Omega^{(t)}$  satisfies Assumption 3.1, then*

$$T_t \xrightarrow{d} \mathcal{N}(0, 1).$$

According to Theorem 5,  $T_t$  converges in distribution to a standard Gaussian random variable under certain asymptotic regularity conditions. Thus,  $\mathbf{P}_{FA}(\Xi_t)$  is guaranteed to remain below  $\pi_0$ , if we choose

$$\zeta_{\pi_0} = Q_{\pi_0}$$

where  $Q_x$  denotes the inverse Gaussian  $Q$ -function at  $x$ , i.e.  $\int_{Q_x}^{\infty} (2\pi)^{-\frac{1}{2}} \exp(-u^2/2) du = x$ .

A close look at  $T_t$  in Eq. (13) reveals that  $T_t$  is a standardized linear combination of non-Gaussian components  $\left\{f(Y_s^{(t,w)}) : s = 1, \dots, p\right\}$ . Further, as discussed in Section 2.2 (see Eq. (11)), the random variables  $Y_{s_1}^{(t,w)}$  and  $Y_{s_2}^{(t,w)}$  are independent if there is no edge between  $s_1$  and  $s_2$  at time  $t$ , i.e.  $\Omega_{s_1, s_2}^{(t)} = 0$ . Therefore,  $T_t$  can be viewed as a linear combination of components in a sparse, stationary non-Gaussian random field. We expect the central limit theorem to hold for  $T_t$  (as  $p \rightarrow \infty$ ), since it is formed from a finite range random field. Lemma 2 in (Bolthausen, 1982), which provides a sufficient condition for asymptotic normality of a sequence of random variables, is a critical tool for obtaining the asymptotic distribution of  $T_t$ . Note that we adapt the proof of Theorem 3.3.1 in (Guyon, 1995) to the case of growing  $d_{\max}$  and  $\bar{d}$  for verifying the sufficient condition in Lemma 2 in (Bolthausen, 1982).

Figure 3 depicts the histogram of  $T_t$  and its kernel density estimate for different values of  $p, w$  and  $d_{\max}$  for  $10^4$  independent replicates. In each case,  $\Omega^{(t)}$  is constructed in the following way: We first generate a matrix  $U \in \mathbb{R}^{p \times p}$  with  $d_{\max}$  non-zero entries in each row. Non-zero elements of  $U$  are independently generated from a uniform distribution on  $(-1, 1)$ . We then choose the symmetric positive definite  $\Omega^{(t)}$  by

$$\begin{aligned}\Omega^{(t)} &\leftarrow U + U^\top + 1.5d_{\max}I_p, \\ \Omega^{(t)} &\leftarrow \frac{\Omega^{(t)}}{\lambda_{\min}(\Omega^{(t)})}.\end{aligned}$$

Thus, in each case  $\bar{d} = d_{\max}$  and  $\alpha_{\min} = 1$ . It is apparent from Figure 3 that  $T_t$  has approximately a standard Gaussian density for all four cases. It is also obvious that the skewness in the histograms decays as  $p$  increases.

**Remark 6** *The condition (b) in Assumption 3.1 restricts the growth rate of  $\bar{d}d_{\max}$ . For instance, it is asymptotically violated for star network topologies, i.e. a tree with a single hub node and  $p-1$  leaves of length one. One can easily verify that for star networks  $d_{\max} = p$  and  $\bar{d} = p^{-1}(2(p-1) + p) \approx 3$ . Therefore,*

$$\frac{d_{\max}\bar{d}}{\sqrt{p}} \asymp \sqrt{p} \rightarrow \infty.$$

*Such graphs are prevalent in centralized sensor or computer networks. Note that  $\bar{d} = \mathcal{O}(1)$  in many applied problems, even for star networks. So roughly speaking, Assumption 3.1 holds whenever  $d_{\max} = o(\sqrt{p})$ , as  $p \rightarrow \infty$ . Namely,  $\hat{\Omega}^{(t)}$  can have a finite number of local hub nodes, whose degrees grow slower than  $\sqrt{p}$ . Extensive numerical work presented in Section 5 aims to understand whether condition (b) in Assumption 3.1 is an artifact of our asymptotic analysis or reflects a shortcoming of the proposed algorithm.*

### 3.2. Distribution of $T_t$ under $\mathbb{H}_{1,t}$

Next, we study the power of  $\Xi_t$ , under the asymptotic regime of Assumption 3.1.

**Assumption 3.2** *The following conditions are imposed on the change-point location and post-change precision matrix.*

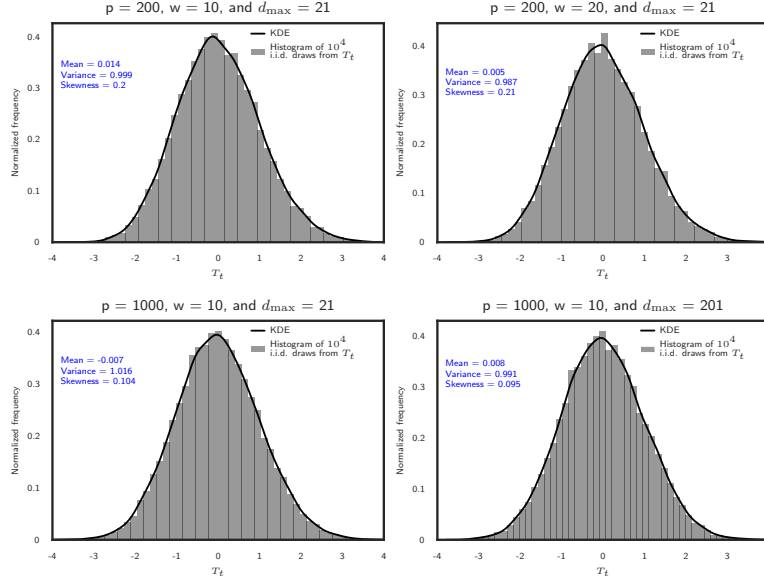


Figure 3: Histogram and kernel density estimator of  $T_t$  for different scenarios regarding  $(p, d_{\max}, w)$ . The condition number of  $\Omega^{(t)}$  equals to 2.05, 2.02, 2.08, and 1.24 from left to right and top to bottom, respectively.

- (a)  $\Omega^{(t)} \neq \Omega^{(t+1)}$  and  $\Omega^{(t+1)} = \dots = \Omega^{(t+w)}$ .  
 (b)  $\Omega^{(t+1)} \in \mathcal{C}_{++}^{p \times p}(\alpha_{\min}, M)$  for some  $M, \alpha_{\min} \in (0, \infty)$ .

For brevity, define  $\Delta = [\Delta_s : s = 1, \dots, p]^\top \in \mathbb{R}^p$  by

$$\Delta_s := \frac{\left[ \Omega^{(t)} (\Omega^{(t+1)})^{-1} \Omega^{(t)} \right]_{ss}}{\Omega_{ss}^{(t)}} - 1, \quad \forall s = 1, \dots, p, \quad (16)$$

with  $\Delta$  encoding the (relative) amount of sudden change in the network. For instance, if  $\Omega^{(t)} = \Omega^{(t+1)}$ , then all entries of  $\Delta$  are zero and  $\|\Delta\|_{\ell_2}$  is small if the dependence structure of GGM experiences a weak change at time  $t$ . Further, define  $\bar{\Psi}_p$  by

$$\bar{\Psi}_p := \frac{1}{p} \sum_{s=1}^p f(1 + \Delta_s).$$

Note that  $\bar{\Psi}_p$  is well-defined as  $\min_{s=1, \dots, p} \Delta_s > -1$ .  $\bar{\Psi}_p$  roughly quantifies the average relative change at time  $t$ . We now present the key result of this section.

**Theorem 7** *Suppose that Assumptions 3.1 and 3.2 hold. Further, assume that  $\pi_0, \pi_1 \leq \frac{1}{2}$ . Consider  $\Xi_t$  with the critical value  $\zeta_{\pi_0} = Q_{\pi_0}$ . Then,  $\Xi_t$  satisfies the following condition in the asymptotic regime*

$$P_{\text{FA}}(\Xi_t) \leq \pi_0, \quad \text{and} \quad P_{\text{MD}}(\Xi_t) \leq \pi_1,$$



as long as

$$\bar{\Psi}_p \geq \frac{4}{w} \sqrt{\frac{1}{p} \left[ \log \left( \frac{1}{2\pi_0} \right) + \log \left( \frac{1}{2\pi_1} \right) \right]}. \quad (17)$$

**Remark 8** *Theorem 7 confirms that detection of changes in the dependence structure of a GGM becomes relatively easier as  $p$  grows. The intuition behind Eq. (17) is that in large GGMs, borrowing information across different edges improves the detection power. Another intuitive aspect of the asymptotic result in Eq. (17) is the capacity of recognizing weaker change-points for larger delay  $w$ . Increasing  $w$  provides more information about the post-change dependence structure of the GGM. We also draw the reader's attention to the different asymptotic behavior of  $p$  and  $w$  in Eq. (17). Recall the formulation of  $\Xi_t$  from Eq. (14). We discussed in Remark 2 that the standard deviation of  $f(Y_s^{(t,w)})$ ,  $s = 1, \dots, p$ , decays at rate  $w^{-1}$  (instead of  $w^{-1/2}$ ), as  $w$  grows. Simply put, the proposed convex barrier function  $f$  plays a critical role in  $T_t$ , as it introduces a test statistic with very small variance under both the null and alternative hypotheses. Therefore, the detection capability of  $\Xi_t$  rapidly increases with  $w$ .*

**Remark 9** *In the offline setting with a single change point located at  $t^*$ , it has been shown in (Atchade and Bybee, 2017; Roy et al., 2016) that  $t^*$  can be estimated with order  $\log p$  accuracy, i.e.  $|\hat{t} - t^*| = \mathcal{O}(\log p)$ . In the online framework,  $w$  (detection delay) plays the role of the estimation error in the change-point location. So in practical scenarios, one can choose  $w = \mathcal{O}(\log p)$ . For this case the detection rate (17) can be rewritten as follows*

$$\bar{\Psi}_p \geq 4 \sqrt{\frac{-\log(2\pi_0) - \log(2\pi_1)}{p \log^2 p}}.$$

Next, we explore different settings in light of Theorem 7..

- (a) *Uniform change in  $\Omega^{(t)}$* : Assume that there is some  $\beta \geq 1$  such that  $-\Omega^{(t+1)} + \beta^{-1}\Omega^{(t)}$  is strictly positive definite. So the change-point affects all the eigenvalues of  $\Omega^{(t)}$ . Namely, the network is subject to a dense abrupt change in the spectral domain. In this case for all  $s \in \{1, \dots, p\}$ ,  $\Delta_s$  admits

$$1 + \Delta_s = \frac{\left[ \Omega^{(t)} (\Omega^{(t+1)})^{-1} \Omega^{(t)} \right]_{ss}}{\Omega_{ss}^{(t)}} \geq \frac{\left[ \Omega^{(t)} (\beta^{-1}\Omega^{(t)})^{-1} \Omega^{(t)} \right]_{ss}}{\Omega_{ss}^{(t)}} = \beta.$$

Hence  $\bar{\Psi}_p \geq f(\beta)$  and so the condition (17) holds, whenever

$$f(\beta) \geq 4 \sqrt{\frac{-\log(2\pi_0) - \log(2\pi_1)}{pw^2}}.$$

- (b) *Rank- $r$  change in  $\Omega^{(t)}$* : Let  $\Omega^{(t)} = U\Lambda U^\top$  be the eigen-decomposition of  $\Omega^{(t)}$ , i.e.,  $UU^\top = U^\top U = I_p$  and  $\Lambda$  is a diagonal matrix with  $\text{diag}(\Lambda) = [\lambda_1, \dots, \lambda_p]$ . For simplicity, also assume that  $\text{diag}(\Omega^{(t)}) = \mathbf{1}_p$ . We assume that only top  $r$  eigenvalues

of  $\Omega^{(t)}$  are impacted by the abrupt change at  $t$ . Particularly,  $\Omega^{(t+1)}$  can be decomposed in the following way.

$$\Omega^{(t+1)} = U \begin{bmatrix} \lambda_1 (1 + \beta_1) & & \\ & \ddots & \\ & & \lambda_p (1 + \beta_p) \end{bmatrix} U^\top \quad \text{s.t.} \quad \beta_{r+1} = \dots = \beta_p = 0.$$

We also assume that  $\min_{i=1,\dots,r} \beta_i \geq \beta_{\min}$  for some strictly positive  $\beta_{\min}$ . Roughly speaking  $\beta_{\min}$  can not be so large as  $\Omega^{(t+1)} \in \mathcal{C}_{++}^{p \times p}(\alpha_{\min}, M)$ . Due to the convexity of  $f$ ,  $\bar{\Psi}_p$  can be controlled from below by  $f(\sum_{s=1}^p (1 + \Delta_s)/p)$ . Lastly, we obtain a sharp upper bound on  $\sum_{s=1}^p (1 + \Delta_s)$ . Observe that

$$\begin{aligned} \frac{1}{p} \sum_{s=1}^p (1 + \Delta_s) &= \sum_{s=1}^p \frac{[\Omega^{(t)} (\Omega^{(t+1)})^{-1} \Omega^{(t)}]_{ss}}{p \Omega_{ss}^{(t)}} = \frac{1}{p} \text{tr} \left( \Omega^{(t)} (\Omega^{(t+1)})^{-1} \Omega^{(t)} \right) = \sum_{s=1}^p \frac{\lambda_s^2}{p \lambda_s (1 + \beta_s)} \\ &= \sum_{s=1}^p \frac{\lambda_s}{p (1 + \beta_s)} = \frac{1}{p} \sum_{s=1}^p \lambda_s - \frac{1}{p} \sum_{s=1}^r \frac{\lambda_s \beta_s}{1 + \beta_s} \stackrel{(a)}{=} 1 - \frac{1}{p} \sum_{s=1}^r \frac{\lambda_s \beta_s}{1 + \beta_s} \\ &\leq 1 - \frac{\beta_{\min}}{1 + \beta_{\min}} \frac{\sum_{s=1}^r \lambda_s}{p}. \end{aligned}$$

Notice that identity (a) is implied from the fact that all diagonal entries of  $\Omega^{(t)}$  equal to one. So,

$$\bar{\Psi}_p \geq f \left( \sum_{s=1}^p \frac{1 + \Delta_s}{p} \right) \geq f \left( 1 - \frac{\beta_{\min}}{1 + \beta_{\min}} \frac{\sum_{s=1}^r \lambda_s}{p} \right).$$

In summary, the condition (17) holds, as long as

$$f \left( 1 - \frac{\beta_{\min}}{1 + \beta_{\min}} \frac{\sum_{s=1}^r \lambda_s}{p} \right) \geq 4 \sqrt{\frac{-\log(2\pi_0) - \log(2\pi_1)}{pw^2}}.$$

It is noteworthy that for small enough  $\beta_{\min}$ ,  $\frac{\beta_{\min}}{1 + \beta_{\min}} \frac{\sum_{s=1}^r \lambda_s}{p}$  has the same asymptotic behavior as  $\frac{\beta_{\min} r}{p}$ .

- (c) *Small relative change*: In this case,  $\Omega^{(t+1)} = \Omega^{(t)} + \Theta$  for a positive semi-definite matrix  $\Theta$  satisfying

$$\left\| \left( \Omega^{(t)} \right)^{-\frac{1}{2}} \Theta \left( \Omega^{(t)} \right)^{-\frac{1}{2}} \right\|_{2 \rightarrow 2} \leq \xi < 1,$$

for a  $\xi \in [0, 1)$ . For any positive semi-definite  $A$  with  $\|A\|_{2 \rightarrow 2} \leq \xi$ , we know that

$$\left( I - (\xi + 1)^{-1} A \right) - (I_p + A)^{-1} \in \mathcal{S}_{++}^{p \times p}.$$

So we can control  $1 + \Delta_s$  from above,  $\forall s = 1, \dots, p$ . Set  $A = (\Omega^{(t)})^{-\frac{1}{2}} \Theta (\Omega^{(t)})^{-\frac{1}{2}}$ , so

$$\begin{aligned} 1 + \Delta_s &= \frac{\left[ \Omega^{(t)} (\Omega^{(t)} + \Theta)^{-1} \Omega^{(t)} \right]_{ss}}{\Omega_{ss}^{(t)}} = \frac{\left[ (\Omega^{(t)})^{\frac{1}{2}} (I_p + A)^{-1} (\Omega^{(t)})^{\frac{1}{2}} \right]_{ss}}{\Omega_{ss}^{(t)}} \\ &\leq 1 - \frac{\left[ (\Omega^{(t)})^{\frac{1}{2}} A (\Omega^{(t)})^{\frac{1}{2}} \right]_{ss}}{(1 + \xi) \Omega_{ss}^{(t)}} = 1 - \frac{\Theta_{ss}}{(1 + \xi) \Omega_{ss}^{(t)}} < 1. \end{aligned}$$

Since  $f$  is a decreasing, convex function in  $(0, 1]$ , we have

$$\bar{\Psi}_p = \frac{1}{p} \sum_{s=1}^p f(1 + \Delta_s) \geq f\left(\frac{\sum_{s=1}^p 1 + \Delta_s}{p}\right) \geq f\left(1 - \frac{1}{(1 + \xi)p} \sum_{s=1}^p \frac{\Theta_{ss}}{\Omega_{ss}^{(t)}}\right).$$

Therefore  $\Xi_t$  has the desirable properties, if

$$f\left(1 - \frac{1}{(1 + \xi)p} \sum_{s=1}^p \frac{\Theta_{ss}}{\Omega_{ss}^{(t)}}\right) \geq 4\sqrt{\frac{-\log(2\pi_0) - \log(2\pi_1)}{pw^2}}.$$

Lastly note that  $p^{-1} \sum_{s=1}^p \frac{\Theta_{ss}}{\Omega_{ss}^{(t)}}$  encodes the mean relative change in the conditional variance of all nodes.

- (d) *Localized change*: In this example, we consider a setting that the change-point only affects a single node and the edges connected to it. Let  $\Theta := \Omega^{(t+1)} - \Omega^{(t)}$  and suppose that  $\text{diag}(\Omega^{(t)}) = \mathbf{1}_p$ . The abrupt change being trapped in some  $s \in \{1, \dots, p\}$  and its neighbors means that

$$\Theta = v e_s^\top + e_s v^\top = [v \ e_s] \begin{bmatrix} e_s^\top \\ v^\top \end{bmatrix},$$

for  $s$ -th coordinate unit vector  $e_s \in \mathbb{R}^p$  and a  $v \in \mathbb{R}^p$  such that for any  $u \in \{1, \dots, p\}$ ,  $v_u = 0$  if  $\Omega_{su}^{(t)} = 0$ . For simplicity define  $\beta := \|v\|_{\ell_2}$ ,  $U := [v \ e_s] \in \mathbb{R}^{p \times 2}$ , and  $V := [e_s \ v] \in \mathbb{R}^{p \times 2}$ . Analogously to parts (a) – (c), our goal is to study the asymptotic behavior of  $\bar{\Psi}_p$  for growing  $p$ . We only consider the case that  $\beta = \mathcal{O}(1)$  when  $p \rightarrow \infty$ . Using *Woodbury's* matrix identity, we get the following equality for any  $u = 1, \dots, p$

$$\Delta_u = \frac{\left[ \Omega^{(t)} (\Omega^{(t)} + UV^\top)^{-1} \Omega^{(t)} \right]_{uu}}{\Omega_{uu}^{(t)}} - 1 = \left[ U \left( I_2 + V^\top (\Omega^{(t)})^{-1} U \right)^{-1} V^\top \right]_{uu}.$$

Observe that if  $\Omega_{su}^{(t)} = 0$ , then the  $u$ -th row of  $U$  and  $V$  are filled with zeros. So  $\Delta_u = 0$  and  $f(1 + \Delta_u) = 0$ . On the other hand for neighbors of  $s$ , we have

$$|\Delta_u| \leq \frac{\|U_{u,:}\|_{\ell_2} \|V_{u,:}\|_{\ell_2}}{\lambda_{\min}\left(I_2 + V^\top (\Omega^{(t)})^{-1} U\right)} = \frac{\|U_{u,:}\|_{\ell_2}^2}{\lambda_{\min}\left(I_2 + V^\top (\Omega^{(t)})^{-1} U\right)}$$

Since the smallest eigenvalue of  $\Omega^{(t+1)} = \Omega^{(t)} + UV^\top$  is greater than  $\alpha_{\min} > 0$  and its largest eigenvalue is less than  $M$ , one can find a bounded positive scalar  $C_{\min}$  depending on  $M$  and  $\alpha_{\min}$ , such that

$$\lambda_{\min} \left( I_2 + V^\top \left( \Omega^{(t)} \right)^{-1} U \right) \geq \sqrt{C_{\min}} \implies |\Delta_u| \leq \frac{\|U_{u,:}\|_{\ell_2}^2}{\sqrt{C_{\min}}}.$$

Thus, as  $p \rightarrow \infty$ , all  $\Delta_u$  terms remain in a small neighborhood of zero, since  $\|U\|_{\ell_2}^2 = 1 + \beta^2 = \mathcal{O}(1)$ . By applying a Second order *Taylor* expansion of  $f$  near one, we get

$$f(1 + \Delta_u) \leq C_{\max} \Delta_u^2, \quad \forall u = 1, \dots, p,$$

for some bounded positive scalar  $C_{\max}$ . Thus, some straightforward algebraic derivations lead to

$$\bar{\Psi}_p = \frac{1}{p} \sum_{u=1}^p f(1 + \Delta_u) \leq \frac{C_{\max}}{p} \sum_{u=1}^p \Delta_u^2 \leq \frac{C_{\max}}{p C_{\min}} \sum_{s=1}^p \|U_{u,:}\|_{\ell_2}^4 \asymp \frac{1}{p} \sum_{s=1}^p \|U_{u,:}\|_{\ell_2}^4 \leq \frac{\|U\|_{\ell_2}^4}{p} \asymp \frac{1}{p} \quad (18)$$

Comparing Eq. (18) with the sufficient detectability condition in Eq. (17) reveals that the considered localized change *can not be detected* with small error.

