Bernoulli **30**(1), 2024, 712–742
https://doi.org/10.3150/23-BEJ1614

Central limit theorems for high dimensional dependent data

JINYUAN CHANG[1,2,a], XIAOHUI CHEN[3,4,c] and MINGCONG WU[1,b]

[1] *Joint Laboratory of Data Science and Business Intelligence, Southwestern University of Finance and Economics, Chengdu, China,* [a]*changjinyuan@swufe.edu.cn,* [b]*wumingcong@smail.swufe.edu.cn*
[2] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China*
[3] *Department of Mathematics, University of Southern California, Los Angeles, CA, USA,* [c]*xiaohuic@usc.edu*
[4] *Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL, USA*

Motivated by statistical inference problems in high-dimensional time series data analysis, we first derive non-asymptotic error bounds for Gaussian approximations of sums of high-dimensional dependent random vectors on hyper-rectangles, simple convex sets and sparsely convex sets. We investigate the quantitative effect of temporal dependence on the rates of convergence to a Gaussian random vector over three different dependency frameworks (α-mixing, m-dependent, and physical dependence measure). In particular, we establish new error bounds under the α-mixing framework and derive faster rate over existing results under the physical dependence measure. To implement the proposed results in practical statistical inference problems, we also derive a data-driven parametric bootstrap procedure based on a kernel-type estimator for the long-run covariance matrices. The unified Gaussian and parametric bootstrap approximation results can be used to test mean vectors with combined ℓ^2 and ℓ^∞ type statistics, do change point detection, and construct confidence regions for covariance and precision matrices, all for time series data.

Keywords: Central limit theorem; dependent data; Gaussian approximation; high-dimensional statistical inference; parametric bootstrap

1. Introduction

High-dimensional dependent data are frequently encountered in current practical problems of finance, biomedical sciences, geological studies and many more areas. Due to the complicated dependency among different components and nonlinear dynamical behaviors in the series, there have been tremendous challenges in developing principled statistical inference procedures for such data. Most existing methods require certain parametric assumptions on the underlying data generation mechanism or structural assumptions on the dependency among different components in order to derive asymptotically pivotal distributions of the involved statistics. Assumptions of this kind are not only difficult to be verified but also often violated in real data. How to derive statistically valid inference procedures that do not rely on specific structural assumptions imposed on the dependency among different components of high-dimensional dependent data has been an urgent demand.

In this paper, we focus on establishing quantitative high-dimensional Central Limit Theorems (CLTs) and related parametric bootstrap approximations for dependent (and possibly non-stationary) data. Let $\mathcal{X}_n = \{X_1, \dots, X_n\}$ be a sequence of p-dimensional dependent random vectors with mean zero, i.e., $\mathbb{E}(X_t) = 0$. Write $S_{n,x} = n^{-1/2} \sum_{t=1}^n X_t$. Denote the instantaneous covariance matrix of X_t at time point t by $\Sigma_t = \mathrm{Cov}(X_t)$ and the long-run covariance matrix of $\{X_t\}_{t=1}^n$ by $\Xi = \mathrm{Cov}(S_{n,x})$. Our main goal is to bound

$$\rho_n(\mathcal{A}) := \sup_{A \in \mathcal{A}} |\mathbb{P}(S_{n,x} \in A) - \mathbb{P}(G \in A)|, \tag{1}$$

where $G \sim N(0, \Xi)$ and $\mathcal{A}$ is a class of Borel subsets in $\mathbb{R}^p$. Gaussian and parametric bootstrap approximation results over a rich index class $\mathcal{A}$ for large p are fundamental tools in developing downstream statistical inference procedures for a wide spectrum of problems in the high-dimensional setting, including, for example, inference of mean vector, change point detection, structure checking of instantaneous covariance matrix, and testing white noise hypothesis. We refer the readers to Section 4 for more details of these applications.

When $X_1, \ldots, X_n$ are independent random vectors in $\mathbb{R}^p$, the problem of bounding $\rho_n(\mathcal{A})$ for a variety of choices $\mathcal{A}$ is a classical research topic in probability theory (Bhattacharya and Rao, 2010, Petrov, 1995). Bounds on $\rho_n(\mathcal{A})$ under mild assumptions yield some useful statistics for a variety of high-dimensional inference problems. For instance, Chen and Qin (2010) and Cai, Liu and Xia (2014) studied the ℓ^2-type statistic and ℓ^∞-type statistic for testing high-dimensional mean vectors, respectively. Asymptotic validity of those test procedures relies on restrictive assumptions such as weak dependence in covariance matrix or sparsity in precision matrix. In contrast, Gaussian approximation results derived in this paper impose no explicit structural assumptions on the component-wise dependence structure, thus allowing to derive associated parametric bootstrap procedures for arbitrary dependence among different components of high-dimensional data. Recent years has witnessed a renewed interest in the accuracy of Gaussian approximation with explicit dependence on the dimension p since such results are particularly useful in modern large-scale statistical inference problems such as change point detection (Yu and Chen, 2021, 2022) and multiple testing for high-dimensional data (Chang et al., 2017a,b). For isotropic distributions with bounded third moments, Bentkus (2003) derived a Berry-Esseen type bound $O(p^{7/4}n^{-1/2})$ and $O(p^{3/2}n^{-1/2})$ over the class of convex subsets and Euclidean balls in $\mathbb{R}^p$, respectively. For independent (not necessarily identically distributed) sums, Chernozhukov, Chetverikov and Kato (2013) considered the problem of approximating maxima of $S_{n,x}$ by its Gaussian analogue and established an error bound that allows the dimension p to grow sub-exponentially fast in the sample size n.

Since the seminal work Chernozhukov, Chetverikov and Kato (2013), there have been substantial progresses being made in several directions. For instances, generalization of the index set from the max-rectangles to hyper-rectangles with improved rates of convergence can be found in Chernozhukov, Chetverikov and Kato (2017), Chernozhuokov, Chetverikov and Koike (2023), Das and Lahiri (2021), Deng (2020), Deng and Zhang (2020), Fang and Koike (2021), Koike (2021), Kuchibhotla and Rinaldo (2020), Lopes (2022), Lopes, Lin and Müller (2020) and Chernozhuokov et al. (2022a); extension from linear sums to U-statistics with nonlinear kernels can be found in Chen (2018), Chen and Kato (2019), Song, Chen and Kato (2019), and Koike (2023); generalization to dependent random vectors over max-rectangles can be found in Zhang and Cheng (2018), Zhang and Wu (2017), and Chernozhukov, Chetverikov and Kato (2019).

In the literature, some popular assumptions imposed on the temporal dependence of the sequence X_n include: (i) strong-mixing (or α-mixing) (Rosenblatt, 1956), (ii) m-dependent sequence (Hoeffding and Robbins, 1948), and (iii) physical (or functional) dependence measure for casual time series (Wu, 2005). Various CLTs for univariate (or fixed dimensional) dependent data have been developed under these dependence frameworks, see Bradley (2007), Doukhan, Massart and Rio (1994), Wu (2007), and Berkes, Liu and Wu (2014). We remark that there are many other mixing coefficients measuring the temporal dependence of the past and future, among which the α-mixing coefficient (see Definition 1 in Section 2.1.1) is the weakest one in the literature (Bradley, 2005). In particular, for $p = 1$, if the time series has finite third moment, the Komolgorov distance between the normalized random variable $\Xi^{-1/2}S_n$ and the standard univariate Gaussian distribution obeys a nearly optimal Berry-Esseen bound $O(n^{-1/2}\log^2 n)$ with geometrically decaying α-mixing coefficients (Sunklodas, 1984), or the sharp Berry-Esseen bound $O(n^{-1/2})$ for either m-dependent sequence with fixed m (Chen and Shao, 2004) or weakly dependent sequence under the physical dependence framework (Hörmann, 2009, Jirak, 2016).