We conclude this section by relaxing condition (b) in Assumption 3.2. Choose  $w_+ \in \mathbb{N}$  strictly smaller than  $w$ . The following result introduces a sufficient condition on  $\bar{\Psi}_p$  for detecting the change point at time  $t_- := t - (w - w_+)$ . Notice that the next  $w$  samples at time  $t_-$  satisfy

$$\Omega^{(t_-)} = \dots = \Omega^{(t)} \neq \Omega^{(t+1)} = \dots = \Omega^{(t+w_+)}. \quad (19)$$

In words, we observe  $w_+$  post-change samples, which is equivalent to having a detection delay of  $w_+$ .

**Theorem 10** *Suppose that Assumption 3.1, constraint (a) in Assumption 3.2, and condition (19) hold. Choose  $\pi_0, \pi_1 \leq \frac{1}{2}$  and consider  $\Xi_{t_-}$  with the critical value  $\zeta_{\pi_0} = Q_{\pi_0}$ . Then, asymptotically*

$$\mathbf{P}_{\text{FA}}(\Xi_{t_-}) \leq \pi_0, \quad \text{and} \quad \mathbf{P}_{\text{MD}}(\Xi_{t_-}) \leq \pi_1,$$

whenever

$$\bar{\Psi}_p \geq \frac{4}{w_+} \sqrt{\frac{1}{p} \left[ \log \left( \frac{1}{2\pi_0} \right) + \log \left( \frac{1}{2\pi_1} \right) \right]}.$$

Due to space considerations, the proof of Theorem 10 is omitted since it follows along the lines of the proof of Theorem 7, with slightly more involved calculations.

#### 4. Asymptotic analysis of $\Xi_t$ : unknown $\Omega^{(t)}$

As discussed in Section 2.3, the pre-change precision matrix needs to be estimated from the observed samples. Our developed theoretical approach for studying  $\hat{\Xi}_t$  (and its associated test statistic  $\hat{T}_t$ ) relies on a large-sample sharp bound on the estimation error of the precision matrix. Specifically, we need operator and *Frobenius* norm convergence

rates in terms of  $p, d_{\max}$  of the network, and sample size. Because of space constraints, the non-asymptotic analysis of estimating sparse precision matrices is beyond the scope of this paper. Consequently, we only employ off-the-shelf theoretical results for this task. For instance, we heavily use Theorems 1 and 4 of (Cai et al., 2011) for studying Algorithm 2.3 with the CLIME estimator.

In Section 2.3, we qualitatively stated that separation of consecutive change-points is needed to guarantee the consistency of Algorithm 2.3. The following condition formalizes this notion.

**Assumption 4.1** *There exists a bounded large enough scalar  $C$  (depending on  $\alpha_{\min}$  and  $M$ ) such that*

$$|t_j^* - t_{j-1}^*| \geq N := Cp d_{\max} \log^2 p (w \vee \log p), \quad \forall j \geq 1.$$

Next, we find the null distribution of  $\hat{T}_t$  in Algorithm 2.3 (with the CLIME estimator).

**Theorem 11** *Suppose that there is no change-point between  $t-N$  and  $t+w$ , i.e.  $\Omega^{(t-N+1)} = \dots = \Omega^{(t)} = \dots = \Omega^{(t+w)}$ . Note that the condition is viable when Assumption 4.1 holds. Further, suppose that Assumption 3.1 holds for  $\Omega^{(t)}$ ; then, asymptotically*

$$\hat{T}_t \xrightarrow{d} \mathcal{N}(0, 1),$$

According to Theorem 11, if Assumption 4.1 holds, then the null distribution of  $\hat{T}_t$  has the same asymptotic behavior as that of  $T_t$  (with a fully-known pre-change precision matrix). Roughly speaking, the null distribution of  $T_t$  remains almost intact, given large enough separation of consecutive change-point locations. Hence, if the critical value  $\zeta_{\pi_0}$  is chosen by  $Q_{\pi_0}$ , then  $\mathbf{P}_{\text{FA}}(\hat{\Xi}_t)$  is ensured to be less than  $\pi_0$ , in the asymptotic regime. We conduct a simple simulation study for verifying Theorem 11. Figure 4 depicts the histogram and associated kernel density estimate of  $\hat{T}_t$  for four scenarios of  $(p, w, d_{\max})$ . In each panel, the histogram summarizes  $10^4$  independent replicates based on the same pre-change true and estimated precision matrices. For constructing each histogram, we use the CLIME procedure with  $N = \lceil pd_{\max} \log p \rceil$  for estimating the pre-change precision matrix. The *scikit-learn* implementation of the graphical lasso with the coordinate descent optimization algorithm is utilized for estimating  $\Omega^{(t)}$ . Finally, note that the regularization parameters are chosen as 0.01, 0.01, 0.01, and 0.005 from left to right and top to bottom panels, respectively.

**Remark 12** *The CLIME estimator requires  $\mathcal{O}(pd_{\max} \log p)$  samples for consistently estimating  $\Omega^{(t)}$ , in the Frobenius norm (see Theorem 4 of (Cai et al., 2011)), which is slightly weaker than the sufficient condition on the interval between two successive change-point in Assumption 4.1. Moreover, according to Theorem 8 of Atchadé et al. (Atchade and Bybee, 2017), estimating the location of change point with order  $\log p$  error in a  $s$ -sparse GGM requires  $\mathcal{O}(s \log p)$  samples. Setting  $w = \log p$  in Assumption 4.1 leads to a slightly stronger condition on  $N$  in the online framework. In particular, we require  $N = pd_{\max} \log^3 p$  observations before the change-point.*

**Remark 13** Notice that  $N$  is an increasing function of  $w$  in Assumption 4.1, which may seem counter-intuitive at first glance. This observation can be spelled out by noting that the  $\text{var}(T_t)$  decays as order  $w^{-2}$ , under the null hypothesis. Simply put as  $w$  increases, even a slight bias introduced by the CLIME estimator can change the null distribution of  $T_t$  (because of its very small variance). In contrast, Theorem 7 suggests that increasing  $w$  improves the power of our proposed test. Therefore, a suitable value for  $w$  is affected by the trade-off between the false alarm and miss-detection rates of  $\hat{\Xi}_t$ . Hence, we posit that selecting  $w = \mathcal{O}(\log p)$  is a reasonable choice in practical settings.

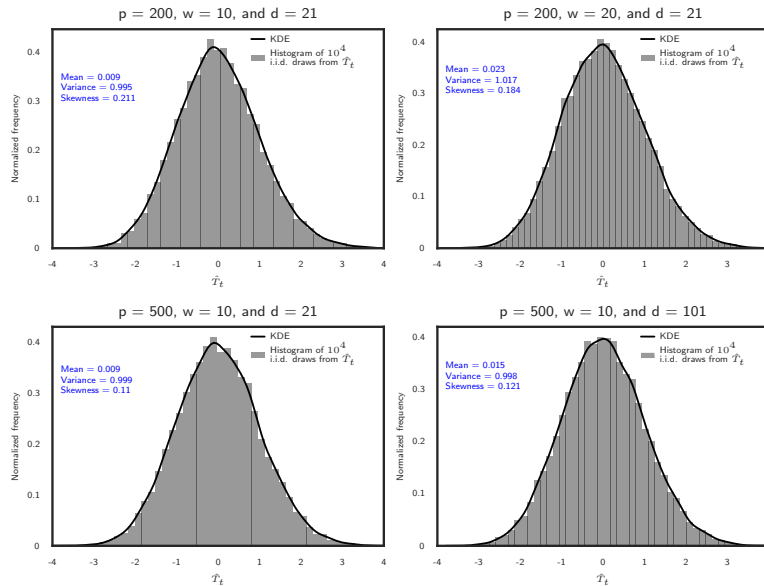


Figure 4: Histogram and kernel density estimator of  $\hat{T}_t$  for different scenarios regarding  $(p, d, w)$ . In each panel,  $\Omega^{(t)}$  is estimated by the graphical LASSO with  $N = \lceil pd_{\max} \log p \rceil$ . The regularization parameters are chosen as 0.01, 0.01, 0.01, and 0.005 from left to right and top to bottom, respectively.

Next, we study the detection rate of  $\hat{\Xi}_t$  under Assumption 4.1. The following result shows that under the same conditions of Theorem 7 and if Assumption 4.1 holds, then  $\hat{\Xi}_t$  enjoys the same detection power as  $\Xi_t$ .

**Theorem 14** Suppose that Assumptions 3.1 and 3.2 hold. Further, assume that  $\pi_0, \pi_1 \leq \frac{1}{2}$ . Consider  $\hat{\Xi}_t$  with the critical value  $\zeta_{\pi_0} = Q_{\pi_0}$ . Then,  $\hat{\Xi}_t$  satisfies the following condition asymptotically

$$\mathbf{P}_{\text{FA}}(\hat{\Xi}_t) \leq \pi_0, \quad \text{and} \quad \mathbf{P}_{\text{MD}}(\hat{\Xi}_t) \leq \pi_1,$$

provided that Assumption 4.1 is satisfied and

$$\bar{\Psi}_p \geq \frac{4}{w} \sqrt{\frac{1}{p} \left[ \log \left( \frac{1}{2\pi_0} \right) + \log \left( \frac{1}{2\pi_1} \right) \right]}.$$

In summary, given a good estimate of  $\Omega^{(t)}$ , for large networks the false alarm and miss-detection rates of the proposed test are not affected by the estimation error. Namely, our detection algorithm is robust against small estimation error, which is highly desirable in applications. It is worth mentioning that under Assumption 4.1, the detection power of  $\Xi_{t-}$  (recall its formulation from Eq. (19) and Theorem 10) is not also affected by the estimation error of the pre-change precision matrix.

## 5. Numerical studies

The next set of numerical experiments aims to:

- (a) Corroborate the developed asymptotic theory for  $\Xi_t$  and  $\hat{\Xi}_t$  in Sections 3 and 4.
- (b) Understand the impact of issues absent in the asymptotic analysis, such as the length of the burn-in period and update window, on  $\mathbf{P}_{\text{FA}}$  and  $\mathbf{P}_{\text{MD}}$  of Algorithm 2.3.
- (c) Assess the capabilities of our proposed method on real-world applications through sequential detection of structural changes in the S&P 500 over the period 2000 – 2016.

In Section 5.1, we evaluate power of  $\Xi_t$  (known pre-change precision matrix) for the three different scenarios introduced in Section 3.2: uniform, rank- $r$  and localized abrupt changes. Section 5.2 gauges the performance of  $\hat{\Xi}_t$  in Algorithm 2.3. Lastly, Section 5.3 is devoted to experiments with S&P 500 data. Throughout this section, the false alarm rate is set to  $\pi_0 = 0.01$  and thus the critical value is  $\zeta_{\pi_0} = 2.3263$ . Further,  $\{G_t \sim \mathcal{N}(\mathbf{0}_p, \Omega^{(t)}) : t = 1, \dots, T\}$  denotes a time-evolving zero-mean GGM of size  $p$ , observed on a time horizon of size  $T$ . The  $\Omega^{(t)}$  is always normalized to be a correlation matrix. In other words,  $\Omega^{(t)} = R^{(t)}$ , where  $R^{(t)}$  denotes the associated partial correlation matrix to  $\Omega^{(t)}$ .

### 5.1. Performance evaluation of $\Xi_t$

We set to  $T = 500$ , with a single change point at  $t^* = 250$ . We use  $\Omega_{\text{BC}}$  and  $\Omega_{\text{AC}}$  for referring the before and after-change precision matrix of  $G_t$ , respectively. Namely

$$\Omega^{(t)} = \Omega_{\text{BC}} \mathbf{1}_{\{t \leq t^*\}} + \Omega_{\text{AC}} \mathbf{1}_{\{t > t^*\}}, \quad \forall t \in \{1, \dots, T\}.$$

**Remark 15** *Given  $p$  and  $d$ , we generate a sparse random matrix  $U \in \mathbb{R}^{p \times p}$  with exactly  $d$  independent standard Gaussian entries per row. The locations of non-zero elements in each row are selected uniformly at random (without replacement). We generate  $H \in \mathbb{R}^{p \times p}$  via*

$$H := \left\| UU^\top \right\|_{\ell_\infty}^{-1} UU^\top$$

where  $H$  is a non-negative definite symmetric matrix whose entries lie in  $[-1, 1]$ . Note that when  $d \ll p$ , distinct rows of  $U$  have non-overlapping support with high probability. This fact ensures the sparsity of  $H$  in our synthetic simulation studies. In practice  $H$  is near singular when  $\frac{d}{p} \ll 1$ . So for controlling the condition number of  $\Omega_{\text{BC}}$ , we choose

$$\Omega_{\text{BC}} \longleftarrow H + \lambda_0 I_p,$$

for some properly chosen  $\lambda_0 > 0$ .

The first experiment assesses the performance of  $\Xi_t$  for detecting a uniform change across the entire network. In this setting, we select  $\lambda_0 = 0.1$  and  $\Omega_{AC} = (1 + \beta) \Omega_{BC}$ . Note that  $p\beta$  encodes the strength of change-point in the whole network. Our objective is to evaluate the sensitivity of false alarm and miss-detection to  $p, \beta, d$ , and delay  $w$ . For any fixed tuple  $(p, w, \beta, d)$ ,  $\pi_0$  and  $\pi_1$  are estimated based on 100 replicates of the network. In particular for the  $i$ -th replication, false alarm and miss-detection rates are respectively estimated by

$$\hat{\pi}_{0,i} = \frac{1}{t^* - w} \sum_{t=1}^{t^*-w} \Xi_t^i, \quad \text{and} \quad \hat{\pi}_{1,i} = \frac{1}{T - w - t^*} \sum_{t=t^*}^{T-w} (1 - \Xi_t^i), \quad (20)$$

where  $\{\Xi_t^i\}_{i=1}^T$  represents the sequence of binary decisions in  $\{1, \dots, T\}$ . The final estimates  $\hat{\pi}_0$  and  $\hat{\pi}_1$  are successively obtained by the sample mean of  $\{\hat{\pi}_{0,i}\}_{i=1}^{100}$  and  $\{\hat{\pi}_{1,i}\}_{i=1}^{100}$ .

Figure 5 depicts  $\log_{10} \hat{\pi}_0$  and  $\log_{10} \hat{\pi}_1$  as a function of  $p, w, \beta$ , and  $\bar{d}$ . It is apparent from Figure 5 that  $\hat{\pi}_0$  lies in a small neighborhood of  $\pi_0 = 0.01$  in all scenarios. However,  $\hat{\pi}_1$  rapidly decays with an increase in signal-to-noise-ratio (SNR),  $w$  and network size  $p$ , which is in accordance with Theorem 7. For instance, as can be seen in the left-bottom panel in Figure 5, when  $p = 800$  and  $w = 15$ , the average mis-detection probability is around 1%. In contrast, when the change point uniformly affects all nodes, for fixed values of  $\beta, p$ , and  $w$ , the detection power does not clearly depend on the sparsity structure of the network (which is encoded by  $\bar{d}$ ), which is again an expected finding based on part (a) of Remark 9.

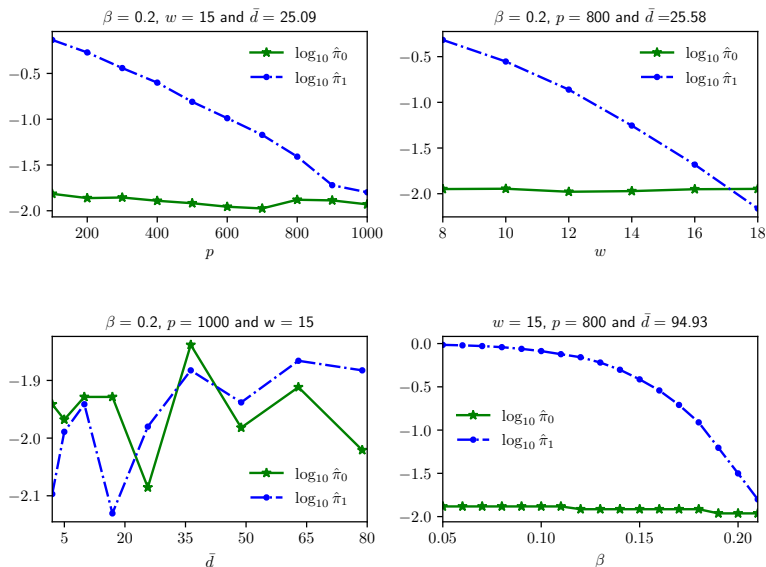


Figure 5:  $\log_{10} \hat{\pi}_0$  and  $\log_{10} \hat{\pi}_1$  for  $\Xi_t$  (fully known pre-change attributes) in a uniform change scenario. The four plots present  $\hat{\pi}_0$  and  $\hat{\pi}_1$  for different  $p, w, \bar{d}$ , and  $\beta$ , respectively from left to right and top to bottom.



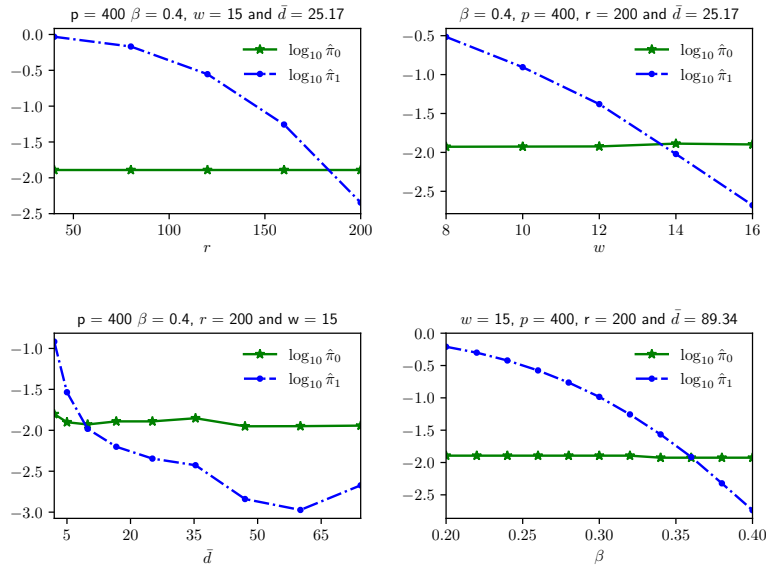


Figure 6:  $\log_{10} \hat{\pi}_0$  and  $\log_{10} \hat{\pi}_1$  for  $\Xi_t$  (fully known pre-change attributes) in low-rank change regime. The four plots present  $\hat{\pi}_0$  and  $\hat{\pi}_1$  to variations in  $r, w, \bar{d}$ , and  $\beta$ , respectively from left to right and top to bottom.

In the next experiment, we focus on a low-rank sudden change in the spectrum of  $\{G_t\}_{t=1}^T$ . Consider the spectral representation of  $\Omega_{BC}$  given by

$$\Omega_{BC} = \sum_{i=1}^p \lambda_{i,BC} v_i v_i^\top, \quad \text{where } \langle v_{j_1}, v_{j_2} \rangle = \mathbb{1}_{\{j_1=j_2\}}, \forall j_1, j_2 \in \{1, \dots, p\}.$$

Here we assume that  $\{\lambda_{i,BC} : i = 1, \dots, p\}$  has a non-increasing order. In this setting, the top  $r$  eigenvalues of  $\Omega_{BC}$  are affected by abrupt change, without any impact on the eigenvectors. Particularly, we choose  $\Omega_{AC}$  in the following way:

$$\Omega_{AC} = \sum_{i=1}^r \lambda_{i,BC} (1 + \beta) v_i v_i^\top + \sum_{i=r+1}^p \lambda_{i,BC} (1 + \beta) v_i v_i^\top.$$

The dependency of  $\hat{\pi}_0$  and  $\hat{\pi}_1$  (over 100 independent replicates) on  $\beta, r, w$ , and  $\bar{d}$  is presented in Figure 6. Analogous to the uniform change framework, the miss-detection rate rapidly decays with  $\beta, r$ , and  $w$ . For example, with window size  $w = 15$ ,  $\Xi_t$  is capable of detecting 20% change ( $\beta = 0.2$ ) in half of the eigenvalues with 99% percent accuracy ( $\hat{\pi}_1 \approx 0.01$ ). Further, based on the bottom-left panel in Figure 6, the proposed algorithm is more powerful for denser graphs (large average degree).