Note that neither a dependence framework in (i)-(iii) implies the others. Thus there is a pressing call for a unified collection for Gaussian approximation tools under these temporal dependence frameworks for high dimensional dependent data.

Previous related works on high-dimensional CLTs for dependent data in the literature are complementary results for different dependence frameworks on max-rectangles, a subclass of hyper-rectangles. For examples, Zhang and Cheng (2018) studied the Gaussian approximation for m-dependent sequences with extension to dependent random vectors satisfying a geometric moment contraction condition (Wu and Shao, 2004); Zhang and Wu (2017) derived the Gaussian approximation result for causal stationary time series under a polynomial decay of the physical dependence measure; Chernozhukov, Chetverikov and Kato (2019) studied the validity of a block multiplier bootstrap under the β-mixing condition. All the aforementioned papers are only applicable to approximating the distributions of the ℓ^∞-type statistics and not applicable to approximating the distributions of some more general and complicated statistics involved in high-dimensional statistical inference. See Section 4 for details.

We conclude the introduction by summarizing our main contributions. Specifically, we develop a comprehensive and off-the-shelf probability toolbox containing the explicit rates of convergence of the high-dimensional CLTs for a combination of different index sets (including hyper-rectangles, simple convex sets, and sparsely convex sets) and different dependence frameworks (including α-mixing, m-dependent, and physical dependence measure). Our error bounds are non-asymptotic in all key parameters, including the sample size n and the data dimension p. In particular, our results established under the α-mixing framework are new in the literature, while results established under the physical dependence measure improve over existing results. In addition, we provide a parametric bootstrap procedure to implement the proposed results with a kernel-type estimator for the long-run covariance matrix. For both Gaussian and parametric bootstrap approximations, the data dimension p is allowed to grow sub-exponentially fast in the sample size n. The rest of the paper is organized as follows. Section 2 presents the error bounds of $\rho_n(\mathcal{A})$ defined as (1) with selecting $\mathcal{A}$ as hyper-rectangles, simple convex sets, and sparsely convex sets, respectively. Section 3 proposes a data-driven parametric bootstrap to approximate the probability $\mathbb{P}(S_{n,x} \in A)$ uniformly over $A \in \mathcal{A}$. Section 4 discusses how to implement the proposed results in several statistical inference problems of interest. Section 5 includes the proofs of high-dimensional CLTs on hyper-rectangles presented in Section 2.1, which provide the backbone for deriving high-dimensional CLTs on simple convex sets and sparsely convex sets stated, respectively, in Sections 2.2 and 2.3. The technical proofs of high-dimensional CLTs on simple convex sets and sparsely convex sets are given in the supplementary material Chang, Chen and Wu (2024).

2. High-dimensional central limit theorems

We define some notation first. For any positive integer m, we write $[m] := \{1,\ldots,m\}$. Denote by $I(\cdot)$ the indicator function. For two sequences of positive numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$ or $b_n \gtrsim a_n$ if there exists a universal constant $c > 0$ such that $\limsup_{n\to\infty} a_n/b_n \le c$. For any two p-dimensional vectors $v = (v_1,\ldots,v_p)^\top$ and $u = (u_1,\ldots,u_p)^\top$, $v \le u$ means that $v_j \le u_j$ for all $j \in [p]$. Given $\alpha > 0$, we define the function $\psi_\alpha(x) := \exp(x^\alpha) - 1$ for any $x > 0$. For a real-valued random variable ξ, we define $\|\xi\|_{\psi_\alpha} := \inf[\lambda > 0 : \mathbb{E}\{\psi_\alpha(|\xi|/\lambda)\} \le 1]$ and write $\xi \in \mathcal{L}^q$ for some $q > 0$ if $\|\xi\|_q := \{\mathbb{E}(|\xi|^q)\}^{1/q} < \infty$. For a thricely differentiable function $f : \mathbb{R}^p \to \mathbb{R}$, we write $\partial_j f(x) = \partial f(x)/\partial x_j$, $\partial_{jk} f(x) = \partial^2 f(x)/\partial x_j \partial x_k$ and $\partial_{jkl} f(x) = \partial^3 f(x)/\partial x_j \partial x_k \partial x_l$ for any $j,k,l \in [p]$. For a $q_1 \times q_2$ matrix $B = (b_{i,j})_{q_1\times q_2}$, let $|B|_\infty = \max_{i\in[q_1],j\in[q_2]} |b_{i,j}|$ be the super-norm, and $\|B\|_2 = \lambda_{\max}^{1/2}(BB^\top)$ be the spectral norm. Specifically, if $q_2 = 1$, we use $|B|_0 = \sum_{i=1}^{q_1} I(b_{i,1} \ne 0)$, $|B|_1 = \sum_{i=1}^{q_1} |b_{i,1}|$ and

$|B|_2 = (\sum_{i=1}^{q_1} b_{i,1}^2)^{1/2}$ to denote the ℓ^0-norm, ℓ^1-norm and ℓ^2-norm of the q_1-dimensional vector B, respectively.

Recall $S_{n,x} = n^{-1/2}\sum_{t=1}^n X_t$ and $\Xi = \mathrm{Cov}(S_{n,x})$. Let $G \sim N(0,\Xi)$ which is independent of $\mathcal{X}_n = \{X_1,\ldots,X_n\}$. We will first consider in Section 2.1 the upper bounds for

$$\varrho_n := \sup_{u\in\mathbb{R}^p,\nu\in[0,1]} |\mathbb{P}(\sqrt{\nu}S_{n,x} + \sqrt{1-\nu}G \le u) - \mathbb{P}(G \le u)| \tag{2}$$

when $\{X_t\}$ is (i) an α-mixing sequence, (ii) an m-dependent sequence, and (iii) a physical dependence sequence, respectively. Based on such derived upper bounds, we can easily translate them to the upper bounds for $\rho_n(\mathcal{A})$ when $\mathcal{A}$ is selected as the class of all hyper-rectangles in $\mathbb{R}^p$. In Sections 2.2 and 2.3, we will consider the upper bounds for $\rho_n(\mathcal{A})$ when $\mathcal{A}$ is selected as the class of simple convex sets and s-sparsely convex sets, respectively. Write $X_t = (X_{t,1},\ldots,X_{t,p})^\top$. Throughout the rest of this paper (unless otherwise explicitly stated), we shall focus on the high-dimensional scenario by assuming that $p \ge n^\kappa$ for some universal constant $\kappa > 0$. Here $\kappa > 0$ can be selected as some sufficiently small constant. Assuming $p \ge n^\kappa$ is a quite mild condition in the literature of high-dimensional data analysis which is not necessary for our theoretical analysis and just used to simplify our presentation. In our theoretical proofs, we need to compare $\log p$ and $\log n$ in lots of places. Without the restriction $p \ge n^\kappa$, some $\log p$ terms in the theoretical results should be replaced by $\log(pn)$.

2.1. High-dimensional CLT for hyper-rectangles

Let $\mathcal{A}^{\mathrm{re}}$ be the class of all hyper-rectangles in $\mathbb{R}^p$; that is, $\mathcal{A}^{\mathrm{re}}$ consists of all sets A of the form $A = \{(w_1,\ldots,w_p)^\top \in \mathbb{R}^p : a_j \le w_j \le b_j \text{ for all } j \in [p]\}$ with some $-\infty \le a_j \le b_j \le \infty$. Define $S_{n,\check{x}} = n^{-1/2}\sum_{t=1}^n \check{X}_t$ with $\check{X}_t = (X_t^\top, -X_t^\top)^\top$ and let $\check{G} \sim N(0,\check{\Xi})$ with $\check{\Xi} = \mathrm{Cov}(n^{-1/2}\sum_{t=1}^n \check{X}_t)$. We then have

$$\rho_n(\mathcal{A}^{\mathrm{re}}) \le \sup_{u\in\mathbb{R}^{2p},\nu\in[0,1]} |\mathbb{P}(\sqrt{\nu}S_{n,\check{x}} + \sqrt{1-\nu}\check{G} \le u) - \mathbb{P}(\check{G} \le u)|,$$

where the term on the right-hand side is a $(2p)$-dimensional analogue of ϱ_n defined as (2) over one-sided hyper-rectangles. To derive the convergence rate of $\rho_n(\mathcal{A}^{\mathrm{re}})$, it suffices to consider that for ϱ_n.