In the sequel, we appraise the detection performance of  $\Xi_t$  for uniform change in star graphs. Recall that the developed asymptotic theory in Section 3 can not be extended to star graphs, as  $d_{\max}$  grows faster than  $\sqrt{p}$ . For modeling star graphs, the following scheme is used for constructing  $\Omega_{BC}$ . Let  $u = [0, u_1, \dots, u_{p-1}]^\top \in \mathbb{R}^p$  be a standard Gaussian

vector padded with a zero in its first position and also define  $e_1 = [1, 0, \dots, 0]^\top \in \mathbb{R}^p$ . The pre-change precision matrix is given by

$$\Omega_{\text{BC}} = 1.1I_p + \frac{e_1 u^\top + u e_1^\top}{p \|u\|_{\ell_\infty}}.$$

Note that  $\Omega_{\text{BC}}$  encodes a star network whose root is set as the first node ( $s = 1$ ). Despite lack of asymptotic analysis, Figure 7 shows that the false alarm rate is still around 0.01, particularly for large graphs. Moreover, the variation of  $\hat{\pi}_1$  in terms of  $w, p$ , and  $\beta$  is analogous to that in preceding simulation studies.

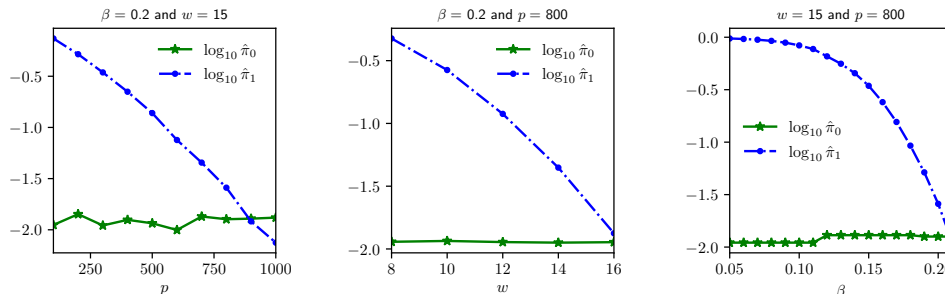


Figure 7:  $\log_{10} \hat{\pi}_0$  and  $\log_{10} \hat{\pi}_1$  for  $\Xi_t$  (fully known pre-change attributes) for uniform change in star graph. From left to right,  $\hat{\pi}_0$  and  $\hat{\pi}_1$  are plotted against  $p, w$ , and  $\beta$ , respectively.

## 5.2. Sensitivity of $\hat{\Xi}_t$ to the burn-in and update periods

Next, we numerically examine the sensitivity of  $\hat{\Xi}_t$  to three choice parameters: burn-in period, precision matrix update frequency, and the frequency of updating the regularization parameter (model selection). The burn-in period  $n_0$  refers to the number of samples used for computing an initial estimate of the post-change precision matrix (with the premise that the first change-point occurs at  $t = 0$ ). We assume that the distance between two consecutive change-points is greater than  $n_0$ . Since increasing  $n_0$  reduces the bias and variance of the estimated precision matrix, it can provide an effective barrier against random fluctuations in  $\hat{\Xi}_t$  time series and reduces the false alarm rate. On the other hand, an unnecessary increase in  $n_0$  can affect the applicability of the proposed detection algorithm in real world scenarios.

False alarms are costly as the algorithm needs to wait for a new burn-in window which in turn may lead to missing a true forthcoming sudden change. Strictly speaking, suppose that  $n_0 = 500$  and the algorithm raises a false alarm at  $t = 1000$ , while the true change-point is located at  $t = 1250$ . Thus, detecting this sudden change with a delay less than 250 is infeasible. One practical solution that guards against false alarms, which comes at the price of slightly increasing the detection delay, is to declare an abrupt change at time  $t$ , whenever

$$\hat{\Xi}_{t+r} \geq \zeta \pi_0, \quad \forall r = 0, \dots, \iota - 1.$$

In other words, we suggest to wait for  $\iota$  successive flags before entering a new burn-in period. Our synthetic simulation studies show that  $\iota = 5$  is a proper choice and so throughout this section we fix  $\iota = 5$ .

Another choice parameter, which has direct impact on the computational complexity of the online detection algorithm, is how often to update the estimated pre-change precision matrix. Let  $B$  denote a pre-specified block size of recent observations. After the burn-in period and until identifying a new change-point, the precision matrix is updated once every  $B$  new samples. In our simulation studies, we exploit the QUIC method (see Algorithm 1 in (Hsieh et al., 2014)) for the estimating pre-change precision matrix because of its smaller error and faster convergence comparing to the graphical lasso. The QUIC algorithm takes advantage of the second order approximation of the Gaussian log-likelihood for solving the following optimization problem.

$$\hat{\Omega}_\tau = \min_{\Omega \in S_{++}^{p \times p}} \left\{ -\log \det \Omega + \langle \Omega, \frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top \rangle + \tau \|\Omega\|_{\ell_1} \right\},$$

in which  $Z_i$  are i.i.d. draws from a zero mean Gaussian vector with precision matrix  $\Omega^*$ . Proper choice of  $\tau$  is essential for controlling the sparsity of the solution and avoiding over-fitting. The optimal value of  $\tau$  is chosen on a grid of size 20 to minimize the Bayesian information criterion (BIC) score. In particular, when  $\hat{t}_{last}$  stands for the last detected abrupt change before  $t$  and  $n := t - \hat{t}_{last}$ , then we optimize the BIC score over the following grid (for finding the best  $i \in \{0, \dots, 19\}$ ).

$$\mathcal{G} := \left\{ 10^{(-1+j/10)} \sqrt{\frac{\log p}{n}} : j = 0, \dots, 19 \right\}.$$

Namely, we assume that  $\tau = \tau_0 \sqrt{n^{-1} \log p}$  as  $n$  increases and the BIC procedure aims to choose the best  $\tau_0$ . Note that the relationship between  $\tau$  and  $n$  is justified by our asymptotic understanding of the QUIC algorithm. Evaluating the BIC score over  $\mathcal{G}$  is a heavy computational burden for large graphs. On the other hand, when  $B$  is considerably smaller than  $p$ , tuning the optimal  $j$  for each update cycle may provide a negligible improvement in the detection accuracy. So for accelerating the whole procedure, we propose to conduct a BIC model selection once every  $\kappa$  times of updating the precision matrix. Thus, we update the optimal value of  $j$  after getting  $\kappa B$  new samples, where  $\kappa^{-1}$  refers to the frequency of conducting BIC model selection.

We now have all the required ingredients for describing our numerical experiments. Our objective is to assess the sensitivity of Algorithm 2.3 to  $n_0, B$ , and  $\kappa$ . In all experiments, we choose  $p = 100, \pi_0 = 0.01$  and  $w = 20$ . Moreover three abrupt changes occur at  $\{3000, 6000, 9000\}$  in  $T = 10^4$  samples. For each choice of parameters  $(n_0, B, \kappa)$ , we employ Algorithm 2.3 on 50 replications of the non-stationary GGM generated from the following procedure. In order to have a fair comparison, all experiments are conducted on the same 50 replications. We denote by  $\Omega_0, \Omega_1, \Omega_2$ , and  $\Omega_3$  the precision matrix of the network in the following four periods  $\{1, \dots, 2999\}, \{3000, \dots, 5999\}, \{6000, \dots, 8999\}$ , and  $\{9000, \dots, 10000\}$ . In each replicate,  $\Omega_0$  is independently generated with the same procedure as in Remark 15 with  $p = 100, d = 20$  and  $\lambda_0 = 0.1$ . The first change point uniformly affects all nodes with  $\beta = 0.2$ , i.e.  $\Omega_1 = (1 + \beta) \Omega_0$ . The second abrupt change only impacts the top  $r = 50$  eigenvalues without changing the eigenvectors. Specifically,

$$\Omega_2 = \sum_{i=1}^r \lambda_i \beta v_i v_i^\top + \Omega_0, \quad \text{with } \beta = 0.4,$$

in which  $\{v_i \in \mathbb{R}^p\}_{i=1}^p$  are orthonormal eigenvectors of  $\Omega_1$  whose associated eigenvalues are sorted in a non-increasing order. For the last change point, we randomly generate a new precision matrix with the same distribution as  $\Omega_0$ . In this case, possibly all entries, eigenvalues, and eigenvectors are affected by the change-point. Roughly speaking, the sudden change signal is more visible compared to the previous ones. So for each replicate, we have access to a multivariate Gaussian time series of length  $T = 10^4$  of  $p = 100$  vertices.

The first simulation study aims to assess the sensitivity of Algorithm 2.3 to the burn-in period  $n_0$ . We set  $B = 50$  and  $\kappa = 4$  for all experiments, while varying  $n_0$  in the set  $\{1100, 1300, 1500, 1700, 1900, 2100\}$ . Note that in all the following experiments,  $\hat{T}_t$  is filled with *NA* if  $t$  is in the burn-in period. Table 1 summarizes the median and interquartile range (IQR) of the detection delays for the three aforementioned type of change points, as well as average number of false alarms. According to Table 1 given enough samples in the burn-in period for estimating pre-change precision matrix, increasing  $n_0$  has negligible impact on the detection delay. However, larger values of  $n_0$  can guard  $\hat{T}_t$  against false alarms. Moreover Figure 8 exhibits the average sample path of  $\hat{T}_t$  time series (over 50 experiments and after skipping the missing values) for each choice of  $n_0$ . Figure 8 shows that the change points randomly affecting all the nodes are easier to detect than uniform or low rank breaks, which corroborates the summary results in Table 1.

		$n_0 = 1100$	$n_0 = 1300$	$n_0 = 1500$	$n_0 = 1700$	$n_0 = 1900$	$n_0 = 2100$
Uniform Change	Median delay	53.5	51	54	54	52.5	51
	IQR of delay	48.75	48.75	52.25	51.25	48.75	48.75
Low rank Change	Median delay	33	32.5	32	32.5	33	33
	IQR of delay	13.25	14.5	12	11.75	13.25	13.25
Random Change	Median delay	4	4	4	4	4	4
	IQR of delay	0.75	0.75	0.75	0.75	0.75	0.75
Average number of false alarms		0.12	0.06	0.08	0.04	0.00	0.00

Table 1: The median and IQR of detection delay, and average number of false alarms for different values of  $n_0$  over 50 independent experiments.

In the second experiment, we evaluate the robustness of false alarm and miss detection rates to variations of  $\kappa$ . Recall that for a fixed mini-batch size  $B$  and given  $n$  samples in the pre-change regime, the precision matrix is estimated using the QUIC algorithm with regularization parameter  $\tau_0 \sqrt{n^{-1} \log p}$ , where  $\tau_0$  is updated after getting  $\kappa B$  new samples. We fix  $B = 50$ , and choose  $(n_0, \kappa)$  pair from  $\{1500, 2000\} \times \{1, 2, 3, 4\}$  (8 different cases). The median and IQR of detection delay, as well as average number of false alarms (computed from 50 independent experiment) are given in Table 2 for each  $(n_0, \kappa)$ . It can be seen that variations in  $\kappa$  barely affects the detection delay and false alarm rate.

		(1500, 1)	(1500, 2)	(1500, 3)	(1500, 4)	(2000, 1)	(2000, 2)	(2000, 3)	(2000, 4)
Uniform Change	Median delay	54	54	54	54	52.5	52.5	52.5	52.5
	IQR of delay	52.25	52.25	52.5	52.25	48	48	48.5	48
Low rank Change	Median delay	33	33	33	32	33	33	32	31
	IQR of delay	13.5	13.25	13.25	12	13.5	13.25	13.25	12
Random Change	Median delay	4	4	4	4	4	4	4	4
	IQR of delay	0	0.75	0.75	0.75	0.75	1	1	1
Average number of false alarms		0.08	0.08	0.08	0.08	0.00	0.00	0.00	0.00

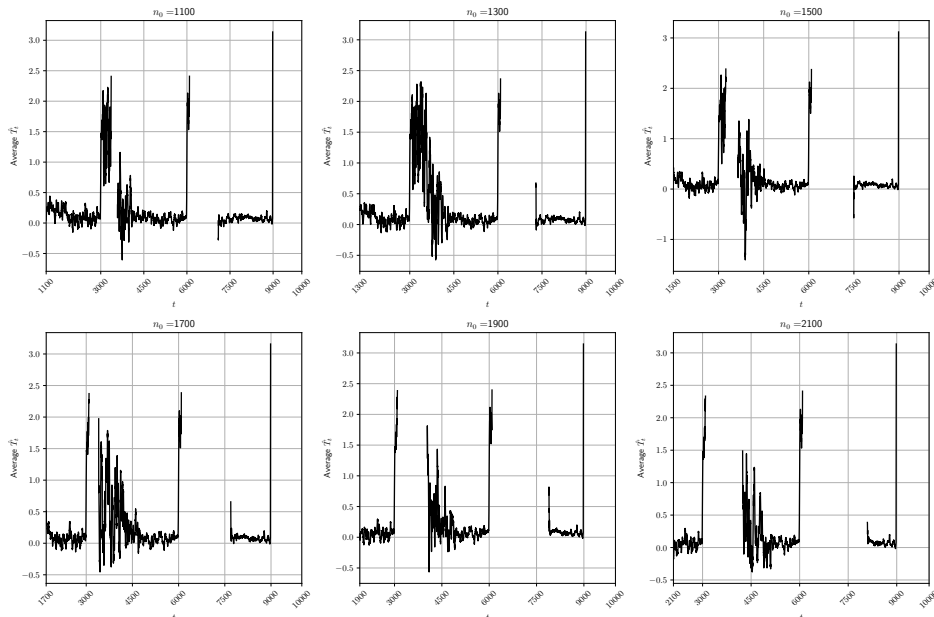

 Figure 8: Average sample path of  $\hat{T}_t$  for different values of  $n_0$ .

 Table 2: The median and IQR of detection delay, and average number of false alarms for different values of  $(n_0, \kappa)$  over 50 independent experiments.

Lastly, we study the role of increasing the min-batch size  $B$  on the performance of  $\hat{T}_t$ . Note that we use the same data set as in the two previous simulation studies. We fix  $\kappa = 4$  for all the experiments and consider 8 different scenarios for  $(n_0, B)$  in  $\{1100, 1500\} \times \{5, 10, 20, 40\}$ . Table 3 summarizes the median and IQR of detection delay, and average number of false alarms. Similar to Table 2, examining the columns of Table 3 shows that if provided with an adequate quality initial estimate of the precision matrix in the burn-in period, increasing  $B$  exhibits a small impact on the detection performance of  $\hat{T}_t$ . Roughly speaking, when  $B$  is small, the estimated precision matrix (and consequently the detection procedure) slightly changes after an update.

		(1100, 5)	(1100, 10)	(1100, 20)	(1100, 40)	(1500, 5)	(1500, 10)	(1500, 20)	(1500, 40)
Uniform Change	Median delay	54.5	54.5	53.5	52.5	54.5	54.5	54.5	54.5
	IQR of delay	50.5	50.5	50.75	48.75	52.75	52.75	52.25	52.25
Low rank Change	Median delay	33	33	32.5	32.5	33	33	32.5	32.5
	IQR of delay	13.5	16.75	16	13.25	16.75	16.75	16	13.25
Random Change	Median delay	4	4	4	4	4	4	4	4
	IQR of delay	0	0	0	0.75	0.75	0	0	0.75
Average number of false alarms		0.12	0.12	0.12	0.14	0.08	0.08	0.08	0.08

 Table 3: The median and IQR of detection delay, and average number of false alarms for different values of  $(n_0, B)$  over 50 independent experiments.

### 5.3. Real data experiment

We assess the performance of  $\hat{\Xi}_t$  in a real-world scenario. The objective is to estimate the time location of abrupt changes in the dependency structure of S&P 500 daily close price data for the period from 2000-01-01 to 2016-03-03 (total of 3814 trading days). Note that the S&P 500 data do not fit into the formulation of time evolving GGMs with independent observations for two main reasons. First, the list of securities in S&P 500 pool evolves over time. For example *Alphabet Inc. Class C* (with ticker symbol *GOOG*) entered the list on 2006-04-03. For circumventing the first issue, we follow the cleaning procedure from Atchade and Bybee (2017) and select a fixed list of 436 ticker symbols from 2004-02-06 to 2016-03-03, consisting of 3039 trading days. The second technical challenge is that the daily close price of each ticker typically exhibits strong temporal dependence with non-Gaussian marginal distribution. The geometric Brownian motion (GBM) is a versatile and popular tool for modeling daily stock prices in mathematical finance (see e.g., Chapter 10 of Ross (2014)). Note that under the GBM model, the daily log-return time series can be well approximated by a Gaussian random walk. In other words, a high-dimensional GGM with 436 vertices provides a good working model for the daily log-returns of S&P 500 components.

We choose  $n_0 = 200$ ,  $w = 22$  (corresponding to the number of trading days in a month),  $\pi_0 = 0.05$ ,  $\kappa = 2$ , and  $B = 10$ . We also employ the same estimation and model selection approach as in Section 5.2. For each experiment, we conduct our sequential detection algorithm on a network of 100 randomly chosen tickers, i.e.,  $p = 100$ . We also set  $\hat{\Xi}_t = 1$  for  $t$  in the burn-in period. The solid black curve and cyan interval in Figure 9, respectively exhibits the sample average and standard deviation of  $\hat{\Xi}_t$  over 300 independent experiments. Note that for each  $t$ , the proximity of the average  $\hat{\Xi}_t$  (over 300 experiments) to one implies that  $t$  is close to a sudden change in most of the experiments. So roughly speaking the solid dark curve captures the strength of nearest change-point.

It becomes apparent from Figure 9 that the first cluster of change-points starts in the last few months of 2008 and lasts for one and half years, which successfully captures the 17-month bear market period from October 2007 to March 2009 (when the S&P 500 index lost approximately 50% of its value). The beginning of the *Great Recession* in December 2007 and the bankruptcy filing of *Lehman Brothers Holdings Inc.* on September 15, 2008 are two notable events during this bear market. Another visible change-point in Figure 9 takes place in late 2010, which can represent the end of bear market in 2009. Our detection algorithm finds the third change-point starting around September 2011, in most of the 300 experiments. The new regime likely corresponds to a sharp drop in stock prices during August 2011 affecting the performance of stock exchanges across the United States, Europe, East Asia, and the Middle East. The fall of stock market in August 2011 was due to fears of contagion of the European sovereign debt crisis to Spain and Italy. The proposed algorithm spots the last change-point in Fall 2014. Note that from September 18 to October 14 of 2014, the S&P 500 experienced a remarkable decline of 7.4% due to market jitters over the rapid spread of the *Ebola* virus beyond Africa, fears of a global economic slowdown, Hong Kong protests (*umbrella revolution*), and the first U.S. strikes against ISIS in Syria.

In the sequel, we introduce a tangible interpretation of the change-points identified in terms of the average volatility of all tickers (nodes in the S&P 500 graphical model). For each node  $s$ , let  $\{X_{t,s}\}_{t=1}^T$  stand for the daily log-return of  $s$ . We approximate the volatility

of  $\{X_{t,s}\}_{t=1}^T$  using

$$V_{t,s} = \text{std}(X_{s,t}, \dots, X_{s,t+w}), \forall t = 1, \dots, T - w.$$

The dashed blue line in Figure 9 represents the sample average of  $\{V_{t,s}\}_{s=1}^p$ , which is a proxy for the return volatility index in the S&P 500 pool. Note that based on Figure 9, all detected change-points are closely related to periods of relatively high values in the return volatility index.

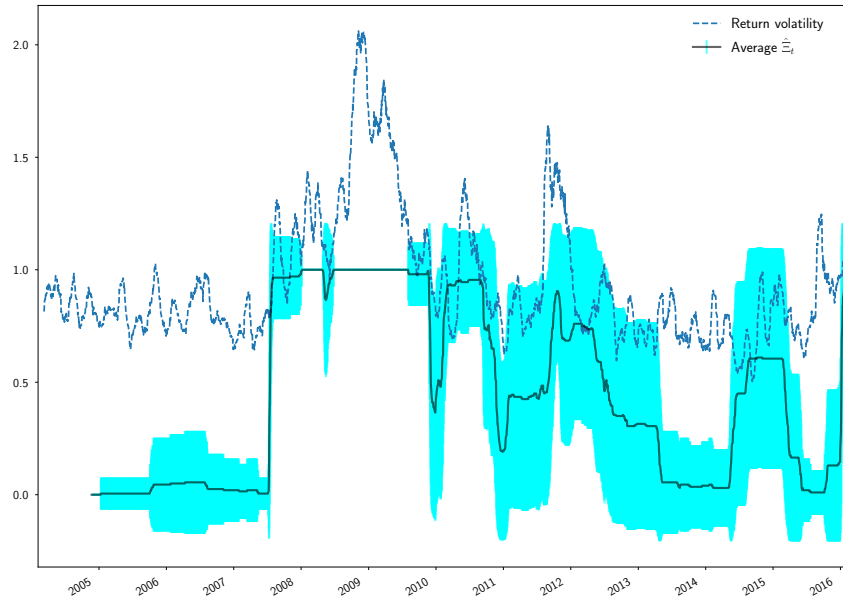


Figure 9: Confidence bound around  $\hat{\Xi}_t$  for 300 random subsets of S&P 500 with 100 stocks, and average return volatility.

## 6. Future directions

We studied the problem of scalable sequential detection of abrupt changes in the precision matrix of high-dimensional sparse GGMs. The objective is to detect a regime-change with maximum delay of  $w$ , while keeping the false alarm rate below some pre-specified threshold  $\delta$ . The proposed test  $\hat{T}_t$  (in Algorithm 2.3) uses a convex barrier function, which is motivated by the formulation of KL-divergence between two zero mean multivariate Gaussian vectors, for comparing the conditional log-likelihood of all nodes before and after a sudden-change. We also established asymptotic performance guarantees of the algorithm under certain regularity conditions on  $p$ ,  $w$ , sample size, and pre- and post-change structure of the graph.

Despite recent progress on scalable offline detection of change-points in the dependence structure of graphical models in the “*large p, small n*” framework, there is still a lot of work to be undertaken in the online setting and this study constitutes a first step in that direction. We conclude this section by addressing several potential extensions of Algorithm 2.3 as topics of future investigation.

- (a) Extending the applicability of  $\hat{T}_t$  to non-Gaussian random fields having a closed-form conditional log-likelihood, such as *Ising* models, is straightforward. However the asymptotic analysis of the associated  $\hat{T}_t$  with Ising models is more technically challenging, since  $Y_s^{(t,w)} : s = 1, \dots, p$  (in Eq. (6)) is not a set of  $\chi_w^2$  random variables under the null hypothesis.
- (b) The pre-change precision matrix can be sequentially updated after receiving a new observation by *regularized dual averaging (RDA)* procedure Xiao (2010) instead of using a mini-batch scheme in Algorithm 2.3. To the best of our knowledge, the consistency of RDA is formulated in terms of the regret function, unlike  $\ell_2$  or operator norm consistency of the CLIME or QUIC mini-batch updates. It is worth mentioning that an RDA-based sequential change-point detection was beyond the scope of this paper, as our core asymptotic focus was on online detection, rather than developing  $\ell_2$  or operator norm estimation rate of the online update for the pre-change precision matrix.
- (c) From a statistical standpoint, investigating optimal choices for the barrier function  $f$ , as well as the minimax lower bound on separability between  $\mathbb{H}_{0,t}$  and  $\mathbb{H}_{1,t}$  constitute interesting subjects of future research.
- (d) We pointed out in part (d) of Remark 9 that  $\hat{T}_t$  is not capable of detecting regime-shifts affecting a small sub-graph of the network. When a sudden-change is confined to a subset of nodes  $B \subset \{1, \dots, p\}$ ,  $f(Y^{(t,w)}_s)$  concentrates around zero for all nodes  $s$  not directly connected to  $B$ , i.e.  $\Omega_{su}^{(t)} = 0, \forall u \in B$ . In other words, the SNR is reduced in  $\hat{T}_t$  by considering all nodes unaffected by the change-point. In this case scanning through a set of potential candidates for  $B$  can improve the detection power. Strictly speaking if  $B$  is known to belong to a class of clusters  $\mathcal{C}$ , then this idea can be formulated as the following.

$$\hat{T}'_t = \max_{B \in \mathcal{C}} \frac{\sum_{s \in B} \left[ f\left(\hat{Y}_s^{(t,w)}\right) - g_1(w) \right]}{g_2(w) \sqrt{\sum_{s_1, s_2 \in B}^p h_w\left(\hat{R}_{s_1, s_2}^{(t)}\right)}}$$

Obtaining the distribution of  $\hat{T}'_t$  under the null and the alternative hypotheses is a challenging analytic task and relies on extreme value theory for dependent random variables, which is beyond the scope of this manuscript.

## 7. Proofs

**Proof** (Proof of Theorem 5) Recall from Eq. (6) and Eq. (13) that

$$Y_s^{(t,w)} := \frac{1}{w\Omega_{ss}^{(t)}} \sum_{r=1}^w \langle \mathbf{X}_{t+r}, \Omega_{:,s}^{(t)} \rangle^2,$$

$$T_t = \frac{\sum_{s=1}^p \left[ f\left(Y_s^{(t,w)}\right) - g_1(w) \right]}{g_2(w) \sqrt{\sum_{s_1, s_2=1}^p h_w\left(R_{s_1, s_2}^{(t)}\right)}}$$



The proof leverages the well-known Stein's result (see e.g. Stein (1972) or Lemma 2 in Bolthausen (1982)) saying that a sequence of  $\mathbb{L}^2$  integrable random variables  $\eta_n$  converges in distribution to a standard normal random variable as  $n \rightarrow \infty$ , if

$$\mathbb{E} [(j\nu - \eta_n) e^{j\nu\eta_n}] \rightarrow 0, \quad \forall \nu \in \mathbb{R}. \quad (21)$$

So we just need to show that  $\mathbb{E} (j\nu - T_t) e^{j\nu T_t}$  converges to zero (as  $p \rightarrow \infty$ ), for any  $\nu \in \mathbb{R}$ . We use a similar technique as in the proof of Theorem 3.3.1 in Guyon (1995). Note that we cannot directly utilize the result in Guyon (1995), as both  $\bar{d}$  and  $d_{\max}$  can grow with  $p$ . We only consider the case of  $w = 1$  and the proof can be extended to any  $w > 1$ . However, we still keep  $w$  as superscript for notational consistency. We introduce new notation for compactness and transparency of algebraic derivations. For any  $s \in \{1, \dots, p\}$ , define

$$\begin{aligned} \bar{f} \left( Y_s^{(t,w)} \right) &:= f \left( Y_s^{(t,w)} \right) - g_1(w), \quad \sigma_{p,w}^2 := g_2^2(w) \sum_{s_1, s_2=1}^p h_w \left( R_{s_1, s_2}^{(t)} \right), \\ \xi_s &:= \sum_{l=1}^p \bar{f} \left( Y_l^{(t,w)} \right) \mathbb{1} (R_{s,l} \neq 0), \quad \xi_s^c := \sum_{l=1}^p \bar{f} \left( Y_l^{(t,w)} \right) \mathbb{1} (R_{s,l} = 0). \end{aligned} \quad (22)$$

Here  $R$  stands for the partial correlation matrix of the network. Notice that with the new notation, we have

$$T_t = \frac{1}{\sigma_{p,w}} \sum_{s=1}^p \bar{f} \left( Y_s^{(t,w)} \right), \quad \text{and} \quad \xi_u + \xi_u^c = \sum_{u=1}^p \bar{f} \left( Y_u^{(t,w)} \right) = \sigma_{p,w} T_t, \quad \forall u = 1, \dots, p. \quad (23)$$

Since  $R$  is a correlation matrix, then

$$\sigma_{p,w}^2 \geq g_2^2(w) \sum_{s=1}^p h_w(R_{ss}) = g_2^2(w) \sum_{s=1}^p h_w(1) = g_2^2(w) p \Rightarrow \sigma_{p,w} \geq g_2(w) \sqrt{p}. \quad (24)$$

Our idea is to decompose  $(j\nu - T_t) e^{j\nu T_t}$  into 3 terms with diminishing expected value. For any  $\nu \in \mathbb{R}$ , define

$$\begin{aligned} A_1 &:= \frac{1}{\sigma_{p,w}} e^{j\nu T_t} \sum_{s=1}^p \bar{f} \left( Y_s^{(t,w)} \right) \left[ 1 - \frac{j\nu}{\sigma_{p,w}} \xi_s - \exp \left( -\frac{j\nu \xi_s}{\sigma_{p,w}} \right) \right], \\ A_2 &:= j\nu e^{j\nu T_t} \left[ 1 - \sigma_{p,w}^{-2} \sum_{s,l=1}^p \bar{f} \left( Y_s^{(t,w)} \right) \bar{f} \left( Y_l^{(t,w)} \right) \mathbb{1} (R_{s,l} \neq 0) \right] = j\nu e^{j\nu T_t} \left[ 1 - \sigma_{p,w}^{-2} \sum_{s=1}^p \bar{f} \left( Y_s^{(t,w)} \right) \xi_s \right], \\ A_3 &:= \sigma_{p,w}^{-1} \sum_{s=1}^p \left( \bar{f} \left( Y_s^{(t,w)} \right) e^{j\nu \xi_s^c / \sigma_{p,w}} \right). \end{aligned}$$

We obtain a new formulation for  $A_1 + A_3$  using Eq. (23), which shows that  $(j\nu - T_t) e^{j\nu T_t} = A_2 - (A_1 + A_3)$ .

$$\begin{aligned}
 A_1 + A_3 &= \frac{e^{j\nu T_t}}{\sigma_{p,w}} \sum_{s=1}^p \bar{f}(Y_s^{(t,w)}) - \frac{j\nu e^{j\nu T_t}}{\sigma_{p,w}^2} \sum_{s=1}^p \bar{f}(Y_s^{(t,w)}) \xi_s - \frac{1}{\sigma_{p,w}} \sum_{s=1}^p \bar{f}(Y_s^{(t,w)}) e^{j\nu(T_t - \frac{\xi_s}{\sigma_{p,w}})} + A_3 \\
 &= e^{j\nu T_t} T_t - \frac{j\nu e^{j\nu T_t}}{\sigma_{p,w}^2} \sum_{s=1}^p \bar{f}(Y_s^{(t,w)}) \xi_s - \frac{1}{\sigma_{p,w}} \sum_{s=1}^p \bar{f}(Y_s^{(t,w)}) e^{j\nu \xi_s / \sigma_{p,w}} + A_3 \\
 &= e^{j\nu T_t} T_t - \frac{j\nu e^{j\nu T_t}}{\sigma_{p,w}^2} \sum_{s=1}^p \bar{f}(Y_s^{(t,w)}) \xi_s - A_3 + A_3 = e^{j\nu T_t} T_t - \frac{j\nu e^{j\nu T_t}}{\sigma_{p,w}^2} \sum_{s=1}^p \bar{f}(Y_s^{(t,w)}) \xi_s.
 \end{aligned}$$

We first show that  $\mathbb{E}A_3 = 0$ . There is a zero mean Gaussian random vector  $Z \in \mathbb{R}^p$  such that

$$Y_s^{(t,w)} = Z_s^2, \quad \forall s \in \{1, \dots, p\}, \quad \text{and} \quad \text{cov}(Z) = R.$$

So for any  $s$ ,  $Y_s^{(t,w)}$  is independent of  $\{Y_v^{(t,w)} : R_{vs} = 0\}$ , implying the independence of  $\bar{f}(Y_s^{(t,w)})$  and  $\xi_s^c$ . Therefore,

$$\mathbb{E}A_3 = \sigma_{p,w}^{-1} \sum_{s=1}^p \mathbb{E} \left( \bar{f}(Y_s^{(t,w)}) e^{j\nu \xi_s^c / \sigma_{p,w}} \right) = \sigma_{p,w}^{-1} \sum_{s=1}^p \text{cov} \left( \bar{f}(Y_s^{(t,w)}), e^{j\nu \xi_s^c / \sigma_{p,w}} \right) = 0.$$

Thus  $\mathbb{E}[(j\nu - T_t) e^{j\nu T_t}] = \mathbb{E}A_2 - \mathbb{E}A_1$ . Next, we show that  $\mathbb{E}|A_1|$  and  $\mathbb{E}|A_2|$  tend to zero. We first deal with  $\mathbb{E}|A_2|$ . Due to the Cauchy-Schwartz inequality, we only show that  $\|A_2\|_2$  goes to zero. It is known from Eq. 24 that  $\sigma_{p,w} \geq g_2(w) \sqrt{p}$ . A lower bound on  $\sigma_{p,w}$  gives an alternative asymptotic representation for  $\|A_2\|_2$ .

$$\begin{aligned}
 \|A_2\|_2^2 &= \nu^2 \mathbb{E} \left[ \left| \sum_{s,l=1}^p \frac{\bar{f}(Y_s^{(t,w)}) \bar{f}(Y_l^{(t,w)})}{\sigma_{p,w}^2} \mathbb{1}_{\{R_{s,l} \neq 0\}} - 1 \right|^2 \right] \\
 &= \frac{\nu^2}{\sigma_{p,w}^4} \text{var} \left[ \sum_{s,l=1}^p \bar{f}(Y_s^{(t,w)}) \bar{f}(Y_l^{(t,w)}) \mathbb{1}_{\{R_{s,l} \neq 0\}} \right] \\
 &\leq \frac{\nu^2}{p^2 g_2^4(w)} \sum_{R_{s_1, l_1} \neq 0} \sum_{R_{s_2, l_2} \neq 0} \text{cov} \left[ \bar{f}(Y_{s_1}^{(t,w)}) \bar{f}(Y_{l_1}^{(t,w)}), \bar{f}(Y_{s_2}^{(t,w)}) \bar{f}(Y_{l_2}^{(t,w)}) \right].
 \end{aligned} \tag{25}$$

We break the summation in Eq. (25) into two parts. We interchangeably use  $B_1$  and  $B_2$  instead of  $(s_1, l_1)$  and  $(s_2, l_2)$ , respectively. Define

$$\begin{aligned}
 \mathcal{C}_1 &:= \{B_1, B_2 : R_{s_1, l_1}, R_{s_2, l_2} \neq 0, \text{ There is no edge between } B_1 \text{ and } B_2\}, \\
 \mathcal{C}_2 &:= \{B_1, B_2 : R_{s_1, l_1}, R_{s_2, l_2} \neq 0, \text{ There is an edge between } B_1 \text{ and } B_2\}.
 \end{aligned}$$

Proposition 23 states that

$$\sum_{B_1, B_2 \in \mathcal{C}_1} \text{cov} \left[ \bar{f}(Y_{s_1}^{(t,w)}) \bar{f}(Y_{l_1}^{(t,w)}), \bar{f}(Y_{s_2}^{(t,w)}) \bar{f}(Y_{l_2}^{(t,w)}) \right] = 0.$$

So we can simplify the upper bound on  $\|A_2\|_2$  in Eq. (25) as follows.

$$\begin{aligned} \|A_2\|_2^2 &\leq \left(\frac{\nu}{pg_2^2(w)}\right)^2 \sum_{B_1, B_2 \in \mathcal{C}_2} \text{cov} \left[ \bar{f}(Y_{s_1}^{(t,w)}) \bar{f}(Y_{l_1}^{(t,w)}), \bar{f}(Y_{s_2}^{(t,w)}) \bar{f}(Y_{l_2}^{(t,w)}) \right] \\ &\leq \left(\frac{\nu}{pg_2^2(w)}\right)^2 |\mathcal{C}_2| \max_{B_1, B_2 \in \mathcal{C}_2} \left| \text{cov} \left[ \bar{f}(Y_{s_1}^{(t,w)}) \bar{f}(Y_{l_1}^{(t,w)}), \bar{f}(Y_{s_2}^{(t,w)}) \bar{f}(Y_{l_2}^{(t,w)}) \right] \right|. \end{aligned} \quad (26)$$

We discussed in Section that  $wg_2(w)$  is uniformly bounded from below on  $\mathbb{N}$ . Namely,  $\inf_{w \in \mathbb{N}} |wg_2(w)| > 0$ . So Eq. (26) can be simplified as the following.

$$\|A_2\|_2^2 \leq \left(\frac{\nu}{p}\right)^2 |\mathcal{C}_2| \max_{B_1, B_2 \in \mathcal{C}_2} w^4 \left| \text{cov} \left[ \bar{f}(Y_{s_1}^{(t,w)}) \bar{f}(Y_{l_1}^{(t,w)}), \bar{f}(Y_{s_2}^{(t,w)}) \bar{f}(Y_{l_2}^{(t,w)}) \right] \right|. \quad (27)$$

We now control the covariance between  $\bar{f}(Y_{s_1}^{(t,w)}) \bar{f}(Y_{l_1}^{(t,w)})$ , and  $\bar{f}(Y_{s_2}^{(t,w)}) \bar{f}(Y_{l_2}^{(t,w)})$  from above. Choose  $B_1, B_2 \in \mathcal{C}_2$  arbitrarily. Observe that

$$\begin{aligned} &\left| \text{cov} \left[ \bar{f}(Y_{s_1}^{(t,w)}) \bar{f}(Y_{l_1}^{(t,w)}), \bar{f}(Y_{s_2}^{(t,w)}) \bar{f}(Y_{l_2}^{(t,w)}) \right] \right| \leq \\ &\text{std} \left[ \bar{f}(Y_{s_1}^{(t,w)}) \bar{f}(Y_{l_1}^{(t,w)}) \right] \text{std} \left[ \bar{f}(Y_{s_2}^{(t,w)}) \bar{f}(Y_{l_2}^{(t,w)}) \right] \leq \\ &\left\| \bar{f}(Y_{s_1}^{(t,w)}) \bar{f}(Y_{l_1}^{(t,w)}) \right\|_2 \left\| \bar{f}(Y_{s_2}^{(t,w)}) \bar{f}(Y_{l_2}^{(t,w)}) \right\|_2 \leq \\ &\left\| \bar{f}(Y_{s_1}^{(t,w)}) \right\|_4 \left\| \bar{f}(Y_{l_1}^{(t,w)}) \right\|_4 \left\| \bar{f}(Y_{s_2}^{(t,w)}) \right\|_4 \left\| \bar{f}(Y_{l_2}^{(t,w)}) \right\|_4 \leq \max_{s=1, \dots, p} \left\| \bar{f}(Y_s^{(t,w)}) \right\|_4^4 \end{aligned} \quad (28)$$

Since  $\{Y_s^{(t,w)} : s = 1, \dots, p\}$  form a set of identically distributed random variables, then

$$\left\| \bar{f}(Y_1^{(t,w)}) \right\|_4 = \max_{s=1, \dots, p} \left\| \bar{f}(Y_s^{(t,w)}) \right\|_4 < \infty.$$

Let  $M_0(w)$  stands for the fourth moment of  $w\bar{f}(Y_1^{(t,w)})$ . We substantiated the uniform boundedness of  $M_0(w)$  in Lemma 19. So substituting Eq. (28) into Eq. (27) leads to

$$(\mathbb{E} |A_2|)^2 \leq \|A_2\|_2^2 \lesssim \left(\frac{|\nu|}{p}\right)^2 |\mathcal{C}_2| M_0(w) \asymp \left(\frac{|\nu|}{p}\right)^2 |\mathcal{C}_2|. \quad (29)$$

**Claim 1**  $|\mathcal{C}_2| \leq 8sd_{\max}^2 = 8p\bar{d}_{\max}^2$

**Proof** (Proof of Claim 1) Consider an arbitrary pair  $(s_1, l_1), (s_2, l_2) \in \mathcal{C}_2$ . There is a direct edge between  $s_1$  and  $l_1$  as  $R_{s_1, l_1} \neq 0$ . Similarly, there should be an edge connecting  $s_2$  and  $l_2$ . A direct connection between  $(s_1, l_1), (s_2, l_2)$ , means that four vertices  $(s_1, l_1, s_2, l_2)$  form a path of length less 1, 2, or 3. Let  $m_i$  denotes the number of distinct paths of length  $i$  in  $G_t$ . Then

$$|\mathcal{C}_2| \leq 4(m_1 + m_2 + m_3).$$

Since each node in network is connected to at most  $d_{\max}$  other nodes, one can easily verify that  $m_1 = s$  and

$$m_i = \sum_{s=1}^p \binom{d_s}{2} d_{\max}^{i-2} \leq \sum_{s=1}^p \frac{d_s^2}{2} d_{\max}^{i-2} \leq \frac{1}{2} d_{\max}^{i-1} \sum_{s=1}^p d_s = \frac{s}{2} d_{\max}^{i-1}.$$

Thus  $|\mathcal{C}_2| \leq 2s(2 + d_{\max} + d_{\max}^2) \leq 8sd_{\max}^2 = 8p\bar{d}d_{\max}^2$  ■

Replacing the upper bound on  $|\mathcal{C}_2|$  in Claim 1 into Eq. (29) shows that

$$\mathbb{E}|A_2| \leq |\nu| \sqrt{\frac{8\bar{d}d_{\max}^2}{p}}.$$

Now we find sufficient condition under which  $\mathbb{E}|A_1| \rightarrow 0$ . Observe that

$$\begin{aligned} |A_1| &= \sigma_{p,w}^{-1} \left| \sum_{s=1}^p \bar{f}(Y_s^{(t,w)}) \left( 1 - \frac{j\nu\xi_s}{\sigma_{p,w}} - e^{-j\nu\xi_s/\sigma_{p,w}} \right) \right| \leq \sigma_{p,w}^{-1} \sum_{s=1}^p \left| \bar{f}(Y_s^{(t,w)}) \right| \left| 1 - \frac{j\nu\xi_s}{\sigma_{p,w}} - e^{-j\nu\xi_s/\sigma_{p,w}} \right| \\ &\stackrel{(a)}{\leq} \frac{1}{2\sigma_{p,w}} \sum_{s=1}^p \left| \bar{f}(Y_s^{(t,w)}) \right| \left( \frac{\nu\xi_s}{\sigma_{p,w}} \right)^2 = \frac{\nu^2}{2\sigma_p^3} \sum_{s=1}^p \xi_s^2 \left| \bar{f}(Y_s^{(t,w)}) \right|. \end{aligned} \quad (30)$$