2.1.1. *α-mixing sequence*

Definition 1 (α-mixing coefficient). Let $\{X_t\}$ be a random sequence. Denote by $\mathcal{F}_{-\infty}^u$ and $\mathcal{F}_u^\infty$ the σ-fields generated respectively by $\{X_t\}_{t\le u}$ and $\{X_t\}_{t\ge u}$. The *α-mixing coefficient* at lag k of the sequence $\{X_t\}$ is defined as

$$\alpha_n(k) := \sup_t \sup_{A\in\mathcal{F}_{-\infty}^t, B\in\mathcal{F}_{t+k}^\infty} |\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|.$$

We say the sequence $\{X_t\}$ is *α-mixing* if $\alpha_n(k) \to 0$ as $k \to \infty$.

The *long-run variance* of the j-th coordinate marginal sequence $\{X_{t,j}\}_{t=1}^n$ is defined as

$$V_{n,j} = \mathrm{Var}\bigg(\frac{1}{\sqrt{n}}\sum_{t=1}^n X_{t,j}\bigg). \tag{3}$$

To investigate the convergence rate of ϱ_n defined as (2) for the α-mixing sequence $\{X_t\}$, we need the following regularity conditions.

Condition 1 (Sub-exponential moment). There exist a sequence of constants $B_n \geq 1$ and a universal constant $\gamma_1 \geq 1$ such that $\|X_{t,j}\|_{\psi_{\gamma_1}} \leq B_n$ for all $t \in [n]$ and $j \in [p]$.

Condition 2 (Decay of α-mixing coefficients). There exist some universal constants $K_1 > 1$, $K_2 > 0$ and $\gamma_2 > 0$ such that $\alpha_n(k) \leq K_1 \exp(-K_2 k^{\gamma_2})$ for any $k \geq 1$.

Condition 3 (Non-degeneracy). There exists a universal constant $K_3 > 0$ such that $\min_{j \in [p]} V_{n,j} \geq K_3$.

Since Condition 1 implies that $\mathbb{E}\{\exp(|X_{t,j}|^{\gamma_1} B_n^{-\gamma_1})\} \leq 2$, it follows from Markov's inequality that $\mathbb{P}(|X_{t,j}| > u) \leq 2\exp(-u^{\gamma_1} B_n^{-\gamma_1})$ for all $u > 0$. If each $X_{t,j}$ is sub-gaussian, we have $\gamma_1 = 2$ and $B_n = O(1)$. On one hand, the sub-exponential moment condition is a widely used condition in the high-dimensional statistics literature as this would generally entail vanishing rates of convergence of the sample mean for mean-zero independent data when the dimension p scales sub-exponentially fast in the sample size n (Wainwright, 2019). On the other hand, it can be relaxed to polynomial moment condition (e.g., 3rd moment condition). The trade-off is that under such condition, we can only expect the dimension p scales polynomially fast in the sample size n (i.e., $p = O(n^c)$ for some constant $c > 0$) to obtain vanishing rates of the sample mean for independent data. We can certainly expect that similar rates can be established for temporally dependent data. However, the proof techniques would be similar to the sub-exponential moment case. Moreover, if we replace the sample mean by its self-normalized version, then in the independent data case the studentized mean still has exponential decay tail under 3rd moment condition. It would be an interesting future work to investigate the self-normalization in the high-dimensional time series setting for the Gaussian approximation.

The α-mixing assumption is mild in the literature. Causal ARMA processes with continuous innovation distributions are α-mixing with exponential decay rates. So are stationary Markov chains satisfying certain conditions. See Section 2.6.1 of Fan and Yao (2003) and references within. In fact stationary GARCH models with finite second moments and continuous innovation distributions are also α-mixing with exponential decay rates. Under certain conditions, VAR processes, multivariate ARCH processes, and multivariate GARCH processes are all α-mixing with exponential decay rates; see Boussama, Fuchs and Stelzer (2011), Hafner and Preminger (2009) and Wong, Li and Tewari (2020). The next two examples also satisfy Condition 2.

- Let $X_t = A_t f_t + \varepsilon_t$, where A_t is a nonrandom loading matrix, $f_t \in \mathbb{R}^r$ is the latent factor with some fixed integer r, and $\{\varepsilon_t\}$ is an independent sequence that is also independent of $\{f_t\}$. If $\{f_t\}$ is selected as VAR processes, multivariate ARCH processes, or multivariate GARCH processes, due to that r is fixed and $\{\varepsilon_t\}$ is an independent sequence, we know such defined $\{X_t\}$ satisfies Condition 2 under certain conditions imposed on the model of $\{f_t\}$.
- Let $X_t = g_t(U_t)$, where $\{U_t\}$ is a q-dimensional latent sequence, and $g_t(\cdot): \mathbb{R}^q \to \mathbb{R}^p$ is a Borel function. Here we do not impose any relationship between p and q, and allow $q = \infty$. Write $U_t = (U_{t,1}, \ldots, U_{t,q})^\top$. Assume the sequence $\{U_{t,j}\}$ is ρ-mixing with exponential decay rates for each $j \in [q]$. If $\{U_{t,1}\}, \ldots, \{U_{t,q}\}$ are independent of each other, Theorem 5.1 of Bradley (2005) implies $\{U_t\}$ is ρ-mixing with exponential decay rates. Due to the relationship between ρ-mixing and α-mixing, we know such defined $\{X_t\}$ satisfies Condition 2.

Condition 3 assumes the partial sum $n^{-1/2} \sum_{t=1}^n X_{t,j}$ is non-degenerated which is required when we apply Nazarov's inequality (Chernozhukov, Chetverikov and Kato, 2017, Lemma A.1) to bound the probability of a Gaussian vector taking values in a small region. When $\{X_{t,j}\}_{t \geq 1}$ is stationary, we know $V_{n,j} = \Gamma_j(0) + 2\sum_{k=1}^{n-1}(1 - kn^{-1})\Gamma_j(k)$, where $\Gamma_j(k) = \mathrm{Cov}(X_{1,j}, X_{1+k,j})$ is the autocovariance of $\{X_{t,j}\}_{t \geq 1}$ at lag k. If each component sequence $\{X_{t,j}\}$ is stationary, Condition 3 holds if $\Gamma_j(0) + 2\sum_{k=1}^{\infty} \Gamma_j(k) \geq C$ holds for any $j \in [p]$, where $C > 0$ is a universal constant. Based on Conditions 1–3, Theorem 1 gives an upper bound for ϱ_n when the underlying sequence $\{X_t\}$ is α-mixing, whose proof is given in Section 5.1.

Theorem 1 (Gaussian approximation for partial sums of the α-mixing sequence). *Assume $\{X_t\}$ is an α-mixing sequence with $p \geq n^{\kappa}$ for some universal constant $\kappa > 0$. Under Conditions 1–3, it holds that*

$$\varrho_n \lesssim \frac{B_n^{2/3}(\log p)^{(1+2\gamma_2)/(3\gamma_2)}}{n^{1/9}} + \frac{B_n(\log p)^{7/6}}{n^{1/9}}$$

provided that $(\log p)^{3-\gamma_2} = o(n^{\gamma_2/3})$.