The inequality (a) is implied from the fact that  $|e^{-jy} + jy - 1| \leq y^2/2$  for any  $y \in \mathbb{R}$ . We can further simplify the upper bound on  $\mathbb{E}|A_1|$  by using Holder's inequality with  $p = 3$  and  $q = 3/2$ .

$$\mathbb{E}|A_1| \leq \frac{\nu^2}{2\sigma_p^3} \sum_{s=1}^p \mathbb{E}\xi_s^2 \left| \bar{f}(Y_s^{(t,w)}) \right| \leq \frac{\nu^2}{2\sigma_p^3} \sum_{s=1}^p \left( \mathbb{E}|\xi_s|^3 \right)^{\frac{2}{3}} \left( \mathbb{E} \left| \bar{f}(Y_s^{(t,w)}) \right|^3 \right)^{\frac{1}{3}}. \quad (31)$$

We previously argued in Eq. (24) that  $\sigma_{p,w} \geq \sqrt{p}g_2(w) \gtrsim w\sqrt{p}$ . This fact is crucial for transforming Eq. (31) into a desirable form.

$$\begin{aligned} \mathbb{E}|A_1| &\lesssim \nu^2 p^{\frac{3}{2}} \sum_{s=1}^p \left( \mathbb{E}|w\xi_s|^3 \right)^{\frac{2}{3}} \left( \mathbb{E} \left| w\bar{f}(Y_s^{(t,w)}) \right|^3 \right)^{\frac{1}{3}} = \nu^2 p^{-\frac{3}{2}} \sum_{s=1}^p \|w\xi_s\|_3^2 \left\| w\bar{f}(Y_s^{(t,w)}) \right\|_3 \\ &\stackrel{(b)}{\lesssim} \nu^2 p^{-\frac{3}{2}} \sum_{s=1}^p \|w\xi_s\|_3^2. \end{aligned} \quad (32)$$

Here inequality (b) is obvious implication of Lemma 19. We finally control  $\|w\xi_s\|_3$  from above. Again by using Holder's inequality with  $p = 3, q = 3/2$  and Lemma 19, we get

$$\begin{aligned} \|w\xi_s\|_3^3 &= \mathbb{E} \left[ \left| \sum_{l=1}^p w\bar{f}(Y_l^{(t,w)}) \mathbb{1}_{\{R_{s,l} \neq 0\}} \right|^3 \right] \leq d_s^2 \sum_{l=1}^p \mathbb{E} \left( \left| w\bar{f}(Y_l^{(t,w)}) \right|^3 \right) \mathbb{1}_{\{R_{s,l} \neq 0\}} \\ &\leq d_s^3 \max_{l=1, \dots, p} \mathbb{E} \left( \left| w\bar{f}(Y_l^{(t,w)}) \right|^3 \right) \lesssim d_s^3. \end{aligned} \quad (33)$$

We finally substitute Eq. (33) into Eq. (33).

$$\mathbb{E} |A_1| \lesssim \nu^2 p^{-\frac{3}{2}} \sum_{s=1}^p d_s^2 \leq \nu^2 p^{-\frac{3}{2}} \sum_{s=1}^p d_s d_{\max} = \nu^2 p^{-\frac{3}{2}} p \bar{d} d_{\max} = \nu^2 \frac{\bar{d} d_{\max}}{\sqrt{p}}.$$

Therefore  $\mathbb{E} |A_1|$  and  $\mathbb{E} |A_2|$  simultaneously tend to zero (for any fixed  $\nu \in \mathbb{R}$ ) whenever

$$\sqrt{\frac{\bar{d} d_{\max}^2}{p}} \vee \frac{\bar{d} d_{\max}}{\sqrt{p}} = \frac{\bar{d} d_{\max}}{\sqrt{p}} \rightarrow 0,$$

which exactly coincides with Assumption 3.1.  $\blacksquare$

**Proof** (Proof of Theorem 7) Let  $Q_\eta$  denotes  $(1 - \eta)$  quantile of  $\mathcal{N}(0, 1)$  distribution, for any  $\eta \in (0, 1)$ . We prove that

$$\mathbb{P}(T_t \geq Q_{\pi_0} \mid \mathbb{H}_{1,t}) \geq 1 - \pi_1.$$

Let  $\mu_{p,w}$  and  $\sigma_{p,w}^2$  respectively stand for the mean and variance of  $T_t$  under alternative hypothesis. Obviously

$$\mathbb{P}(T_t \geq Q_{\pi_0} \mid \mathbb{H}_{1,t}) = \mathbb{P}\left(\frac{T_t - \mu_{p,w}}{\sigma_{p,w}} \geq \frac{Q_{\pi_0} - \mu_{p,w}}{\sigma_{p,w}} \mid \mathbb{H}_{1,t}\right). \quad (34)$$

As smallest eigenvalue of  $\Omega^{(t+1)}$  is greater than  $\alpha_{\min}$  and its  $\ell_1$ -operator norm is bounded above by  $M$ , we can show that  $(T_t - \mu_{p,w})/\sigma_{p,w}$  has asymptotically a standard Gaussian distribution by employing exact same techniques as in the proof of Theorem 5. So the condition in Eq. (34) is satisfied if we can introduce a sufficient condition under which

$$\frac{Q_{\pi_0} - \mu_{p,w}}{\sigma_{p,w}} \leq -Q_{\pi_1}, \quad \text{or equivalently} \quad \frac{\mu_{p,w}}{\sigma_{p,w}} \left(1 - \frac{Q_{\pi_0}}{\mu_{p,w}}\right) \geq Q_{\pi_1}. \quad (35)$$

The inequality (35) trivially holds if we can prove that

$$\mu_{p,w} \geq 2Q_{\pi_0}, \quad \text{and} \quad \mu_{p,w} \geq 2\sigma_{p,w}Q_{\pi_1}. \quad (36)$$

It is easy to verify that  $\sigma_{p,w}$  is of order 1 as  $p, w \rightarrow \infty$ . So, we just need to find sufficient conditions under which  $\mu_{p,w} \gtrsim Q_{\pi_0} + Q_{\pi_1}$ . Recall  $\Delta$  from Eq. (16) and define

$$\bar{\Psi}_p := \frac{1}{p} \sum_{s=1}^p f(1 + \Delta_s), \quad \bar{\Psi}'_p := \frac{1}{p} \sum_{s=1}^p \Delta_s^2.$$

We first note that all terms  $\Delta_s$ ,  $s = 1, \dots, p$  are uniformly bounded, as for any  $s \in \{1, \dots, p\}$

$$\Delta_s = \frac{1}{\Omega_{ss}^{(t)}} \Omega_{s,\cdot}^{(t)} \left(\Omega^{(t+1)}\right)^{-1} \Omega_{\cdot,s}^{(t)} - 1 \leq \frac{\left\|\Omega_{\cdot,s}^{(t)}\right\|_{\ell_2}^2}{\lambda_{\min}(\Omega^{(t+1)}) \Omega_{ss}^{(t)}} - 1 \leq \frac{\left\|\Omega_{\cdot,s}^{(t)}\right\|_{\ell_2}^2}{\lambda_{\min}(\Omega^{(t+1)}) \lambda_{\min}(\Omega^{(t)})} - 1 \leq \frac{M^2}{\alpha_{\min}^2} - 1.$$

So we can find a bounded scalar  $C$  (depending on  $\frac{M^2}{\alpha_{\min}^2} - 1$ ) such that

$$\Delta_s^2 \leq Cf(1 + \Delta_s), \quad \forall s = 1, \dots, p \implies \bar{\Psi}'_p \leq C\bar{\Psi}_p.$$

We now have required tools for introducing a sufficient condition that guarantees two inequalities in Eq. (36). We begin by obtaining a sharp lower bound on  $\mu_{p,w}$ . Recall  $T_t$  from Eq. (13). Using the first part of Lemma 22 yields

$$\mu_{p,w} = \mathbb{E}(T_t | \mathbb{H}_{1,t}) = \frac{\sum_{s=1}^p f(1 + \Delta_s)}{g_2(w) \sqrt{\sum_{s_1, s_2=1}^p h_w(R_{s_1, s_2})}} \stackrel{(a)}{\gtrsim} \frac{pw\bar{\Psi}_p}{\sqrt{\sum_{s_1, s_2=1}^p h_w(R_{s_1, s_2})}}. \quad (37)$$

Notice that (a) in Eq. (37) holds since  $g_2(w) \geq c_0 w$  for some universal constant  $c_0 > 0$ . Moreover, it is easy to verify the existence of a bounded constant  $C_1(M, \alpha_{\min})$  for which

$$\sum_{s_1, s_2=1}^p h_w(R_{s_1, s_2}) \leq C_1(M, \alpha_{\min}) p.$$

Thus  $\mu_{p,w}$  in Eq. (37) satisfies  $\mu_{p,w} \gtrsim w\sqrt{p}\bar{\Psi}_p$ . So the first condition in Eq. (36) holds if

$$\bar{\Psi}_p \geq C_2(M, \alpha_{\min}) \frac{Q_{\pi_0} + Q_{\pi_1}}{w\sqrt{p}}, \quad (38)$$

which is same as the desired condition in Theorem 7.  $\blacksquare$

**Proof** (Proof of Theorem 11) Under Assumption 3.1 and,  $n_t, p$ , and possibly  $w, \bar{d}, d_{\max}$  asymptotically grow in such a way that

$$\frac{\bar{d}d_{\max}}{\sqrt{p}} \rightarrow 0, \quad \text{and} \quad \frac{pd_{\max} \log^2 p}{n} (w \vee \log p) \rightarrow 0. \quad (39)$$

All the statements regarding probabilistic convergence results in the proof are restricted to large  $p$  and  $n_t$  framework described in Eq. (39). According to Theorem 5,  $T_t$  converges in distribution to a standard Gaussian random variable. Therefore, we only require to show that  $|\hat{T}_t - T_t|$  converges to zero, in probability.

Recall  $\sigma_{p,w}$  and  $\bar{f}(\cdot)$  from Eq. (22) and define  $\hat{\sigma}_{p,w}$  (stands for estimated  $\sigma_{p,w}$ ) by

$$\hat{\sigma}_{p,w}^2 := g_2^2(w) \sum_{s_1, s_2=1}^p h_w(\hat{R}_{s_1, s_2}^{(t)}).$$

Using triangle inequality leads to

$$\begin{aligned} |\hat{T}_t - T_t| &= \left| \frac{1}{\hat{\sigma}_{p,w}} \sum_{s=1}^p \bar{f}(\hat{Y}_s^{(t,w)}) - \frac{1}{\sigma_{p,w}} \sum_{s=1}^p \bar{f}(Y_s^{(t,w)}) \right| \\ &\leq \frac{1}{\sigma_{p,w}} \left| \sum_{s=1}^p \bar{f}(\hat{Y}_s^{(t,w)}) - \bar{f}(Y_s^{(t,w)}) \right| + \left| \frac{1}{\hat{\sigma}_{p,w}} - \frac{1}{\sigma_{p,w}} \right| \left| \sum_{s=1}^p \bar{f}(Y_s^{(t,w)}) \right| \\ &= \frac{1}{\sigma_{p,w}} \left| \sum_{s=1}^p \bar{f}(\hat{Y}_s^{(t,w)}) - \bar{f}(Y_s^{(t,w)}) \right| + \left| \left( \frac{\sigma_{p,w}}{\hat{\sigma}_{p,w}} - 1 \right) T_t \right|. \end{aligned} \quad (40)$$

The following facts provides a simpler upper bound on  $|\hat{T}_t - T_t|$ .

- $\sigma_{p,w}$  is (obviously) greater than  $g_2(w) \sqrt{p}$ .
- Based upon Theorem 5,  $T_t$  converges in distribution to a standard Gaussian random variable. Therefore, in the asymptotic regime

$$\mathbb{P}\left(|T_t| \geq 4\sqrt{\log p}\right) \leq p^{-1}.$$

Applying these facts on inequality (40) yields

$$\mathbb{P}\left(\left|\frac{\hat{T}_t - T_t}{4}\right| \leq \left|\sum_{s=1}^p \frac{\bar{f}\left(\hat{Y}_s^{(t,w)}\right) - \bar{f}\left(Y_s^{(t,w)}\right)}{g_2(w) \sqrt{p}}\right| + \left|\frac{\sigma_{p,w}}{\hat{\sigma}_{p,w}} - 1\right| \sqrt{\log p}\right) \geq 1 - \frac{1}{p}. \quad (41)$$

For further simplification of the upper bound on  $|\hat{T}_t - T_t|$ , we need to control  $|\hat{\sigma}_{p,w} - \sigma_{p,w}|$  from above.

**Claim 1** *There exists a bounded constant  $C(\alpha_{\min}, M)$  such that*

$$\mathbb{P}\left(\left|\frac{\hat{\sigma}_{p,w}}{\sigma_{p,w}} - 1\right| \geq C(\alpha_{\min}, M) \varepsilon_{n_t, p, d_{\max}}\right) \leq \frac{2}{p}, \quad \text{where } \varepsilon_{n_t, p, d_{\max}} := d_{\max}^{\frac{1}{4}} \sqrt{\frac{\log p}{n_t}}.$$

Before proving Claim 1, we use it for simplifying Eq. (41). Under conditions in Eq. (39),  $\varepsilon_{n_t, p, d_{\max}} \sqrt{\log p}$  converges to zero, as

$$\varepsilon_{n_t, p, d_{\max}} \sqrt{\log p} = \frac{d_{\max}^{\frac{1}{4}} \log p}{\sqrt{n_t}} \lesssim \frac{d_{\max}^{\frac{1}{4}} \log p}{\sqrt{p d_{\max} \log^2 p}} = \left(p \sqrt{d_{\max}}\right)^{-\frac{1}{2}} \rightarrow 0.$$

Therefore for large  $p$  and  $n$ ,  $2\hat{\sigma}_{p,w} \geq \sigma_{p,w}$  with probability at least  $1 - p^{-1}$ , and

$$\begin{aligned} \left|\frac{\sigma_{p,w}}{\hat{\sigma}_{p,w}} - 1\right| \sqrt{\log p} &= \frac{|\hat{\sigma}_{p,w} - \sigma_{p,w}|}{\hat{\sigma}_{p,w}} \sqrt{\log p} \leq \frac{2|\hat{\sigma}_{p,w} - \sigma_{p,w}|}{\sigma_{p,w}} \sqrt{\log p} \leq 2C(\alpha_{\min}, M) \varepsilon_{n_t, p, d_{\max}} \sqrt{\log p} \\ &\asymp \varepsilon_{n_t, p, d_{\max}} \sqrt{\log p} \rightarrow 0. \end{aligned} \quad (42)$$

Replacing Eq. (42) into Eq. (41) ensures that

$$|\hat{T}_t - T_t| = \mathcal{O}_{\mathbb{P}}\left(\left|\sum_{s=1}^p \frac{\bar{f}\left(\hat{Y}_s^{(t,w)}\right) - \bar{f}\left(Y_s^{(t,w)}\right)}{g_2(w) \sqrt{p}}\right| \vee \varepsilon_{n_t, p, d_{\max}} \sqrt{\log p}\right).$$

So  $|\hat{T}_t - T_t| \rightarrow 0$  (in probability), which terminates the proof, if we prove that

$$\sum_{s=1}^p \left|\bar{f}\left(\hat{Y}_s^{(t,w)}\right) - \bar{f}\left(Y_s^{(t,w)}\right)\right| = o_{\mathbb{P}}\left(g_2(w) \sqrt{p}\right).$$

Since  $wg_2(w)$  is strictly less than some bounded universal constant  $C'_0$ , we just need to show that

$$\frac{w}{\sqrt{p}} \sum_{s=1}^p \left| \bar{f} \left( \hat{Y}_s^{(t,w)} \right) - \bar{f} \left( Y_s^{(t,w)} \right) \right| = o_{\mathbb{P}}(1). \quad (43)$$

Define

$$\eta_s := \frac{\hat{\Omega}_{s,:}^{(t)} (\Omega^{(t)})^{-1} \hat{\Omega}_{:,s}^{(t)}}{\hat{\Omega}_{ss}^{(t)}}, \quad \forall s = 1, \dots, p. \quad (44)$$

$\eta_s$  is an important quantity in our analysis as it captures the conditional expected value of  $\hat{Y}_s^{(t,w)}$  given  $X_1, \dots, X_t$ . Particularly, it is easy to show that

$$\mathbb{E} \left[ Y_s^{(t,w)} \mid X_1, \dots, X_t \right] = 1, \quad \text{and} \quad \mathbb{E} \left[ \hat{Y}_s^{(t,w)} \mid X_1, \dots, X_t \right] = \eta_s.$$

A vigilant reader notices that despite random nature of  $\eta_s$ , it does not depend on  $X_{t+1}, \dots, X_{t+w}$ . So using Corollary 25 yields

$$\mathbb{P} \left[ \left| \bar{f} \left( \hat{Y}_s^{(t,w)} \right) - \bar{f} \left( Y_s^{(t,w)} \right) \right| \geq f(\eta_s) + |\eta_s - 1| \sqrt{\frac{8 \log p}{w}} \left( 1 \vee \sqrt{\frac{8 \log p}{w}} \right) \right] \leq p^{-2}, \quad \forall s = 1, \dots, p.$$

Using union bound technique, we get

$$\mathbb{P} \left[ \left| \bar{f} \left( \hat{Y}_s^{(t,w)} \right) - \bar{f} \left( Y_s^{(t,w)} \right) \right| \geq f(\eta_s) + |\eta_s - 1| \sqrt{\frac{8 \log p}{w}} \left( 1 \vee \sqrt{\frac{8 \log p}{w}} \right), \quad \forall s = 1, \dots, p \right] \leq p^{-1}.$$

In summary, we showed that

$$\frac{w}{\sqrt{p}} \sum_{s=1}^p \left| \bar{f} \left( \hat{Y}_s^{(t,w)} \right) - \bar{f} \left( Y_s^{(t,w)} \right) \right| = \mathcal{O}_{\mathbb{P}} \left( \frac{w \sum_{s=1}^p f(\eta_s)}{\sqrt{p}} + \frac{\log p \vee \sqrt{w \log p}}{\sqrt{p}} \sum_{s=1}^p |\eta_s - 1| \right). \quad (45)$$

Using Cauchy-Schwartz inequality, we can rewrite Eq. (45) as the following.

$$\frac{w}{\sqrt{p}} \sum_{s=1}^p \left| \bar{f} \left( \hat{Y}_s^{(t,w)} \right) - \bar{f} \left( Y_s^{(t,w)} \right) \right| = \mathcal{O}_{\mathbb{P}} \left( \frac{w \sum_{s=1}^p f(\eta_s)}{\sqrt{p}} + \left( \log p \vee \sqrt{w \log p} \right) \left( \sum_{s=1}^p |\eta_s - 1|^2 \right)^{\frac{1}{2}} \right). \quad (46)$$

Let us investigate the behaviour of  $f(\eta_s)$  for  $s = 1, \dots, p$ . We assumed that  $\lambda_{\min}(\Omega^{(t)})$  is greater than  $\alpha_{\min}$ . It is also known that (see Theorem 1 of Cai et al. (2011)) with probability at least  $1 - 1/p$

$$\left\| \hat{\Omega}^{(t)} - \Omega^{(t)} \right\|_{2 \rightarrow 2} \leq CM^2 d_{\max} \sqrt{\frac{\log p}{n}} = \mathcal{O}_{\mathbb{P}} \left( d_{\max} \sqrt{\frac{\log p}{n}} \right) \rightarrow 0.$$

for some bounded universal constant  $C$ . Thus

$$\left( 1 - \frac{CM^2 d_{\max}}{\alpha_{\min}} \sqrt{\frac{\log p}{n}} \right) \hat{\Omega}^{(t)} \preceq \Omega^{(t)} \preceq \left( 1 + \frac{CM^2 d_{\max}}{\alpha_{\min}} \sqrt{\frac{\log p}{n}} \right) \hat{\Omega}^{(t)}. \quad (47)$$



The inequality (47) is essential for studying the behaviour of  $\eta_s$ , as

$$|\eta_s - 1| \leq \left(1 + \frac{CM^2 d_{\max}}{\alpha_{\min}} \sqrt{\frac{\log p}{n}}\right) \frac{\hat{\Omega}_{s,:}^{(t)} \left(\hat{\Omega}^{(t)}\right)^{-1} \hat{\Omega}_{:,s}^{(t)}}{\hat{\Omega}_{ss}} - 1 = \frac{CM^2 d_{\max}}{\alpha_{\min}} \sqrt{\frac{\log p}{n}} \rightarrow 0, \quad \forall s = 1, \dots, p.$$

So all  $\eta_s$  are in a small neighbourhood around one, which means that for large enough  $p$  and  $n_t$

$$f(\eta_s) \leq 2(\eta_s - 1)^2.$$

From Proposition 26 we know that if we replace the last inequality into Eq. (45), then we get

$$\frac{w}{\sqrt{p}} \sum_{s=1}^p \left| \bar{f}(\hat{Y}_s^{(t,w)}) - \bar{f}(Y_s^{(t,w)}) \right| = \mathcal{O}_{\mathbb{P}} \left( \frac{\sum_{s=1}^p (\eta_s - 1)^2}{w^{-1} \sqrt{p}} + (\log p \vee \sqrt{w \log p}) \left( \sum_{s=1}^p |\eta_s - 1|^2 \right)^{\frac{1}{2}} \right). \quad (48)$$

**Claim 2**  $\sum_{s=1}^p (\eta_s - 1)^2 = \mathcal{O}_{\mathbb{P}} \left( \frac{pd_{\max} \log p}{n_t} \right)$ . The constant involved in  $\mathcal{O}_{\mathbb{P}}$  depends on  $M$  and  $\alpha_{\min}$ .

Given Claim 2, by applying Proposition 26 we can rewrite Eq. (48) in the following way.

$$\begin{aligned} \frac{w}{\sqrt{p}} \sum_{s=1}^p \left| \bar{f}(\hat{Y}_s^{(t,w)}) - \bar{f}(Y_s^{(t,w)}) \right| &= \mathcal{O}_{\mathbb{P}} \left( \frac{pwd_{\max} \log p}{\sqrt{p}} + \sqrt{\frac{pd_{\max} \log^2 p (w \vee \log p)}{n_t}} \right) \\ &= \mathcal{O}_{\mathbb{P}} \left( \sqrt{\frac{pd_{\max} \log^2 p (w \vee \log p)}{n_t}} \right) \stackrel{(a)}{=} \mathcal{O}_{\mathbb{P}}(1). \end{aligned}$$

Here identity (a) is implied from assumptions in Eq. (39). Notice that the last equation is identical as Eq. (43), which concludes the proof. Finally, we establish Claim 1 and 2.

**Proof** (Proof of Claim 1) Lemma 28 states that as long as  $\alpha_{\min} > 0$ , the following result holds with probability at least  $1 - p^{-1}$ , and for some bounded  $C'(\alpha_{\min})$ .

$$\begin{aligned} \left| \frac{\hat{\sigma}_{p,w}}{\sigma_{p,w}} - 1 \right| &\leq C'(\alpha_{\min}) \left( \|R\|_{\ell_4}^{-1} \sqrt{\|\hat{\Omega}^{(t)} - \Omega^{(t)}\|_{\ell_{\infty}} \|\hat{\Omega}^{(t)} - \Omega^{(t)}\|_{\ell_2}} \vee \|\hat{\Omega}^{(t)} - \Omega^{(t)}\|_{\ell_{\infty}} \right) \\ &\stackrel{(a)}{\lesssim} p^{-\frac{1}{4}} \sqrt{\|\hat{\Omega}^{(t)} - \Omega^{(t)}\|_{\ell_{\infty}} \|\hat{\Omega}^{(t)} - \Omega^{(t)}\|_{\ell_2}} \vee \|\hat{\Omega}^{(t)} - \Omega^{(t)}\|_{\ell_{\infty}}. \end{aligned} \quad (49)$$

The inequality (a) is valid as all diagonal entries of  $R$  are equal to one. It is also known that (see Theorem 4 of Cai et al. (2011)) with probability at least  $1 - p^{-1}$

$$\|\hat{\Omega}^{(t)} - \Omega^{(t)}\|_{\ell_{\infty}} \leq C_0 M^2 \sqrt{\frac{\log p}{n_t}}, \quad \text{and} \quad \|\hat{\Omega}^{(t)} - \Omega^{(t)}\|_{\ell_2} \leq C_0 M^2 \sqrt{\frac{pd_{\max} \log p}{n_t}}, \quad (50)$$

for some bounded universal constant  $C_0$ . Combining Eq. (50) and Eq. (49) ends the proof.  $\blacksquare$

**Proof** [Proof of Claim 2] Recall  $\eta_s$  from Eq. (44). Observe that

$$\eta_s = \frac{\hat{\Omega}_{s,:}^{(t)} (\Omega^{(t)})^{-1} \hat{\Omega}_{:,s}^{(t)}}{\hat{\Omega}_{ss}^{(t)}} = \frac{\|(\Omega^{(t)})^{-1/2} \hat{\Omega}_{:,s}^{(t)}\|_{\ell_2}^2}{\hat{\Omega}_{ss}^{(t)}} = \left( \frac{\|(\Omega^{(t)})^{-1/2} \hat{\Omega}_{:,s}^{(t)}\|_{\ell_2}}{\sqrt{\hat{\Omega}_{ss}^{(t)}}} \right)^2$$

Using Eq. (47), we showed that all  $\delta_s$ ,  $s = 1, \dots, p$  are in a small neighborhood of one, with probability at least  $1 - 1/p$ . Thus,  $|\eta_s - 1| \leq 3|\sqrt{\eta_s} - 1|$  in the asymptotic framework and hence

$$\begin{aligned} |\sqrt{\eta_s} - 1| &= \left| \frac{\|(\Omega^{(t)})^{-1/2} \hat{\Omega}_{:,s}^{(t)}\|_{\ell_2}}{\sqrt{\hat{\Omega}_{ss}^{(t)}}} - 1 \right| \stackrel{(b)}{=} \left| \frac{\|(\Omega^{(t)})^{-1/2} \hat{\Omega}_{:,s}^{(t)}\|_{\ell_2}}{\sqrt{\hat{\Omega}_{ss}^{(t)}}} - \frac{\|(\Omega^{(t)})^{-1/2} \Omega_{:,s}^{(t)}\|_{\ell_2}}{\sqrt{\Omega_{ss}^{(t)}}} \right| \\ &\leq \left\| (\Omega^{(t)})^{-1/2} \left( \frac{\hat{\Omega}_{:,s}^{(t)}}{\sqrt{\hat{\Omega}_{ss}^{(t)}}} - \frac{\Omega_{:,s}^{(t)}}{\sqrt{\Omega_{ss}^{(t)}}} \right) \right\|_{\ell_2} \leq \sqrt{\frac{1}{\lambda_{\min}(\Omega^{(t)})}} \left\| \frac{\hat{\Omega}_{:,s}^{(t)}}{\sqrt{\hat{\Omega}_{ss}^{(t)}}} - \frac{\Omega_{:,s}^{(t)}}{\sqrt{\Omega_{ss}^{(t)}}} \right\|_{\ell_2} \\ &\leq \alpha_{\min}^{-\frac{1}{2}} \left\| \frac{\hat{\Omega}_{:,s}^{(t)}}{\sqrt{\hat{\Omega}_{ss}^{(t)}}} - \frac{\Omega_{:,s}^{(t)}}{\sqrt{\Omega_{ss}^{(t)}}} \right\|_{\ell_2}. \end{aligned} \quad (51)$$

Note that identity (b) is an immediate consequence of the fact that

$$\Omega_{ss}^{(t)} = \left[ \Omega^{(t)} (\Omega^{(t)})^{-1} \Omega^{(t)} \right]_{s,s} = \Omega_{:,s}^{(t)} (\Omega^{(t)})^{-1} \Omega_{:,s}^{(t)} = \left\| (\Omega^{(t)})^{-\frac{1}{2}} \Omega_{:,s}^{(t)} \right\|_{\ell_2}^2.$$

For simplicity define the following two matrices

$$L^{(t)} := \left[ \frac{\Omega_{s_1, s_2}^{(t)}}{\sqrt{\Omega_{s_2, s_2}^{(t)}}} \right]_{s_1, s_2=1}^p, \quad \hat{L}^{(t)} := \left[ \frac{\hat{\Omega}_{s_1, s_2}^{(t)}}{\sqrt{\hat{\Omega}_{s_2, s_2}^{(t)}}} \right]_{s_1, s_2=1}^p.$$

Next, we find an upper bound on  $\sum_{s=1}^p (\eta_s - 1)^2$  by using Eq. (51).

$$\sum_{s=1}^p (\eta_s - 1)^2 \leq 9 \sum_{s=1}^p (\sqrt{\eta_s} - 1)^2 \leq \frac{9}{\alpha_{\min}} \sum_{s=1}^p \left\| \hat{L}_{:,s}^{(t)} - L_{:,s}^{(t)} \right\|_{\ell_2}^2 = \frac{9}{\alpha_{\min}} \left\| \hat{L}^{(t)} - L^{(t)} \right\|_{\ell_2}^2.$$

Lemma 29 provides an upper bound on  $\left\| \hat{L}^{(t)} - L^{(t)} \right\|_{\ell_2}$ . There is  $C_1 < \infty$  (depending on  $\alpha_{\min}$ ) such that

$$\begin{aligned} \left\| \hat{L}^{(t)} - L^{(t)} \right\|_{\ell_2}^2 &\leq C_1 \left( \left\| \Omega^{(t)} \right\|_{\ell_2}^2 \left\| \hat{\Omega}^{(t)} - \Omega^{(t)} \right\|_{\ell_\infty}^2 + \left\| \hat{\Omega}^{(t)} - \Omega^{(t)} \right\|_{\ell_2}^2 \right) \\ &\stackrel{(c)}{=} \mathcal{O}_{\mathbb{P}} \left( M^2 \frac{\log p}{n_t} \left\| \Omega^{(t)} \right\|_{\ell_2}^2 + M^2 \frac{pd_{\max} \log p}{n_t} \right). \end{aligned}$$

Note that identity (c) is implied by the rates obtained in Theorem 4 of Cai et al. (2011). Recall that  $\ell_1$  operator norm of  $\Omega^{(t)}$  is less than  $M$ . Therefore

$$\left\| \Omega^{(t)} \right\|_{\ell_2}^2 \leq p \left\| \Omega^{(t)} \right\|_{2 \rightarrow 2}^2 \leq p \left\| \Omega^{(t)} \right\|_{2 \rightarrow 2}^2 \leq pM^2.$$

We conclude the proof by combining the last two inequalities.

$$\left\| \hat{L}^{(t)} - L^{(t)} \right\|_{\ell_2}^2 = \mathcal{O}_{\mathbb{P}} \left( M^4 \frac{p \log p}{n_t} + M^2 \frac{pd_{\max} \log p}{n_t} \right) = \mathcal{O}_{\mathbb{P}} \left( \frac{pd_{\max} \log p}{n_t} \right).$$

■

■

## Appendix A. Auxiliary results on moments of a Chi-square random variable

We used  $Y_s^{(t,w)}$  (recall its definition from Eq. (6)) to denote the sample mean of conditional log-likelihood of node  $s$  given other nodes of  $G_t$ , over a window of size  $w$ . We argued in Section 2.2 that if GGM does not experience a sudden change at  $t + 1$ , then  $wY_s^{(t,w)}$  is distributed as a Chi-square random variable with  $w$  degrees of freedom. Also recall  $f : [0, \infty] \mapsto \mathbb{R}$  from Eq. (3). In this section we formulate central moments of  $f\left(Y_s^{(t,w)}\right)$ ,  $s = 1, \dots, p$  under  $\mathbb{H}_{0,t}$  (no change at  $t + 1$ ) and  $\mathbb{H}_{1,t}$  (abrupt change at  $t + 1$ ). The two following results, Lemma 16 and 17, are crucial in the proof of principal results in this section.

**Lemma 16** *Let  $X$  and  $Y$  be two real-valued independent zero-mean random variables. For arbitrary functions  $f_1, f_2 : \mathbb{R} \mapsto \mathbb{R}$ , we have*

$$\text{cov}\left(f_1(X), Y f_2(X)\right) = 0.$$

**Proof** Independence of  $X$  and  $Y$  means that

$$\begin{aligned} \text{cov}\left(f_1(X), Y f_2(X)\right) &= \mathbb{E}f_1(X) f_2(X) Y - \mathbb{E}f_1(X) \mathbb{E}f_2(X) Y \\ &= \mathbb{E}f_1(X) f_2(X) \mathbb{E}Y - \mathbb{E}f_1(X) \mathbb{E}f_2(X) \mathbb{E}Y = \mathbb{E}Y \text{cov}\left(f_1(X), f_2(X)\right). \end{aligned}$$

Thus  $f_1(X)$  and  $Y f_2(X)$  are obviously uncorrelated, since  $\mathbb{E}Y = 0$ . ■

**Lemma 17** *Let  $(X, X')$  be a standard bi-variate Gaussian random vector with  $\text{corr}(X, X') = r$ . Let  $\{(X_k, X'_k)\}_{k=1}^w$  be i.i.d. draws from  $(X, X')$  and define*

$$Z_w := \sum_{k=1}^w X_k^2, \quad Z'_w := \sum_{k=1}^w X_k'^2.$$

Then,

$$\begin{aligned} p(r, w) &:= \text{cov}\left[(Z_w - w)^2, (Z'_w - w)^2\right] = 8r^2 [4w + r^2 w^2 + 2r^2], \\ q(r, w) &:= \text{cov}\left[(Z_w - w)^2, (Z'_w - w)^3\right] = 72r^4 (w^2 + w + 2) + 240r^2 w. \end{aligned}$$

**Proof**  $Z_w$  and  $Z'_w$  are identically distributed chi-square random variables with  $w$  degrees of freedom. We aim to find a recursive formula for  $p(r, w)$  in terms of  $p(r, w - 1)$  and  $p(r, 1)$ .  $Z_w$  and  $Z'_w$  can be formulated in a recursive way.

$$Z_w - w = \left(Z_{w-1} - (w-1)\right) + X_w^2 - 1, \quad Z'_w - w = \left(Z'_{w-1} - (w-1)\right) + (X_w'^2 - 1).$$

For simplicity, define  $\bar{Z}_w = Z_w - w$  and  $\bar{Z}'_w = Z'_w - w$ . Notice that the pair of random variables  $(\bar{Z}_{w-1}, \bar{Z}'_{w-1})$  are independent from  $X_w'$  and  $X_w$ . Thus,

$$\text{cov}\left[\bar{Z}_{w-1}^2, (X_w'^2 - 1)^2\right] = 0, \quad \text{cov}\left[\bar{Z}'_{w-1}^2, (X_w^2 - 1)^2\right] = 0. \quad (52)$$

Furthermore, Lemma 16 implies that

$$\begin{aligned} \text{cov} \left[ (X_w^2 - 1)^2, \bar{Z}'_{w-1} (X_w'^2 - 1) \right] &= 0, & \text{cov} \left[ (X_w'^2 - 1)^2, \bar{Z}_{w-1} (X_w^2 - 1) \right] &= 0, \\ \text{cov} \left[ \bar{Z}_{w-1}^2, \bar{Z}'_{w-1} (X_w'^2 - 1) \right] &= 0, & \text{cov} \left[ \bar{Z}'_{w-1}{}^2, \bar{Z}_{w-1} (X_w^2 - 1) \right] &= 0. \end{aligned} \quad (53)$$

The identities (52) and (53) are crucial for finding a simple alternative way of expressing  $p(r, w)$ .

$$\begin{aligned} p(r, w) &= \text{cov} \left[ (X_w^2 - 1)^2 + \bar{Z}_{w-1}^2 + 2\bar{Z}_{w-1} (X_w^2 - 1), (X_w'^2 - 1)^2 + \bar{Z}'_{w-1}{}^2 + 2\bar{Z}'_{w-1} (X_w'^2 - 1) \right] \\ &= \text{cov} \left[ (X_w^2 - 1)^2, (X_w'^2 - 1)^2 \right] + \text{cov} \left[ \bar{Z}_{w-1}^2, \bar{Z}'_{w-1}{}^2 \right] + 4 \text{cov} \left[ \bar{Z}_{w-1} (X_w^2 - 1), \bar{Z}'_{w-1} (X_w'^2 - 1) \right] \\ &= p(r, 1) + p(r, w - 1) + 4 \text{cov} \left[ \bar{Z}_{w-1} (X_w^2 - 1), \bar{Z}'_{w-1} (X_w'^2 - 1) \right]. \end{aligned}$$

Next we evaluate  $\text{cov} \left[ \bar{Z}_{w-1} (X_w^2 - 1), \bar{Z}'_{w-1} (X_w'^2 - 1) \right]$  in terms of  $w$  and  $r$ . Since  $X_w^2 - 1$  is a zero mean random variable, so

$$\mathbb{E} \bar{Z}_{w-1} (X_w^2 - 1) = \mathbb{E} \bar{Z}_{w-1} \mathbb{E} (X_w^2 - 1) = 0.$$

Therefore,

$$\begin{aligned} \text{cov} \left[ \bar{Z}_{w-1} (X_w^2 - 1), \bar{Z}'_{w-1} (X_w'^2 - 1) \right] &= \mathbb{E} (\bar{Z}_{w-1} \bar{Z}'_{w-1}) \mathbb{E} \left[ (X_w^2 - 1) (X_w'^2 - 1) \right] = \\ \text{cov} (\bar{Z}_{w-1}, \bar{Z}'_{w-1}) \text{cov} (X_w^2, X_w'^2) &= \sum_{i,j=1}^{w-1} \text{cov} (X_i^2, X_j'^2) \text{cov} (X_w^2, X_w'^2) \stackrel{(b)}{=} \\ (w-1) [\text{cov} (X_1^2, X_1'^2)]^2 &\stackrel{(c)}{=} 4r^4 (w-1). \end{aligned}$$

Identity (b) is an obvious consequence of the independence of  $X_i$  and  $X_j'$ , for any distinct pair  $(i, j)$ . Finally equality (c) is implied from the fact that  $\text{cov} (X_1^2, X_1'^2) = 2r^2$ . Thus,

$$p(r, w) = p(r, w - 1) + p(r, 1) + 4r^4 (w - 1), \quad \forall w > 1. \quad (54)$$

Eq. (54) can be easily reformulated as the following way

$$p(r, w) = p(r, 1) + (w - 1) p(r, 1) + 4r^4 \sum_{k=1}^w (k - 1) = wp(r, 1) + 8r^4 (w - 1) (w - 2). \quad (55)$$

In the last step, we calculate  $p(r, 1)$ . The cumbersome algebraic details (which is mainly based on using moment identities for standard Gaussian distribution) are omitted due to space constraint. The fact that  $Z_1$  and  $Z_1'$  has the same distribution yields

$$\begin{aligned} p(r, 1) &= \mathbb{E} \left[ (Z_1 - 1)^2 (Z_1' - 1)^2 \right] - \text{var} (Z_1) \text{var} (Z_1') = \mathbb{E} \left[ (Z_1 - 1)^2 (Z_1' - 1)^2 \right] - 4 \\ &= \mathbb{E} Z_1^2 Z_1'^2 - 4\mathbb{E} Z_1^2 Z_1' + 2\mathbb{E} Z_1^2 + 1 - 4\mathbb{E} Z_1 + 4\mathbb{E} Z_1 Z_1' - 4 \\ &= \mathbb{E} Z_1^2 Z_1'^2 - 4\mathbb{E} Z_1^2 Z_1' + 4\mathbb{E} Z_1 Z_1' - 1. \end{aligned} \quad (56)$$

Furthermore, using *Isserlis' Theorem* implies that

$$\begin{aligned}\mathbb{E}Z_1Z_1' &= \mathbb{E}Z_1\mathbb{E}Z_1' + 2(\mathbb{E}X_1X_1')^2 = 1 + 2r^2, \\ \mathbb{E}Z_1^2Z_1' &= \frac{1}{2}\binom{4}{2}(\mathbb{E}Z_1)^2\mathbb{E}Z_1' + 2 \times \binom{4}{2}\mathbb{E}Z_1(\mathbb{E}X_1X_1')^2 = 3 + 12r^2, \\ \mathbb{E}Z_1^2Z_1'^2 &= \mathbb{E}Z_1^2\mathbb{E}Z_1'^2 + 4!(\mathbb{E}X_1X_1')^4 + \frac{4^2 \times 3^2}{2}(\mathbb{E}X_1X_1')^2\mathbb{E}Z_1\mathbb{E}Z_1' = 9 + 72r^2 + 24r^4.\end{aligned}\quad (57)$$

Substituting Eq. (57) into (56) yields

$$p(r, 1) = \mathbb{E}Z_1^2Z_1'^2 - 9 - 40r^2 = 24r^4 + 72r^2 + 9 - 9 - 40r^2 = 8r^2(4 + 3r^2). \quad (58)$$

We terminate the proof by replacing Eq. (58) into Eq. (55). As the second claim in the statement of Lemma 17 can be substantiated by similar techniques, we skip its proof for avoiding repetitions.  $\blacksquare$

The succeeding Lemma comes in handy at Section 2.2 for standardizing  $T_t$  under  $\mathbb{H}_{0,t}$  regime.