Remark 1 (Comparison with existing results under mixing dependence measure). Appendix B of Chernozhukov, Chetverikov and Kato (2019) derived the validity of a block multiplier bootstrap (BMB) under the β-mixing assumption. There are several differences between our Theorem 1 and Theorem B.1 in Chernozhukov, Chetverikov and Kato (2019). First, since the β-mixing assumption implies the α-mixing assumption, our Theorem 1 is applicable for wider class of dependent data. Second, Theorem B.1 in Chernozhukov, Chetverikov and Kato (2019) is proved and stated with the "large-and-small-blocks" argument, where conditions of their Theorem B.1 involve the "tuning parameter" of the block sizes. It is empirically known that the performance of BMB is sensitive to the block sizes. Although the BMB procedure given in Chernozhukov, Chetverikov and Kato (2019) is theoretically valid with suitable divergence rates imposed on the block sizes, how to propose a valid data-driven procedure to select the two involved tuning parameters is unclear in the framework of Gaussian approximation. Thus, it is an undesirable feature of Theorem B.1 in Chernozhukov, Chetverikov and Kato (2019) to rule out bootstraps without a hard truncation block size to estimate the long-run covariance matrices. In Section 3, we consider the kernel-type estimator of Andrews (1991) to estimate the long-run covariance matrix of $S_{n,x}$, which is more appealing from a practical standview (e.g., with the optimal quadratic spectral kernel and optimal data-driven bandwidth formula). Although the optimal data-driven bandwidth is obtained in the fixed dimensional scenario, extensive numerical studies in Chang, Jiang and Shao (2023) indicate that such formula still works well in high-dimensional setting and the associated performance is quite robust when the bandwidth is selected in a large range. Third, result from Chernozhukov, Chetverikov and Kato (2019) holds only for max-norm statistics, while our paper derives the convergence rates of Gaussian and parametric bootstrap approximations under much broader classes of index sets for high-dimensional dependent data (see Sections 2.2 and 2.3 below) that can be applied to approximate the distributions of more general and complicated statistics used in high-dimensional statistical inference.

Remark 2. Theorem 1 extends the Gaussian approximation result for independent data in Chernozhukov, Chetverikov and Kato (2017) to dependent data. When the eigenvalues of Ξ are bounded below from zero (i.e., strongly non-degenerate case), Chernozhuokov, Chetverikov and Koike (2023) derived a nearly optimal rate of convergence for independent data. Our analysis can be adapted with the sharper results from Chernozhuokov, Chetverikov and Koike (2023) to yield an improved error bound in Theorem 1 under stronger conditions.

2.1.2. *m-dependent sequence*

Based on the temporal dependency among $\{X_t\}_{t=1}^n$, we can define an undirected graph $G_n = (V_n, E_n)$, where $V_n = [n]$ is a set of nodes with node t denoting X_t, and E_n is a set of undirected edges connecting the nodes such that X_t and X_s are independent whenever $(t,s) \notin E_n$. Here we adopt the convention

$(t,t) \in E_n$ for any $t \in V_n$. We call such defined G_n is the *dependency graph* of the sequence $\{X_t\}_{t=1}^n$. The dependency graph is a flexible model to study CLTs with increasing dependence strength (Baldi and Rinott, 1989) that covers the m-dependent sequence as a special case. For any $t \in [n]$, let $\mathcal{N}_t = \{s \in V_n : (t,s) \in E_n\}$ be the neighbor nodes of node t in G_n. Define $D_n = \max_{t\in[n]} \sum_{s=1}^n I\{(t,s) \in E_n\}$ as the maximum degree of the first-degree connections in G_n, and $D_n^* = \max_{t\in[n]} \sum_{s=1}^n I\{s \in \cup_{\ell\in\mathcal{N}_t} \mathcal{N}_\ell\}$ as the maximum degree of the second-degree connections in G_n. Theorem 2 gives an upper bound for ϱ_n defined as (2) based on the maximum degrees D_n and D_n^* of the dependency graph determined by the underlying sequence $\{X_t\}_{t=1}^n$, whose proof is given in Section 5.2.

Theorem 2 (Gaussian approximation for partial sums of a sequence under dependency graph). *Assume $p \geq n^\kappa$ for some universal constant $\kappa > 0$. Under Conditions 1 and 3, it holds that*

$$\varrho_n \lesssim \frac{B_n(D_nD_n^*)^{1/3}(\log p)^{7/6}}{n^{1/6}},$$

where D_n and D_n^* are the maximum degrees of the first-degree and second-degree of connections in the dependency graph generated by the sequence $\{X_t\}_{t=1}^n$, respectively.

If $\{X_t\}_{t=1}^n$ is a centered m-dependent sequence, i.e., X_t and X_s are independent for all $|t-s| > m$, then $\{X_t\}_{t=1}^n$ has a dependency graph with $D_n = 2m+1$ and $D_n^* = 4m+1$. The next corollary states a result for m-dependent sequences.

Corollary 1 (Gaussian approximation for partial sums of an m-dependent sequence). *Assume $\{X_t\}_{t=1}^n$ is an m-dependent sequence with $p \geq n^\kappa$ for some universal constant $\kappa > 0$. Under Conditions 1 and 3, it holds that*

$$\varrho_n \lesssim \frac{B_n(m\vee 1)^{2/3}(\log p)^{7/6}}{n^{1/6}}.$$

Since the 0-dependent sequence reduces to the independent sequence, Corollary 1 for $m = 0$ reads $O(B_nn^{-1/6}\log^{7/6} p)$, which has the same sample complexity in n and dimension dependence in p as the independent data case $O(B_n^{1/3}n^{-1/6}\log^{7/6} p)$ up to a moment factor $B_n^{2/3}$ (cf. Proposition 2.1 in Chernozhukov, Chetverikov and Kato (2017)). The extra cost $B_n^{2/3}$ is due to the argument that we need to decouple the distribution tail and dependence simultaneously. In particular, for the data with $B_n = O(1)$, the rate obtained from our m-dependent CLT achieves the CLT rate for independent data derived in Chernozhukov, Chetverikov and Kato (2017).

Corollary 1 is a stepping stone to study the Gaussian approximation under the physical dependence framework with better rate of convergence than the best known results in Zhang and Wu (2017) based on the large-and-small-blocks technique in the weaker temporal dependence regime. See Theorem 3 and the discussions in Section 2.1.4 for more details.

2.1.3. *Sequence with physical dependence*

Let $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ be a sequence of independent and identically distributed random elements. Consider the (causal) time series model

$$X_t = f_t(\varepsilon_t, \varepsilon_{t-1}, \dots), \quad t \geq 1, \tag{4}$$

where $f_t(\cdot)$ is a jointly measurable function taking values in $\mathbb{R}^p$ and $\mathbb{E}(X_t) = 0$. Here $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ are *innovations* that can be viewed as the input of the non-linear system (4). Since the data generation mechanism

$f_t(\cdot)$ may change over time, X_t is allowed to be *non-stationary*. Non-linear time series of the form (4) are first introduced in Wu (2005) for $p = 1$ and $f_t(\cdot) \equiv f(\cdot)$ for some measurable function $f(\cdot)$ (i.e., stationary univariate time series), and their temporal dependence can be quantified by the *functional dependence measure* based on the idea of coupling. In particular, let ε_i' be an independent copy of ε_i and

$$X'_{t,\{m\}} = f_t(\varepsilon_t, \dots, \varepsilon_{t-m+1}, \varepsilon'_{t-m}, \varepsilon_{t-m-1}, \dots)$$

be the coupled version of X_t at the time lag m with ε_{t-m} replaced by ε'_{t-m}. By causality, $X'_{t,\{m\}} = X_t$ for $m < 0$. Write $X'_{t,\{m\}} = (X'_{t,1,\{m\}}, \dots, X'_{t,p,\{m\}})^\top$. The (uniform) functional dependence measure for the j-th coordinate marginal sequence $\{X_{t,j}\}$ is defined as

$$\theta_{m,q,j} = \sup_{t \ge 1} \|X_{t,j} - X'_{t,j,\{m\}}\|_q , \quad q > 0 .$$