**Lemma 18** *Let  $Z_w$  be a chi-square random variable with  $w$  degrees of freedom. Then the mean and standard deviation of  $f(Z_w/w)$  is given by*

$$g_1(w) := \mathbb{E}f\left(\frac{Z_w}{w}\right) = \log\left(\frac{w}{2}\right) - \psi^{(0)}\left(\frac{w}{2}\right), \quad g_2(w) := \text{std } f\left(\frac{Z_w}{w}\right) = \sqrt{\psi^{(1)}\left(\frac{w}{2}\right) - \frac{2}{w}}.$$

in which  $\psi^{(r)}$  stands for the poly-gamma function of order  $r$ .

**Proof** It is known that the expected value and variance of  $\log Z_w$  are respectively given by  $\log 2 + \psi^{(0)}(w/2)$  and  $\psi^{(1)}(w/2)$  Pav (2015). Therefore,

$$\mathbb{E}f\left(\frac{Z_w}{w}\right) = \mathbb{E}\left[\frac{Z_w}{w} - 1 - \log\left(\frac{Z_w}{w}\right)\right] = 1 - 1 + \log w - \mathbb{E}\log Z_w = \log\left(\frac{w}{2}\right) - \psi\left(\frac{w}{2}\right).$$

We now focus on calculating the variance of  $f\left(\frac{Z_w}{w}\right)$ . Observe that

$$\begin{aligned}\text{var } f\left(\frac{Z_w}{w}\right) &= \text{var}\left(\frac{Z_w}{w} - \log Z_w\right) = \frac{\text{var } Z_w}{w^2} + \text{var } \log Z_w - \frac{2 \text{cov}(Z_w, \log Z_w)}{w} \\ &= \frac{2}{w} + \psi^{(1)}\left(\frac{w}{2}\right) - \frac{2}{w} \text{cov}(Z_w, \log Z_w) \\ &= \frac{2}{w} + \psi^{(1)}\left(\frac{w}{2}\right) - \frac{2}{w} [\mathbb{E}(Z_w \log Z_w) - w\mathbb{E}\log Z_w].\end{aligned}\quad (59)$$

So we just need to evaluate the expected value of  $Z_w \log Z_w$ . Using the fact that  $\Gamma(x+1) = x\Gamma(x)$  for any  $x > 0$  leads to

$$\begin{aligned}\mathbb{E}(Z_w \log Z_w) &= \int_0^\infty x \log x \frac{x^{w/2-1} \exp(x/2)}{2^{w/2}\Gamma(w/2)} dx = \int_0^\infty \log x \frac{x^{(w+2)/2-1} \exp(x/2)}{2^{w/2}\Gamma(w/2)} dx \\ &= w \int_0^\infty \log x \frac{x^{(w+2)/2-1} \exp(x/2)}{2^{(w+2)/2}\Gamma((w+2)/2)} dx = w\mathbb{E}\log Z_{w+2}.\end{aligned}$$

Thus the covariance of  $Z_w$  and  $\log Z_w$  can be rewritten as

$$\begin{aligned} \text{cov}(Z_w, \log Z_w) &= w(\mathbb{E} \log Z_{w+2} - \mathbb{E} \log Z_w) = w \left[ \psi^{(0)}\left(\frac{w+2}{2}\right) + \log 2 - \psi^{(0)}\left(\frac{w}{2}\right) - \log 2 \right] \\ &= w \left[ \psi^{(0)}\left(1 + \frac{w}{2}\right) - \psi^{(0)}\left(\frac{w}{2}\right) \right] \stackrel{(a)}{=} w \frac{2}{w} = 2. \end{aligned} \quad (60)$$

In which the identity (a) is a direct implication of the fact that  $\psi^{(0)}(x+1) - \psi^{(0)}(x) = 1/x$ . Replacing Eq. (60) into Eq. (59) yields

$$\text{var} f\left(\frac{Z_w}{w}\right) = \frac{2}{w} + \psi^{(1)}\left(\frac{w}{2}\right) - \frac{4}{w} = \psi^{(1)}\left(\frac{w}{2}\right) - \frac{2}{w},$$

which concludes the proof.  $\blacksquare$

The following result, which is needed for proving Theorem 5, shows that all moments of  $wf\left(\frac{Z_w}{w}\right)$  are bounded

**Lemma 19** *There exist universal constant  $C_1, C_2 > 0$  such that*

$$\begin{aligned} \left\| wf\left(\frac{Z_w}{w}\right) \right\|_k &= w \left| \mathbb{E} f^k\left(\frac{Z_w}{w}\right) \right|^{1/k} \leq C_1 k, \\ \left\| w \left[ f\left(\frac{Z_w}{w}\right) - g_1(w) \right] \right\|_k &\leq C_2 k, \quad \forall k \in \mathbb{N}. \end{aligned}$$

**Proof** Let  $G_w(t)$  denotes the moments generating function of  $f(Z_w/w)$ . Due to non-negativity of  $f$ , if  $G_w$  is well defined for a strictly positive  $t$ , then all moments of  $f(Z_w/w)$  can be controlled from above by  $G_w(t)$ . We first show that  $G_w$  is well defined at  $t = w/4$ . Observe that

$$\begin{aligned} G_w\left(\frac{w}{4}\right) &= \mathbb{E} \left[ \exp\left(\frac{w}{4} f\left(\frac{Z_w}{w}\right)\right) \right] = \mathbb{E} \left[ \exp\left(\frac{Z_w}{4} - \frac{w}{4} - \frac{w}{4} \log\left(\frac{Z_w}{w}\right)\right) \right] \\ &= e^{-\frac{w}{4}} w^{\frac{w}{4}} \mathbb{E} \left[ Z_w^{-\frac{w}{4}} \exp\left(\frac{Z_w}{4}\right) \right] = e^{-\frac{w}{4}} w^{\frac{w}{4}} \int_0^\infty x^{-\frac{w}{4}} e^{\frac{x}{4}} \frac{x^{\frac{w}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{w}{2}} \Gamma\left(\frac{w}{2}\right)} dx \\ &= e^{-\frac{w}{4}} w^{\frac{w}{4}} \int_0^\infty \frac{x^{\frac{w}{4}-1} e^{-\frac{x}{4}}}{2^{\frac{w}{2}} \Gamma\left(\frac{w}{2}\right)} dx. \end{aligned}$$

The last integral can be simplified by introducing  $u := \frac{x}{4}$ .

$$G_w\left(\frac{w}{4}\right) = e^{-\frac{w}{4}} w^{\frac{w}{4}} \int_0^\infty \frac{u^{\frac{w}{4}-1} e^{-u}}{\Gamma\left(\frac{w}{2}\right)} du = \exp\left[-\frac{w}{4} + \frac{w}{4} \log w + \log \Gamma\left(\frac{w}{4}\right) - \log \Gamma\left(\frac{w}{2}\right)\right]. \quad (61)$$

This identity ensures that  $G_w\left(\frac{w}{4}\right)$  is well defined as long as  $w$  does not grow. Strictly speaking we can find bounded  $w_0 \in \mathbb{N}$  and  $C_0 > 0$  (depending on  $w_0$ ) so that  $G_w(w/4)$  is

smaller than  $C_0$  for all  $w \leq w_0$ . Next we study the behaviour of  $G_w\left(\frac{w}{4}\right)$  for large  $w$ . It is known that <sup>2</sup>

$$\lim_{x \rightarrow \infty} \log \Gamma(x) - \left(x - \frac{1}{2}\right) \log x + x = \frac{\log(2\pi)}{2}.$$

Applying the asymptotic identity of  $\log \Gamma$  function on Eq. (61) implies that

$$G_w\left(\frac{w}{4}\right) = \exp\left(\frac{1}{2} \log 2\right) = \sqrt{2}.$$

In summary, we guaranteed the existence of a bounded universal constant  $C_1 \geq 1$  for which  $G_w\left(\frac{w}{4}\right) \leq C_1$ , for all  $w \in \mathbb{N}$ . Elementary properties of moment generating function implies that

$$\frac{1}{k!} \left(\frac{w}{4}\right)^k \left\| f\left(\frac{Z_w}{w}\right) \right\|_k^k \leq G_w\left(\frac{w}{4}\right), \quad \forall k \in \mathbb{N} \implies \left\| \frac{w}{4} f\left(\frac{Z_w}{w}\right) \right\|_k \leq \sqrt[k]{k! G_w\left(\frac{w}{4}\right)} \leq \sqrt[k]{k!} C_1^{\frac{1}{k}}, \quad \forall k \in \mathbb{N}.$$

We conclude the proof by mentioning that  $\sqrt[k]{k!} \leq k$  for any  $k \in \mathbb{N}$ . Now we turn to the proof of second claim. Using Minkowski's inequality, we get

$$\left\| w \left[ f\left(\frac{Z_w}{w}\right) - g_1(w) \right] \right\|_k \leq \left\| w f\left(\frac{Z_w}{w}\right) \right\|_k + |w g_1(w)| \leq Ck + |w g_1(w)|.$$

So the desired result obviously holds, since there is a bounded universal scalar  $C' > 0$  such that  $|w g_1(w)| \leq C'$  for all  $w \in \mathbb{N}$  <sup>3</sup>. ■

We now find a polynomial approximation for  $h_w$  function defined in Eq. (12).

**Lemma 20** *Under the same notation and conditions as Lemma 17, define  $h_w[-1, 1] \mapsto [-1, 1]$  by*

$$h_w(r) := \text{corr} \left[ f\left(\frac{Z_w}{w}\right), f\left(\frac{Z'_w}{w}\right) \right]$$

Then

1.  $h_w$  is an even function, i.e.  $h_w(x) = h_w(-x)$  for any  $x \in [0, 1]$ , with  $h_w(0) = 0$  and  $h_w(1) = 1$ .
2. There exist a bounded universal scalar  $C > 0$  and  $w_0 \in \mathbb{N}$  such that for any  $w \geq w_0$

$$\max_{r \in [-1, 1]} |h_w(r) - r^4| \leq \frac{C}{w}.$$

**Proof** The first claim is obvious. So we focus on finding a uniform upper bound on the difference between  $h_w(r)$  and  $r^4$ , for  $r \in [-1, 1]$ . Choose an arbitrary  $r \in [-1, 1]$ . Define random variables  $\bar{Z}_w$  and  $\bar{Z}'_w$  by

$$\bar{Z}_w := (Z_w - w), \quad \bar{Z}'_w := (Z'_w - w). \tag{62}$$

2. See <http://functions.wolfram.com/GammaBetaErf/LogGamma/introductions/Gammas/ShowAll.html>

3. We refer the reader to 5.11.2 of <https://dlmf.nist.gov/5.11>



Taylor expansion of  $f$  around 1 gives us

$$f\left(\frac{Z_w}{w}\right) = f\left(1 + \frac{\bar{Z}_w}{w}\right) = \sum_{k=2}^{\infty} \frac{(-1)^k \bar{Z}_w^k}{kw^k}.$$

Therefore,

$$\begin{aligned} \text{cov}\left[f\left(\frac{Z_w}{w}\right), f\left(\frac{Z'_w}{w}\right)\right] &= \text{cov}\left[\sum_{k=2}^{\infty} \frac{(-1)^k \bar{Z}_w^k}{kw^k}, \sum_{k=2}^{\infty} \frac{(-1)^k \bar{Z}'_w{}^k}{kw^k}\right] = \sum_{j,k=2}^{\infty} \frac{(-1)^{j+k}}{jkw^{j+k}} \text{cov}\left(\bar{Z}_w^j, \bar{Z}'_w{}^k\right) \\ &= \frac{\text{cov}\left(\bar{Z}_w^2, \bar{Z}'_w{}^2\right)}{4w^4} - \frac{\text{cov}\left(\bar{Z}_w^3, \bar{Z}'_w{}^2\right)}{3w^5} + \sum_{j+k \geq 6}^{\infty} \frac{(-1)^{j+k}}{jkw^{j+k}} \text{cov}\left(\bar{Z}_w^j, \bar{Z}'_w{}^k\right) \end{aligned} \quad (63)$$

Lemma 17 gives an equivalent representation for the first two terms in the second line of Eq. (63). One line of algebra guarantees the existence of a large enough constant  $C_1$  such that

$$\left| \frac{\text{cov}\left(\bar{Z}_w^2, \bar{Z}'_w{}^2\right)}{4w^4} - \frac{\text{cov}\left(\bar{Z}_w^3, \bar{Z}'_w{}^2\right)}{3w^5} - \frac{2r^4}{w^2} \right| \leq \frac{C_1}{w^3}, \quad \forall w \geq 1. \quad (64)$$

We also use  $\xi$  to denote the third term in the second line of Eq. (63). Observe that

$$|\xi| = w^{-3} \left| \sum_{j+k \geq 6}^{\infty} \frac{(-1)^{j+k}}{jk} \left(\frac{2}{w}\right)^{(j+k-6)/2} \text{cov}\left[\left(\frac{\bar{Z}_w}{\sqrt{2w}}\right)^j, \left(\frac{\bar{Z}'_w}{\sqrt{2w}}\right)^k\right] \right|$$

We know from central limit Theorem that as  $w$  tends to infinity, then

$$\left(\frac{\bar{Z}_w}{\sqrt{2w}}, \frac{\bar{Z}'_w}{\sqrt{2w}}\right) \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & r^2 \\ r^2 & 1 \end{bmatrix}\right).$$

Thus, there exist a large enough  $w_0 \in \mathbb{N}$  and finite positive scalar  $C_2$  such that for any  $w \geq w'_0$

$$|\xi| \leq \frac{C_2}{w^3}. \quad (65)$$

Combining Eq. (63), (64) and (65) gives us

$$\left| \text{cov}\left[f\left(\frac{Z_w}{w}\right), f\left(\frac{Z'_w}{w}\right)\right] - \frac{2r^4}{w^2} \right| \leq \frac{C_1 + C_2}{w^3}, \quad \forall w \geq w'_0.$$

According to Lemma 18,  $\text{var}\left[f\left(\frac{Z_w}{w}\right)\right] = \text{var}\left[f\left(\frac{Z'_w}{w}\right)\right] = \psi^{(1)}\left(\frac{w}{2}\right) - \frac{2}{w}$ . So

$$\left| h_w(r) - \frac{2r^4}{w^2} \left(\psi^{(1)}\left(\frac{w}{2}\right) - \frac{2}{w}\right)^{-1} \right| \leq \frac{C_1 + C_2}{w^3} \left(\psi^{(1)}\left(\frac{w}{2}\right) - \frac{2}{w}\right)^{-1}. \quad (66)$$

It is known that<sup>4</sup> there are two finite scalars  $C'_1$  and  $w''_0$  such that

$$\left| \psi^{(1)}\left(\frac{w}{2}\right) - \frac{2}{w} - \frac{2}{w^2} \right| \leq \frac{C'_1}{w^3}, \quad \forall w \geq w''_0.$$

4. See 5.15.8 at <https://dlmf.nist.gov/5.15>

So inequality (66) can be simplified as

$$|h_w(r) - r^4| \leq \frac{C_1 + C_2}{2w - C'_1} + \frac{2C'_1 r^4}{w} \leq \frac{C}{w}.$$

Here  $C$  is a finite scalar depending on  $C_1, C_2, C'_1, w'_0$  and  $w''_0$ . In summary we showed that for any  $r \in [-1, 1]$ ,

$$|h_w(r) - r^4| \leq \frac{C}{w}, \quad \forall w \geq w_0 := w'_0 \vee w''_0.$$

■

**Lemma 21** *There exists a bounded constant  $C_w > 1$  such that  $\sup_{r \in [-1, 1]} \left| \frac{h_w(r)}{r^4} - 1 \right| \leq C_w$ .*

**Proof** Since  $h_w$  is a continuous function on a compact space, we only need to show that  $\limsup_{r \rightarrow 0} \frac{h_w(r)}{r^4} \leq C'_w := 1 + C_w$ . According to Lemma 19,  $h_w(r) \asymp r^4$  as  $w \rightarrow \infty$ , so we only present the proof for  $w = 1$ . ■

The following result is beneficiary for finding the expected value and standard deviation of  $T_t$ , under alternative hypothesis  $\mathbb{H}_{1,t}$ .

**Lemma 22** *Let  $(X, X')$  be a standard bi-variate Gaussian random vector with  $\text{corr}(X, X') = r$ . Let  $\{(X_k, X'_k)\}_{k=1}^w$  be i.i.d. draws from  $(X, X')$  and define*

$$Z_w := \sum_{k=1}^w X_k^2, \quad Z'_w := \sum_{k=1}^w X_k'^2.$$

Recall  $g_1(\cdot)$  and  $g_2(\cdot)$  from Lemma 18. For two arbitrary positive scalars  $\alpha, \alpha'$ ,

1.  $\mathbb{E}f\left(\frac{\alpha Z_w}{w}\right) = g_1(w) + f(\alpha)$ .
2.  $\text{cov}\left[f\left(\frac{\alpha Z_w}{w}\right), f\left(\frac{\alpha' Z'_w}{w}\right)\right] = \frac{2r^2}{w}(\alpha - 1)(\alpha' - 1) + h_w(r)g_2^2(w)$ .

**Proof** We employ a similar approach as the proof of Lemma 18 and 20. Notice that

$$f\left(\frac{\alpha Z_w}{w}\right) = f\left(\frac{Z_w}{w}\right) + (\alpha - 1)\frac{Z_w}{w} - \log \alpha. \quad (67)$$

taking expected value from both sides substantiates the first claim.

$$\mathbb{E}f\left(\frac{\alpha Z_w}{w}\right) = \mathbb{E}f\left(\frac{Z_w}{w}\right) + \alpha - 1 - \log \alpha = \mathbb{E}f\left(\frac{Z_w}{w}\right) + f(\alpha) = g_1(w) + f(\alpha).$$

We now aim to prove the second claim.  $h_w$  is defined to satisfy the following identity.

$$\text{cov}\left[f\left(\frac{Z_w}{w}\right), f\left(\frac{Z'_w}{w}\right)\right] = h_w(r)g_2^2(w).$$

Since covariance operator is bilinear, using Eq. (67) one can show that

$$\begin{aligned} & \text{cov} \left[ f \left( \frac{\alpha Z_w}{w} \right), f \left( \frac{\alpha' Z'_w}{w} \right) \right] - h_w(r) g_2^2(w) = \\ & (\alpha\alpha' - 1) \text{cov} \left( \frac{Z_w}{w}, \frac{Z'_w}{w} \right) - (\alpha - 1) \text{cov} \left( \frac{Z_w}{w}, \log \frac{Z'_w}{w} \right) - (\alpha' - 1) \text{cov} \left( \frac{Z'_w}{w}, \log \frac{Z_w}{w} \right). \end{aligned} \quad (68)$$

Let  $\{Y_k\}_{k=1}^w$  be a set of i.i.d. standard Gaussian random variables. Obviously

$$(Z_w, Z'_w) = \left( Z_w, r^2 Z_w, (1 - r^2) \sum_{k=1}^w Y_k^2 + 2r\sqrt{1 - r^2} \sum_{k=1}^w X_k Y_k \right).$$

This distributional identity alongside with Lemma 16 shows that

$$\begin{aligned} \text{cov}(Z_w, Z'_w) &= r^2 \text{var} Z_w = 2r^2 w, \\ \text{cov} \left( \frac{Z_w}{w}, \frac{Z'_w}{w} \right) &= \text{cov} \left( \frac{Z'_w}{w}, \log \frac{Z_w}{w} \right) = r^2 \text{cov} \left( \frac{Z_w}{w}, \log \frac{Z_w}{w} \right). \end{aligned}$$

So we can rewrite Eq. (68) as the following.

$$\text{cov} \left[ f \left( \frac{\alpha Z_w}{w} \right), f \left( \frac{\alpha' Z'_w}{w} \right) \right] - h_w(r) g_2^2(w) = \frac{2r^2}{w} (\alpha\alpha' - 1) - \frac{r^2 (\alpha + \alpha' - 2)}{w} \text{cov}(Z_w, \log Z_w).$$

We conclude the proof by applying Eq. (60).

$$\text{cov} \left[ f \left( \frac{\alpha Z_w}{w} \right), f \left( \frac{\alpha' Z'_w}{w} \right) \right] - h_w(r) g_2^2(w) = \frac{2r^2}{w} (\alpha\alpha' - 1) - \frac{2r^2}{w} (\alpha + \alpha' - 2) = \frac{2r^2}{w} (\alpha - 1) (\alpha' - 1). \quad \blacksquare$$

## Appendix B. Some technical probabilistic results

This section contains auxiliary results that are useful for proving main theoretical contributions of this manuscript in Section 7. The following Proposition comes in handy for substantiating Theorems 5 and 11.