In essence, $\theta_{m,q,j}$ quantifies the uniform impact of coupling on the j-th coordinate marginal time series at lag m. For any $m \ge 0$, write $\Theta_{m,q,j} = \sum_{i=m}^{\infty} \theta_{i,q,j}$. For $\alpha \in (0, \infty)$, define the *dependence adjusted norm* introduced in Wu and Wu (2016) as

$$\|X_{\cdot,j}\|_{q,\alpha} = \sup_{m \ge 0} (m+1)^{\alpha} \Theta_{m,q,j} \quad \text{and} \quad \|X_{\cdot,j}\|_{\psi_\nu,\alpha} = \sup_{q \ge 2} q^{-\nu} \|X_{\cdot,j}\|_{q,\alpha} ,$$

whenever the supremums are finite. Define further the aggregated norms as follows:

$$\Psi_{q,\alpha} = \max_{j \in [p]} \|X_{\cdot,j}\|_{q,\alpha} \quad \text{and} \quad \Phi_{\psi_\nu,\alpha} = \max_{j \in [p]} \|X_{\cdot,j}\|_{\psi_\nu,\alpha} . \tag{5}$$

Theorem 3 (Gaussian approximation for maxima of partial sums of time series under functional dependence). *Assume $\{X_t\}$ satisfies the model (4) with $p \ge n^{\kappa}$ for some universal constant $\kappa > 0$. Let $\Phi_{\psi_\nu,\alpha} < \infty$ for some $\alpha, \nu \in (0, \infty)$.*

(i) *Under Condition 3, it holds that*

$$\varrho_n \lesssim \frac{\Phi_{\psi_\nu,0} (\log p)^{7/6}}{n^{\alpha/(3+9\alpha)}} + \frac{\Psi_{2,\alpha}^{1/3} \Psi_{2,0}^{1/3} (\log p)^{2/3}}{n^{\alpha/(3+9\alpha)}} + \frac{\Phi_{\psi_\nu,\alpha} (\log p)^{1+\nu}}{n^{\alpha/(1+3\alpha)}}$$

provided that $(\log p)^{\max\{6\nu-1,(5+6\nu)/4\}} = o\{n^{\alpha/(1+3\alpha)}\}$.

(ii) *Under Conditions 1 and 3, it holds that*

$$\varrho_n \lesssim \frac{B_n (\log p)^{7/6}}{n^{\alpha/(12+6\alpha)}} + \frac{\Psi_{2,\alpha}^{1/3} \Psi_{2,0}^{1/3} (\log p)^{2/3}}{n^{\alpha/(12+6\alpha)}} + \frac{\Phi_{\psi_\nu,\alpha} (\log p)^{1+\nu}}{n^{\alpha/(4+2\alpha)}} .$$

The proof of Theorem 3 is given in Section 5.3. We remark that the two rates of convergence given in Theorems 3(i) and 3(ii) are based on the *large-and-small-blocks* and m-dependent approximation techniques, respectively. The large-and-small-blocks technique is widely used in time series analysis to approximate the sum of a time series sequence by the sum over its large blocks. It is interesting to note that the large-and-small-blocks technique gives a faster (or slower) rate than the m-dependent argument when $0 < \alpha < 3$ (or $\alpha > 3$). In particular, when the temporal dependence is weak (for large values of α), the improvement of Theorem 3(ii) than Theorem 3(i) is more significant. The intuition is that the large-and-small-blocks technique used to establish Theorem 3(i) may lose sample size efficiency when the temporal dependence is weak. In such regime, throwing away the data in small blocks may reduce the *effective* sample size, while the m-dependent approximation directly approximates the sequence $\mathcal{X}_n$ without throwing away data. On the other hand, when the temporal dependence is strong (for small values of α), we need to use much larger values of m for constructing an m-dependent sequence in Section 2.1.2, so the m-dependent approximation becomes less effective than throwing a reasonable amount of small blocks to reduce the dependence.

Now combining the two parts of Theorem 3, we obtain the overall rate of convergence under the physical dependence measure.

Corollary 2 (Overall rate of convergence under physical dependence). *Assume $\{X_t\}$ satisfies the model (4) with $p \ge n^{\kappa}$ for some universal constant $\kappa > 0$, and $\Phi_{\psi_\nu,\alpha} < \infty$ for some $\alpha, \nu \in (0,\infty)$. Let $\alpha' = \alpha/\min\{1+3\alpha, 4+2\alpha\}$. Under Conditions 1 and 3, it holds that*

$$\varrho_n \lesssim \frac{\max\{\Phi_{\psi_\nu,0}, B_n\}(\log p)^{7/6}}{n^{\alpha'/3}} + \frac{\Psi_{2,\alpha}^{1/3}\Psi_{2,0}^{1/3}(\log p)^{2/3}}{n^{\alpha'/3}} + \frac{\Phi_{\psi_\nu,\alpha}(\log p)^{1+\nu}}{n^{\alpha'}}$$

provided that $(\log p)^{\max\{6\nu-1,(5+6\nu)/4\}} = o\{n^{\alpha/(1+3\alpha)}\}$.

2.1.4. *Comparison with existing result under physical dependence measure*

Under the physical dependence and a sub-exponential moment condition, Zhang and Wu (2017) derived a Gaussian approximation result for the ℓ^∞-norm of normalized sums of a class of stationary time series:

$$\omega_n := \sup_{u\ge 0} |\mathbb{P}(|D^{-1}S_{n,x}|_\infty \ge u) - \mathbb{P}(|D^{-1}G|_\infty \ge u)|,$$

where $D = \{\mathrm{diag}(\Xi)\}^{1/2}$ and $G \sim N(0,\Xi)$. Specifically, Theorem 7.4 in Zhang and Wu (2017) gives the following error bound: for any $\lambda \in (0,1)$ and $\eta > 0$,

$$\omega_n \lesssim f^\diamond(\sqrt{n}\eta) + \eta(\log p)^{1/2} + h\{\lambda, u_m^\diamond(\lambda)\} + \pi\{\chi(m,M)\} \tag{6}$$

with

$$\begin{aligned} f^\diamond(y) &= p\exp(-C_\beta y^\beta m^{\alpha\beta} n^{-\beta/2}\Phi_{\psi_\nu,\alpha}^{-\beta}) + p\exp\{-C_\beta y^\beta (mw)^{-\beta/2}\Phi_{\psi_\nu,0}^{-\beta}\}, \\ h\{\lambda, u_m^\diamond(\lambda)\} &= \lambda + w^{-1/8}\max\{\Psi_{3,0}^{3/4}, \Psi_{4,0}^{1/2}\}\log^{7/8}(pw\lambda^{-1}) \\ &\quad + w^{-1/2}\max\{\Phi_{\psi_\nu,0}\log^{1/\beta}(pw\lambda^{-1}), \log^{1/2}(pw\lambda^{-1})\}\log^{3/2}(pw\lambda^{-1}), \\ \pi(x) &= x^{1/3}\max\{1, \log^{2/3}(px^{-1})\}, \\ \chi(m,M) &= \Psi_{2,\alpha}\Psi_{2,0}\{m^{-\alpha} + v(M)\} + wmn^{-1}, \end{aligned}$$

where $\beta = 2/(1+2\nu)$, $v(M) = M^{-1}I(\alpha > 1) + (M^{-1}\log M)I(\alpha = 1) + M^{-\alpha}I(0 < \alpha < 1)$, and (m, M, w) are tuning parameters involved in the "large and small blocks" technique for deriving (6) with m and M being, respectively, the sizes of small blocks and large blocks satisfying $m = o(M)$, and $w = \lfloor n/(M+m)\rfloor$. They first approximated $S_{n,x}$ by the sum of an m-dependent sequence, and then applied the large-and-small-blocks technique to approximate the sum of the m-dependent sequence by the sum over large blocks.

To simplify the convergence rate of ω_n specified in (6), we assume $\Phi_{\psi_\nu,\alpha} = O(1)$ for some $\alpha, \nu \in (0,\infty)$ and $p \ge n^\kappa$ for some $\kappa > 0$. Choose $\lambda = n^{-c_1}, \eta = n^{-c_2}$, $w \asymp n^{c_3}$ and $m \asymp n^{c_4}$ for some constants $c_1,c_2,c_3,c_4 > 0$. By optimizing (c_1,c_2,c_3,c_4) according to the right-hand side of (6), we have the following proposition whose proof is given in Section S4 of the supplementary material.

Proposition 1 (Rate of convergence under physical dependence in Zhang and Wu (2017)). *Assume* $\Phi_{\psi_\nu,\alpha} = O(1)$ *for some* $\alpha,\nu \in (0,\infty)$ *and* $p \geq n^\kappa$ *for some* $\kappa > 0$. *Then the upper bound of* ω_n *given in* (6) *can be simplified as*

$$\omega_n \lesssim \frac{\mathrm{polylog}(p)}{n^{\alpha/(3+11\alpha)}},$$

where $\mathrm{polylog}(p)$ *is a polynomial factor of* $\log p$.

Contrasting Proposition 1 with Corollary 2 specialized to max-rectangles, we see that, up to a $\mathrm{polylog}(p)$ factor, our rate of convergence reads $\mathrm{polylog}(p)\cdot n^{-\alpha/[3\min\{1+3\alpha,4+2\alpha\}]}$, which is uniformly faster than that given in Proposition 1 for all $\alpha > 0$. In other words, our rate has a better sample size dependence than Zhang and Wu (2017). The reason can be seen that the optimal choice of Zhang and Wu (2017) throws away $w \asymp n^{8\alpha/(3+11\alpha)}$ small blocks of size $m \asymp n^{3/(3+11\alpha)}$, which leads a total reduction of $O\{n^{(3+8\alpha)/(3+11\alpha)}\}$ data points in the sample size. In our result, we only throw away $w \asymp n^{2\alpha/(1+3\alpha)}$ small blocks of size $m \asymp n^{1/(1+3\alpha)}$, leading to a total reduction of $O\{n^{(1+2\alpha)/(1+3\alpha)}\}$ data points in the sample size. Moreover, the improvement of our result over Zhang and Wu (2017) is more significant for larger values of $\alpha > 3$.

2.2. High-dimensional CLT for simple convex sets

In this section, we consider the class of *simple convex sets* introduced by Chernozhukov, Chetverikov and Kato (2017). Formally, a simple convex set can be well approximated by a convex polytope with a controlled number of facets. Simple convex sets serve an important intermediate step to derive similar error bounds in Gaussian approximation on the class of s-sparsely convex sets considered in Section 2.3. Geometrically, s-sparsely convex sets can be represented as an intersection of possibly many convex sets whose indicator functions depend at most on s elements of their coordinates.

For a closed convex set $A \subset \mathbb{R}^p$, we define its support function:

$$\mathcal{S}_A : \mathbb{S}^{p-1} \mapsto \mathbb{R}\cup\{\infty\}, \qquad v \mapsto \mathcal{S}_A(v) := \sup\{w^\top v : w \in A\},$$

where $\mathbb{S}^{p-1}$ is the unit sphere in $\mathbb{R}^p$. Specially, if A^K is K-*generated* (that is, A^K is generated by the intersection of K half-spaces), we could characterize A^K by its support function for the set $\mathcal{V}(A^K)$ consisting K unit normal vectors outward to the facets of A^K:

$$A^K = \bigcap_{v\in\mathcal{V}(A^K)} \{w \in \mathbb{R}^p : w^\top v \leq \mathcal{S}_{A^K}(v)\}.$$

Moreover, for $\epsilon > 0$ and a K-generated convex set A^K, we also define

$$A^{K,\epsilon} = \bigcap_{v\in\mathcal{V}(A^K)} \{w \in \mathbb{R}^p : w^\top v \leq \mathcal{S}_{A^K}(v) + \epsilon\}.$$

Definition 2 (Simple convex set). We say A is a *simple convex set*, if there exist two constants $a \geq 0$, $d > 0$ and an K-*generated* A^K satisfying $K \leq (pn)^d$ such that

$$A^K \subset A \subset A^{K,\epsilon} \tag{7}$$

with $\epsilon = a/n$. In this case, A^K provides an approximation to A with precision ϵ.

Let $\mathcal{A}^{\mathrm{si}}(a,d)$ be the class of all sets A satisfying (7) with $K \le (pn)^d$ and $\epsilon = a/n$. For any $v \in \mathbb{R}^p$, define

$$V_n(v) = \mathrm{Var}\bigg(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} v^\top X_t\bigg). \tag{8}$$

In the sequel, we shall slightly abuse the notation by using $A^K(\cdot)$ to also denote the operator defined on $\mathcal{A}^{\mathrm{si}}(a,d)$ such that $A^K(A) = A^K$ with A^K specified in (7) for any $A \in \mathcal{A}^{\mathrm{si}}(a,d)$. To construct the upper bounds for $\rho_n(\mathcal{A})$ for some $\mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a,d)$, we need the following condition that imposes the moment assumption on $v^\top X_t$ for $v \in \mathcal{V}\{A^K(A)\}$.

Condition 4. There exist a sequence of constants $B_n \ge 1$ and a universal constant $\gamma_1 \ge 1$ such that $\|v^\top X_t\|_{\psi_{\gamma_1}} \le B_n$ for all $t \in [n]$ and $v \in \mathcal{V}\{A^K(A)\}$.

2.2.1. *α-mixing sequence*

To obtain an upper bound for $\rho_n(\mathcal{A})$ for some $\mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a,d)$, we need the next condition that requires the long-run variance of the sequence $\{v^\top X_t\}_{t=1}^n$ is not degenerated for any $v \in \mathcal{V}\{A^K(A)\}$.

Condition 5. There exists a universal constant $K_4 > 0$ such that $V_n(v) \ge K_4$ for any $v \in \mathcal{V}\{A^K(A)\}$.

Condition 5 holds automatically if the smallest eigenvalue of $\Xi = \mathrm{Cov}(n^{-1/2}\sum_{t=1}^n X_t)$ is uniformly bounded away from zero.

Theorem 4 (Gaussian approximation for partial sums of the α-mixing sequence for simple convex sets). *Assume $\{X_t\}$ is an α-mixing sequence with $p \ge n^\kappa$ for some universal constant $\kappa > 0$ and Condition 2 being satisfied. Let $\mathcal{A}$ be a subclass of $\mathcal{A}^{\mathrm{si}}(a,d)$ such that Conditions 4 and 5 are satisfied for any $A \in \mathcal{A}$. Then*

$$\rho_n(\mathcal{A}) \lesssim \frac{a(d\log p)^{1/2}}{n} + \frac{B_n^{2/3}(d\log p)^{(1+2\gamma_2)/(3\gamma_2)}}{n^{1/9}} + \frac{B_n(d\log p)^{7/6}}{n^{1/9}}$$

provided that $(d\log p)^{3-\gamma_2} = o(n^{\gamma_2/3})$.

The proof of Theorem 4 is given in Section S5 of the supplementary material.

2.2.2. *m-dependent sequence*

Theorem 5 gives the high-dimensional CLT for simple convex sets based on the maximum degrees D_n and D_n^* of the dependency graph determined by the underlying sequence $\{X_t\}_{t=1}^n$, whose proof is given in Section S6 of the supplementary material. See the definitions of dependency graph and its associated maximum degrees in Section 2.1.2.

Theorem 5 (Gaussian approximation for partial sums of a sequence under dependency graph for simple convex sets). *Assume $p \ge n^\kappa$ for some universal constant $\kappa > 0$. Let $\mathcal{A}$ be a subclass of $\mathcal{A}^{\mathrm{si}}(a,d)$ such that Conditions 4 and 5 are satisfied for any $A \in \mathcal{A}$. Then*

$$\rho_n(\mathcal{A}) \lesssim \frac{a(d\log p)^{1/2}}{n} + \frac{B_n(D_nD_n^*)^{1/3}(d\log p)^{7/6}}{n^{1/6}},$$

where D_n and D_n^* are the maximum degrees of the first-degree and second-degree of connections in the dependency graph generated by the sequence $\{X_t\}_{t=1}^n$, respectively.