**Proposition 23** *Let  $Z \in \mathbb{R}^p$  be a zero mean GGM with inverse covariance matrix  $\Omega$ . Define another (zero mean) GGM  $Y \in \mathbb{R}^p$  by*

$$Y_s = \frac{\langle \Omega_{s,:}, Z \rangle}{\sqrt{\Omega_{ss}}}, \quad \forall s \in \{1, \dots, p\}.$$

*Choose four distinct points  $s_1, t_1, s_2, t_2 \in \{1, \dots, p\}$  such that there is no edge between  $B_1 = \{s_1, t_1\}$  and  $B_2 = \{s_2, t_2\}$ . Then  $[Y_{s_1}, Y_{t_1}]$  is independent of  $[Y_{s_2}, Y_{t_2}]$ .*

**Proof** Let  $R$  be the partial correlation matrix of  $Z$ . It is easy to show that  $\text{cov}(Y) = R$ . Thus  $[Y_{s_1}, Y_{t_1}]$  is independent of  $[Y_{s_2}, Y_{t_2}]$  if and only if

$$\text{cov}([Y_{s_1}, Y_{t_1}], [Y_{s_2}, Y_{t_2}]) = R_{B_1, B_2} = \mathbf{0}_{2 \times 2}.$$

Since there is no edge between  $\{s_1, t_1\}$  and  $\{s_2, t_2\}$ , then  $\Omega_{B_1, B_2} = \mathbf{0}_{2 \times 2}$ . We conclude the proof by reminding the reader that  $R$  has the same sparsity pattern as  $\Omega$ , i.e.  $\text{supp}(R) = \text{supp}(\Omega)$ .  $\blacksquare$

Next we conduct a probabilistic sensitivity analysis for  $f$  with a stochastic argument (Chi-square random variable with  $w$ -degrees of freedom). Such result is crucial in the proof of Theorem 11.

**Lemma 24** *Let  $\alpha$  be a bounded strictly positive scalar and let  $Z_w$  be a chi-square random variables with  $w$  degrees of freedom. Then for any  $\xi > 0$ ,*

$$\mathbb{P} \left[ \left| f \left( \frac{\alpha Z_w}{w} \right) - f \left( \frac{Z_w}{w} \right) - f(\alpha) \right| \geq |\alpha - 1| \sqrt{\frac{8\xi \log p}{w}} \left( 1 \vee \sqrt{\frac{8\xi \log p}{w}} \right) \right] \leq p^{-2\xi}.$$

**Proof** Without loss of generality, assume that  $\alpha \neq 1$ . The following identity holds for any  $x > 0$ .

$$f(\alpha x) - f(x) - f(\alpha) = [\alpha x - 1 - \log(\alpha x)] - (x - 1 - \log x) - (\alpha - 1 - \log \alpha) = (\alpha - 1)(x - 1).$$

Thus

$$\frac{\sqrt{\frac{w}{2}}}{|\alpha - 1|} \left| f \left( \frac{\alpha Z_w}{w} \right) - f \left( \frac{Z_w}{w} \right) - f(\alpha) \right| = \frac{|Z_w - w|}{\sqrt{2w}}.$$

Therefore we just need a concentration inequality for a standardized version of  $Z_w$ . We know from Remark 2.11 of Boucheron et al. (2013) that for any  $t > 0$

$$\mathbb{P} \left( \frac{|Z_w - w|}{\sqrt{2w}} \geq \sqrt{2t} + \sqrt{\frac{2}{w}t} \right) \leq e^{-t}. \quad (69)$$

Replacing  $t = 2\xi \log p$  in Eq. (69) concludes the proof.  $\blacksquare$

One can extend Lemma 24 to random  $\alpha$ , as long as it is independent from  $Z_w$ . We skip the proof due to its simplicity.

**Corollary 25** *Lemma 24 is satisfied for a strictly positive random variable  $\alpha$ , independent of  $Z_w$ .*

**Proposition 26** *Let  $X$  and  $Y$  be random variables jointly distributed as  $\mathbb{P}$ . Let  $B$  be a measurable set with  $\mathbb{P}(Y \in B) \geq 1 - \epsilon_1$ , for some  $\epsilon_1 \in (0, 1)$ . If there exist a measurable set  $A$  and  $\epsilon_2 \in (0, 1)$  such that*

$$\mathbb{P}(A \mid Y = y) \leq \epsilon_2, \quad \forall y \in B, \quad (70)$$

then  $\mathbb{P}(A) \leq \epsilon_1 + \epsilon_2$

**Proof** Let  $\mathbb{P}_Y$  stands for the marginal distribution of  $Y$ . Observe that

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap [Y \notin B]) + \mathbb{P}(A \cap [Y \in B]) \leq \mathbb{P}(Y \notin B) + \mathbb{P}(A \cap [Y \in B]) \leq \epsilon_1 + \mathbb{P}(A \cap [Y \in B]) \\ &= \epsilon_1 + \int_B \mathbb{P}(A \mid Y = y) d\mathbb{P}_Y(y) \leq \epsilon_1 + \sup_{y \in B} \mathbb{P}(A \mid Y = y). \end{aligned}$$

Finally condition (70) trivially substantiates the desired upper bound on  $\mathbb{P}(A)$ .  $\blacksquare$

**Lemma 27** Let  $\Omega, \Omega' \in S_{++}^{p \times p}$  be two inverse covariance matrices such that

$$\lambda_{\min}(\Omega) \wedge \lambda_{\min}(\Omega') \geq \alpha_{\min}, \quad (71)$$

for a strictly positive scalar  $\alpha_{\min}$ . We use  $R$  and  $R'$  to denote the partial correlation matrices associated to  $\Omega$  and  $\Omega'$ , respectively. If  $\|\Omega' - \Omega\|_{\ell_\infty} \leq \alpha_{\min}$ , then there exists a bounded universal constant  $C$  such that

$$\frac{\|R' - R\|_{\ell_4}}{\|R\|_{\ell_4}} \leq \frac{C}{\alpha_{\min}^2} \left[ \frac{\sqrt{\|\Omega' - \Omega\|_{\ell_\infty} \|\Omega' - \Omega\|_{\ell_2}}}{\|R\|_{\ell_4}} + \|\Omega' - \Omega\|_{\ell_\infty} \right]$$

**Proof** For simplicity define  $\Delta := R' - R$ . Our objective is control  $|\Delta_{st}|$ ,  $s, t = 1, \dots, p$  from above. Using triangle inequality yields

$$\begin{aligned} |\Delta_{st}| &= \left| \frac{\Omega'_{st}}{\sqrt{\Omega'_{ss}\Omega'_{tt}}} - \frac{\Omega_{st}}{\sqrt{\Omega_{ss}\Omega_{tt}}} \right| \leq \frac{|\Omega'_{st} - \Omega_{st}|}{\sqrt{\Omega'_{ss}\Omega'_{tt}}} + |\Omega_{st}| \left| \frac{1}{\sqrt{\Omega'_{ss}\Omega'_{tt}}} - \frac{1}{\sqrt{\Omega_{ss}\Omega_{tt}}} \right| \\ &\leq \frac{|\Omega'_{st} - \Omega_{st}|}{\sqrt{\Omega'_{ss}\Omega'_{tt}}} \sqrt{\frac{\Omega_{tt}}{\Omega'_{tt}}} \sqrt{\frac{\Omega_{ss}}{\Omega'_{ss}}} + |R_{st}| \left| \sqrt{\frac{\Omega_{tt}}{\Omega'_{tt}}} \sqrt{\frac{\Omega_{ss}}{\Omega'_{ss}}} - 1 \right| \end{aligned}$$

Let us simplify the right hand side term in preceding inequality. The right hand side condition in Eq. (71) guarantees that all diagonal entries of  $\Omega$  are greater than  $\alpha_{\min}$ . Thus for all  $s \in \{1, \dots, p\}$ ,

$$\sqrt{\frac{\Omega_{ss}}{\Omega'_{ss}}} = \left( 1 + \frac{\Omega_{ss} - \Omega'_{ss}}{\Omega'_{ss}} \right)^{1/2} \leq 1 + \frac{|\Omega_{ss} - \Omega'_{ss}|}{2\Omega'_{ss}} \leq 1 + \frac{\|\Omega' - \Omega\|_{\ell_\infty}}{2\Omega'_{ss}} \leq 1 + \frac{\|\Omega' - \Omega\|_{\ell_\infty}}{2\alpha_{\min}}. \quad (72)$$

Moreover we know that  $\|\Omega' - \Omega\|_{\ell_\infty} \leq \alpha_{\min}$ . So, the upper bound on  $|\Delta_{st}|$  can be rewritten as

$$\begin{aligned} |\Delta_{st}| &\leq \frac{|\Omega'_{st} - \Omega_{st}|}{\alpha_{\min}} \left( 1 + \frac{\|\Omega' - \Omega\|_{\ell_\infty}}{2\alpha_{\min}} \right)^2 + |R_{st}| \left| \left( 1 + \frac{\|\Omega' - \Omega\|_{\ell_\infty}}{2\alpha_{\min}} \right)^2 - 1 \right| \\ &\leq \frac{9|\Omega'_{st} - \Omega_{st}|}{4\alpha_{\min}} + \frac{5|R_{st}|}{4\alpha_{\min}} \|\Omega' - \Omega\|_{\ell_\infty} \leq \frac{9\sqrt{|\Omega'_{st} - \Omega_{st}|}}{4\alpha_{\min}} \sqrt{\|\Omega' - \Omega\|_{\ell_\infty}} + \frac{5|R_{st}|}{4\alpha_{\min}} \|\Omega' - \Omega\|_{\ell_\infty}. \end{aligned}$$

We now apply the following result to find an upper bound on  $|\Delta_{st}|^4$ .

$$(x + y)^4 \leq 8(x^4 + y^4), \quad \forall x, y \geq 0.$$

The proof of preceding result is left to the reader. So,

$$|\Delta_{st}|^4 \leq \frac{206}{\alpha_{\min}^4} |\Omega'_{st} - \Omega_{st}|^2 \|\Omega' - \Omega\|_{\ell_\infty}^2 \vee \frac{20}{\alpha_{\min}^4} \|\Omega' - \Omega\|_{\ell_\infty}^4 R_{st}^4.$$

Now summing up over all entries of  $\Delta$  (omitting the straightforward algebra), we get

$$\|\Delta\|_{\ell_4}^4 \leq \frac{C}{\alpha_{\min}^4} \left[ \|\Omega' - \Omega\|_{\ell_\infty}^2 \|\Omega' - \Omega\|_{\ell_2}^2 + \|\Omega' - \Omega\|_{\ell_\infty}^4 \|R\|_{\ell_4}^4 \right], \quad (73)$$

for some bounded universal constant  $C$ . Taking forth root from both sides of Eq. (73) ends the proof.  $\blacksquare$

The following result is an easy consequence of Lemma 21 and Lemma 27.

**Lemma 28** *The following inequality holds under the same conditions of Lemma 27.*

$$\left| \frac{\sum_{s_1, s_2=1}^p h_w(R'_{s_1, s_2})}{\sum_{s_1, s_2=1}^p h_w(R_{s_1, s_2})} - 1 \right| \leq \frac{C}{\alpha_{\min}^2} \left[ \frac{\sqrt{\|\Omega' - \Omega\|_{\ell_\infty} \|\Omega' - \Omega\|_{\ell_2}}}{\|R\|_{\ell_4}} + \|\Omega' - \Omega\|_{\ell_\infty} \right].$$

**Proof** Lemma 21 ensures the existence of two bounded strictly positive constants  $C_w, C'_w$  such that

$$h_w(r) \geq C'_w r^4, \quad \text{and} \quad h_w(r) \leq C_w r^4, \quad \forall r \in [-1, 1].$$

Therefore,

$$\begin{aligned} \left( \sum_{s_1, s_2=1}^p h_w(R'_{s_1, s_2}) \right)^{\frac{1}{4}} &\leq C_w^{\frac{1}{4}} \|R'\|_{\ell_4} \leq C_w^{\frac{1}{4}} \|R\|_{\ell_4} \left( 1 + \frac{\|R' - R\|_{\ell_4}}{\|R\|_{\ell_4}} \right) \\ &\leq \left( \frac{C_w}{C'_w} \right)^{\frac{1}{4}} \left( \sum_{s_1, s_2=1}^p h_w(R_{s_1, s_2}) \right)^{\frac{1}{4}} \left( 1 - \frac{\|R' - R\|_{\ell_4}}{\|R\|_{\ell_4}} \right). \end{aligned} \quad (74)$$

Applying triangle inequality on Eq. (74) leads to

$$\frac{\sum_{s_1, s_2=1}^p h_w(R'_{s_1, s_2})}{\sum_{s_1, s_2=1}^p h_w(R_{s_1, s_2})} - 1 \lesssim \left( 1 - \frac{\|R' - R\|_{\ell_4}}{\|R\|_{\ell_4}} \right)^4 - 1 \leq \left| 1 - \frac{\|R' - R\|_{\ell_4}}{\|R\|_{\ell_4}} \right| - 1 \leq \frac{\|R' - R\|_{\ell_4}}{\|R\|_{\ell_4}}.$$

Employing similar techniques, one can show that  $1 - \frac{\sum_{s_1, s_2=1}^p h_w(R'_{s_1, s_2})}{\sum_{s_1, s_2=1}^p h_w(R_{s_1, s_2})} \leq \frac{\|R' - R\|_{\ell_4}}{\|R\|_{\ell_4}}$ . Thus,

$$\left| \frac{\sum_{s_1, s_2=1}^p h_w(R'_{s_1, s_2})}{\sum_{s_1, s_2=1}^p h_w(R_{s_1, s_2})} - 1 \right| \lesssim \frac{\|R' - R\|_{\ell_4}}{\|R\|_{\ell_4}}. \quad (75)$$

We conclude the proof by applying Lemma 27 to Eq. (75).  $\blacksquare$

**Lemma 29** Let  $\Omega, \Omega' \in S_{++}^{p \times p}$  be two precision matrices such that

$$\lambda_{\min}(\Omega) \wedge \lambda_{\min}(\Omega') \geq \alpha_{\min},$$

for a strictly positive scalar  $\alpha_{\min}$ . Define  $L_1, L_2 \in \mathbb{R}^{p \times p}$  by

$$L := \left[ \frac{\Omega_{st}}{\sqrt{\Omega_{tt}}} \right]_{s,t=1}^p, \quad \text{and} \quad L' := \left[ \frac{\Omega'_{st}}{\sqrt{\Omega'_{tt}}} \right]_{s,t=1}^p.$$

If  $\|\Omega' - \Omega\|_{\ell_\infty} \leq \alpha_{\min}$ , then there exists a bounded universal constant  $C$  such that

$$\|L' - L\|_{\ell_2} \leq C \left( \frac{\|\Omega' - \Omega\|_{\ell_2}}{\sqrt{\alpha_{\min}}} + \frac{\|\Omega\|_{\ell_2} \|\Omega' - \Omega\|_{\ell_\infty}}{\sqrt{\alpha_{\min}^3}} \right).$$

**Proof** We apply similar techniques as the proof of Lemma 27. For simplicity define  $\Delta := L' - L$ . Triangle inequality leads to

$$|\Delta_{st}| \leq \frac{|\Omega'_{st} - \Omega_{st}|}{\sqrt{\Omega'_{tt}}} + \frac{|\Omega_{st}|}{\sqrt{\Omega_{tt}}} \left| \sqrt{\frac{\Omega_{tt}}{\Omega'_{tt}}} - 1 \right|, \quad \forall s, t \in \{1, \dots, p\}.$$

Using Eq. (72) and the fact that  $\|\Omega' - \Omega\|_{\ell_\infty} \leq \alpha_{\min}$  are critical for simplifying the obtained upper bound on  $|\Delta_{st}|$ . We skip the algebraic details as it is exactly the same in Eq. (72).

$$|\Delta_{st}|^2 \leq \left[ \frac{|\Omega'_{st} - \Omega_{st}|}{\sqrt{\alpha_{\min}}} + \frac{|\Omega_{st}| \|\Omega' - \Omega\|_{\ell_\infty}}{2\sqrt{\alpha_{\min}^3}} \right]^2 \leq 2 \left( \frac{(\Omega'_{st} - \Omega_{st})^2}{\alpha_{\min}} + \frac{\Omega_{st}^2 \|\Omega' - \Omega\|_{\ell_\infty}^2}{4\alpha_{\min}^3} \right), \quad (76)$$

for any  $s, t \in \{1, \dots, p\}$ . Combining Eq. (76) for all pairs  $(s, t)$  in  $\{1, \dots, p\} \times \{1, \dots, p\}$ , we get

$$\|\Delta\|_{\ell_2}^2 = \sum_{s,t=1}^p |\Delta_{st}|^2 \leq \frac{2}{\alpha_{\min}} \|\Omega' - \Omega\|_{\ell_2}^2 + \frac{1}{2\alpha_{\min}^3} \|\Omega\|_{\ell_2}^2 \|\Omega' - \Omega\|_{\ell_\infty}^2.$$

This inequality can be easily reformulated as the desirable upper bound on  $\|\Delta\|_{\ell_2}$ . ■

## References

- Yves Atchade and Leland Bybee. A scalable algorithm for gaussian graphical models with change-points. *arXiv preprint arXiv:1707.04306*, 2017.
- Alexander Aue, Siegfried H\"ormann, Lajos Horv\'ath, Matthew Reimherr, et al. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37(6B):4046–4087, 2009.
- Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78, 1998.

- Jushan Bai and Pierre Perron. Computation and analysis of multiple structural change models. *Journal of applied econometrics*, 18(1):1–22, 2003.
- Ian Barnett and Jukka-Pekka Onnela. Change point detection in correlation networks. *Scientific reports*, 6:18893, 2016.
- Michèle Basseville and Igor V Nikiforov. *Detection of abrupt changes: theory and application*, volume 104. Prentice Hall Englewood Cliffs, 1993.
- David A Bessler and Jian Yang. The structure of interdependence in international stock markets. *Journal of international money and finance*, 22(2):261–287, 2003.
- Erwin Bolthausen. On the central limit theorem for stationary mixing random fields. *The Annals of Probability*, pages 1047–1050, 1982.
- Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.
- Peter Bühlmann and Sara Van De Geer. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.
- Tony Cai, Weidong Liu, and Xi Luo. A constrained  $\ell_1$  minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.
- Tony Cai, Weidong Liu, and Yin Xia. Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association*, 108(501):265–277, 2013.
- Hao Chen and Nancy Zhang. Graph-based change-point detection. *The Annals of Statistics*, 43(1):139–176, 2015.
- Alex Gibberd and Sandipan Roy. Consistent multiple changepoint estimation with fused gaussian graphical models. *Annals of the Institute of Statistical Mathematics*, pages 1–27, 2020.
- Xavier Guyon. *Random fields on a network: modeling, statistics, and applications*. Springer Science & Business Media, 1995.
- Rikkert Hindriks, Mohit H Adhikari, Yusuke Murayama, Marco Ganzetti, Dante Mantini, Nikos K Logothetis, and Gustavo Deco. Can sliding-window correlations reveal dynamic functional connectivity in resting-state fmri? *Neuroimage*, 127:242–256, 2016.
- Lajos Horváth and Gregory Rice. Extensions of some classical methods in change point analysis. *Test*, 23(2):219–255, 2014.
- Cho-Jui Hsieh, Mátyás A Sustik, Inderjit S Dhillon, and Pradeep Ravikumar. Quic: quadratic approximation for sparse inverse covariance estimation. *Journal of Machine Learning Research*, 15(1):2911–2947, 2014.



- R Matthew Hutchison, Thilo Womelsdorf, Joseph S Gati, Stefan Everling, and Ravi S Menon. Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques. *Human brain mapping*, 34(9):2154–2177, 2013.
- Beatrix Jones and Mike West. Covariance decomposition in undirected gaussian graphical models. *Biometrika*, 92(4):779–786, 2005.
- Michael G Kallitsis, Shrijita Bhattacharya, Stilian Stoev, and George Michailidis. Adaptive statistical detection of false data injection attacks in smart grids. In *Signal and Information Processing (GlobalSIP), 2016 IEEE Global Conference on*, pages 826–830. IEEE, 2016.
- Hossein Keshavarz, Clayton Scott, and XuanLong Nguyen. Optimal change point detection in gaussian processes. *Journal of Statistical Planning and Inference*, 193:151–178, 2018.
- Mladen Kolar and Eric P Xing. Estimating networks with jumps. *Electronic journal of statistics*, 6:2069, 2012.
- Mladen Kolar, Le Song, Amr Ahmed, and Eric P Xing. Estimating time-varying networks. *The Annals of Applied Statistics*, pages 94–123, 2010.
- Jun Li and Song Xi Chen. Two sample tests for high-dimensional covariance matrices. *The Annals of Statistics*, 40(2):908–940, 2012.
- Steven E Pav. Moments of the log non-central chi-square distribution. *arXiv preprint arXiv:1503.06266*, 2015.
- Sheldon M Ross. *Introduction to probability models*. Academic press, 2014.
- Sandipan Roy, Yves Atchadé, and George Michailidis. Change point estimation in high dimensional markov random-field models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2016.
- Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*. The Regents of the University of California, 1972.
- Martin J Wainwright and Michael I Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.
- Xindong Wu, Xingquan Zhu, Gong-Qing Wu, and Wei Ding. Data mining with big data. *IEEE transactions on knowledge and data engineering*, 26(1):97–107, 2014.
- Lin Xiao. Dual averaging methods for regularized stochastic learning and online optimization. *Journal of Machine Learning Research*, 11(Oct):2543–2596, 2010.