The next corollary states a result for m-dependent sequences.

Corollary 3. *Assume $\{X_t\}_{t=1}^n$ is an m-dependent sequence with $p \geq n^\kappa$ for some universal constant $\kappa > 0$. Let $\mathcal{A}$ be a subclass of $\mathcal{A}^{\mathrm{si}}(a,d)$ such that Conditions 4 and 5 are satisfied for any $A \in \mathcal{A}$. Then*

$$\rho_n(\mathcal{A}) \lesssim \frac{a(d\log p)^{1/2}}{n} + \frac{B_n(m \vee 1)^{2/3}(d\log p)^{7/6}}{n^{1/6}}\,.$$

2.2.3. Sequence with physical dependence

For a K-generated convex set A^K, write $\mathcal{V}(A^K) = \{v_1,\ldots,v_K\}$. Given $\mathcal{V}(A^K)$ and $\{X_t\}$, we can define a new K-dimensional sequence $X_t(A^K) = (v_1^\top X_t,\ldots,v_K^\top X_t)^\top$. If $\{X_t\}_{t=1}^n$ satisfies model (4) with the jointly measurable function $f_t(\cdot)$, $\{X_t(A^K)\}$ also satisfies (4) with a jointly measurable function $\tilde{f}_t(\cdot) = \{v_1^\top f_t(\cdot),\ldots,v_K^\top f_t(\cdot)\}^\top$. We further define $\Psi_{q,\alpha}(A^K)$ and $\Phi_{\psi_\nu,\alpha}(A^K)$ for any K-generated convex set A^K in the same manner as $\Psi_{q,\alpha}$ and $\Phi_{\psi_\nu,\alpha}$ in (5) by replacing $\{X_t\}$ with $\{X_t(A^K)\}$. Given $\mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a,d)$, let

$$\Psi_{q,\alpha,\mathcal{A}} = \sup_{A\in\mathcal{A}} \Psi_{q,\alpha}\{A^K(A)\} \quad \text{and} \quad \Phi_{\psi_\nu,\alpha,\mathcal{A}} = \sup_{A\in\mathcal{A}} \Phi_{\psi_\nu,\alpha}\{A^K(A)\}\,. \tag{9}$$

Theorem 6 (Gaussian approximation for partial sums of time series under functional dependence for simple convex sets). *Assume the sequence $\{X_t\}$ satisfies the model (4) with $p \geq n^\kappa$ for some universal constant $\kappa > 0$. Let $\mathcal{A}$ be a subclass of $\mathcal{A}^{\mathrm{si}}(a,d)$ such that Condition 5 is satisfied for any $A \in \mathcal{A}$, and $\Phi_{\psi_\nu,\alpha,\mathcal{A}} < \infty$ for some $\alpha,\nu \in (0,\infty)$.*

(i) *It holds that*

$$\begin{aligned}\rho_n(\mathcal{A}) \lesssim{}& \frac{a(d\log p)^{1/2}}{n} + \frac{\Phi_{\psi_\nu,0,\mathcal{A}}(d\log p)^{7/6}}{n^{\alpha/(3+9\alpha)}}\\ &+ \frac{\Psi_{2,\alpha,\mathcal{A}}^{1/3}\Psi_{2,0,\mathcal{A}}^{1/3}(d\log p)^{2/3}}{n^{\alpha/(3+9\alpha)}} + \frac{\Phi_{\psi_\nu,\alpha,\mathcal{A}}(d\log p)^{1+\nu}}{n^{\alpha/(1+3\alpha)}}\end{aligned}$$

provided that $(d\log p)^{\max\{6\nu-1,(5+6\nu)/4\}} = o\{n^{\alpha/(1+3\alpha)}\}$.

(ii) *If Condition 4 is satisfied for any $A \in \mathcal{A}$, it holds that*

$$\begin{aligned}\rho_n(\mathcal{A}) \lesssim{}& \frac{a(d\log p)^{1/2}}{n} + \frac{B_n(d\log p)^{7/6}}{n^{\alpha/(12+6\alpha)}}\\ &+ \frac{\Psi_{2,\alpha,\mathcal{A}}^{1/3}\Psi_{2,0,\mathcal{A}}^{1/3}(d\log p)^{2/3}}{n^{\alpha/(12+6\alpha)}} + \frac{\Phi_{\psi_\nu,\alpha,\mathcal{A}}(d\log p)^{1+\nu}}{n^{\alpha/(4+2\alpha)}}\,.\end{aligned}$$

The proof of Theorem 6 is given in Section S7 of the supplementary material.

2.3. High-dimensional CLT for sparsely convex sets

We consider sparsely convex sets here, as a generalization of hyper-rectangles, that can be represented as intersections of convex sets whose indicator functions depend only on a small subset of their coordinates.

Definition 3 (s-sparsely convex set). For an integer $s > 0$, we say $A \subset \mathbb{R}^p$ is an *s-sparsely convex set* if (i) A admits a sparse representation $A = \cap_{q=1}^{K_*} A_q$ for some positive integer K_* and convex sets $A_1, \ldots, A_{K_*} \subset \mathbb{R}^p$, and (ii) the indicator function $I(w \in A_q)$ depends on at most s components of the vector $w \in \mathbb{R}^p$ (which we call the main components of A_q).

Denote by $\mathcal{A}^{\mathrm{sp}}(s)$ the class of all s-sparsely convex sets in $\mathbb{R}^p$. In this section, we target on deriving the upper bounds for $\rho_n\{\mathcal{A}^{\mathrm{sp}}(s)\}$ when the observed data $\{X_t\}$ are (i) an α-mixing sequence, (ii) an m-dependent sequence, and (iii) a physical dependence sequence.

2.3.1. α-mixing sequence

Condition 6. For $V_n(v)$ defined in (8), there exists a universal constant $K_5 > 0$ such that $V_n(v) \geq K_5$ for any $v \in \mathbb{S}^{p-1}$ with $|v|_0 \leq s$.

Condition 6 holds automatically if the smallest eigenvalue of $\Xi = \mathrm{Cov}(n^{-1/2} \sum_{t=1}^n X_t)$ is uniformly bounded away from zero.

Theorem 7 (Gaussian approximation for partial sums of the α-mixing sequence for s-sparsely convex sets). *Assume $\{X_t\}$ is an α-mixing sequence with $p \geq n^\kappa$ for some universal constant $\kappa > 0$. Under Conditions 1, 2 and 6, it holds that*

$$\rho_n\{\mathcal{A}^{\mathrm{sp}}(s)\} \lesssim \frac{B_n^{2/3} s^{(2+6\gamma_2)/(3\gamma_2)} (\log p)^{(1+2\gamma_2)/(3\gamma_2)}}{n^{1/9}} + \frac{B_n s^{10/3} (\log p)^{7/6}}{n^{1/9}}$$

provided that $(s^2 \log p)^{3-\gamma_2} = o(n^{\gamma_2/3})$.

The proof of Theorem 7 is given in Section S8 of the supplementary material.

2.3.2. m-dependent sequence

Theorem 8 (Gaussian approximation for partial sums of a sequence under dependency graph for s-sparsely convex sets). *Assume $p \geq n^\kappa$ for some universal constant $\kappa > 0$. Under Conditions 1 and 6, it holds that*

$$\rho_n\{\mathcal{A}^{\mathrm{sp}}(s)\} \lesssim \frac{s^{10/3} B_n (D_n D_n^*)^{1/3} (\log p)^{7/6}}{n^{1/6}},$$

where D_n and D_n^* are the maximum degrees of the first-degree and second-degree of connections in the dependency graph generated by the sequence $\{X_t\}_{t=1}^n$, respectively.

The proof of Theorem 8 is given in Section S10 of the supplementary material. The next corollary states a result for m-dependent sequences.

Corollary 4 (Gaussian approximation for partial sums of an m-dependent sequence for s-sparsely convex sets). *Assume $\{X_t\}_{t=1}^n$ is an m-dependent sequence with $p \geq n^\kappa$ for some universal constant $\kappa > 0$. Under Conditions 1 and 6, it holds that*

$$\rho_n\{\mathcal{A}^{\mathrm{sp}}(s)\} \lesssim \frac{s^{10/3} (m \vee 1)^{2/3} B_n (\log p)^{7/6}}{n^{1/6}}.$$

2.3.3. Sequence with physical dependence

For any $c > 0$, define

$$\Omega_{s,c} = \left\{ A \in \mathcal{A}^{\mathrm{si}}(1, cs^2) : \max_{v \in \mathcal{V}\{A^K(A)\}} |v|_0 \le s \right\}.$$

Analogous to $\Psi_{q,\alpha,\mathcal{A}}$ and $\Phi_{\psi_\nu,\alpha,\mathcal{A}}$ defined in (9), we also define

$$\Psi_{q,\alpha,\Omega_{s,c}} = \sup_{A \in \Omega_{s,c}} \Psi_{q,\alpha}\{A^K(A)\} \quad \text{and} \quad \Phi_{\psi_\nu,\alpha,\Omega_{s,c}} = \sup_{A \in \Omega_{s,c}} \Phi_{\psi_\nu,\alpha}\{A^K(A)\}$$

with $\Psi_{q,\alpha}(A^K)$ and $\Phi_{\psi_\nu,\alpha}(A^K)$ defined in Section 2.2.3. It then holds that $\Psi_{q,\alpha,\Omega_{s,c}} \ge \Psi_{q,\alpha}$ and $\Phi_{\psi_\nu,\alpha,\Omega_{s,c}} \ge \Phi_{\psi_\nu,\alpha}$.

Theorem 9 (Gaussian approximation for partial sums of time series under functional dependence for s-sparsely convex sets). *Assume the sequence $\{X_t\}$ satisfies the model (4) with $p \ge n^\kappa$ for some universal constant $\kappa > 0$. For some sufficiently large constant $c > 0$, let $\Phi_{\psi_\nu,\alpha,\Omega_{s,c}} < \infty$ with some $\alpha, \nu \in (0,\infty)$.*

(i) *Under Condition 6, it holds that*

$$\rho_n\{\mathcal{A}^{\mathrm{sp}}(s)\} \lesssim \frac{\Phi_{\psi_\nu,0,\Omega_{s,c}}(s^2 \log p)^{7/6}}{n^{\alpha/(3+9\alpha)}} + \frac{\Psi^{1/3}_{2,\alpha,\Omega_{s,c}} \Psi^{1/3}_{2,0,\Omega_{s,c}} (s^2 \log p)^{2/3}}{n^{\alpha/(3+9\alpha)}} + \frac{\Phi_{\psi_\nu,\alpha,\Omega_{s,c}}(s^2 \log p)^{1+\nu}}{n^{\alpha/(1+3\alpha)}} + \frac{\Phi^3_{\psi_\nu,0} s^{17/4}}{n^{\alpha/(2+4\alpha)}}$$

provided that $(s^2 \log p)^{\max\{6\nu-1,(5+6\nu)/4\}} = o\{n^{\alpha/(1+3\alpha)}\}$.

(ii) *Under Conditions 1 and 6, it holds that*

$$\rho_n\{\mathcal{A}^{\mathrm{sp}}(s)\} \lesssim \frac{s^{10/3} B_n (\log p)^{7/6}}{n^{\alpha/(12+6\alpha)}} + \frac{\Psi^{1/3}_{2,\alpha,\Omega_{s,c}} \Psi^{1/3}_{2,0,\Omega_{s,c}} (s^2 \log p)^{2/3}}{n^{\alpha/(12+6\alpha)}} + \frac{\Phi_{\psi_\nu,\alpha,\Omega_{s,c}}(s^2 \log p)^{1+\nu}}{n^{\alpha/(4+2\alpha)}}.$$

The proof of Theorem 9 is given in Section S11 of the supplementary material.

3. Parametric bootstrap

In Section 2, we have established the error bounds for $\rho_n(\mathcal{A})$ defined as (1) when $\{X_t\}$ is (i) an α-mixing sequence, (ii) an m-dependent sequence, and (iii) a physical dependence sequence. Since $\mathbb{P}(G \in A)$ with $G \sim N(0,\Xi)$ depends on the unknown long-run covariance matrix Ξ, to approximate $\mathbb{P}(S_{n,x} \in A)$ in practice, we need to construct a data-dependent Gaussian analogue $\hat{G}$ of G. In this section, we propose a parametric bootstrap procedure to construct $\hat{G} \sim N(0,\hat{\Xi}_n)$ for some covariance matrix $\hat{\Xi}_n$ that is close to Ξ and establish its theoretical validity. Define

$$\hat{\rho}_n(\mathcal{A}) := \sup_{A \in \mathcal{A}} |\mathbb{P}(S_{n,x} \in A) - \mathbb{P}(\hat{G} \in A \,|\, \mathcal{X}_n)|$$

with $\mathcal{X}_n = \{X_1, \ldots, X_n\}$. Let $\Delta_{n,r} = |\hat{\Xi}_n - \Xi|_\infty$ and

$$\Delta_n(\mathcal{A}) = \sup_{A \in \mathcal{A}} \sup_{v_1, v_2 \in \mathcal{V}\{A^K(A)\}} |v_1^\top(\hat{\Xi}_n - \Xi)v_2|$$

for any $\mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a,d)$. Theorem 10 establishes primitive error bounds for $\hat{\rho}_n(\mathcal{A})$ when $\mathcal{A}$ is selected as the class of all hyper-rectangles, the class of simple convex sets and the class of s-sparsely convex sets, respectively, whose proof is given in Section S13 of the supplementary material.

Theorem 10 (Rates of convergence for parametric bootstrap). *Assume $p \geq n^\kappa$ for some universal constant $\kappa > 0$.*

(i) *Under Condition 3, it holds that*

$$\hat{\rho}_n(\mathcal{A}^{\mathrm{re}}) \lesssim \rho_n(\mathcal{A}^{\mathrm{re}}) + \Delta_{n,r}^{1/3}(\log p)^{2/3}\,.$$

(ii) *Let $\mathcal{A}$ be a subclass of $\mathcal{A}^{\mathrm{si}}(a,d)$ such that Condition 5 is satisfied for every $A \in \mathcal{A}$. It holds that*

$$\hat{\rho}_n(\mathcal{A}) \lesssim \rho_n(\mathcal{A}) + an^{-1}(d\log p)^{1/2} + \Delta_n^{1/3}(\mathcal{A})(d\log p)^{2/3}\,.$$

(iii) *Under Conditions 1 and 6, it holds that*

$$\hat{\rho}_n\{\mathcal{A}^{\mathrm{sp}}(s)\} \lesssim \rho_n\{\mathcal{A}^{\mathrm{sp}}(s)\} + s^2\Delta_{n,r}^{1/3}(\log p)^{2/3} + \{B_n + s(\log p)^{1/2}\}n^{-1}\,.$$

Remark 3. (i) For the case of hyper-rectangles, $\Delta_{n,r} = o_{\mathrm{p}}\{(\log p)^{-2}\}$ is necessary to guarantee $\hat{\rho}_n(\mathcal{A}^{\mathrm{re}}) = o_{\mathrm{p}}(1)$. (ii) For $\mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a,d)$, to make $\hat{\rho}_n(\mathcal{A}) = o_{\mathrm{p}}(1)$, $\hat{\Xi}_n$ should satisfy $\Delta_n(\mathcal{A}) = o_{\mathrm{p}}\{(d\log p)^{-2}\}$. Notice that $\Delta_n(\mathcal{A}) \leq \Delta_{n,r}\sup_{A\in\mathcal{A}}\sup_{v\in\mathcal{V}\{A^K(A)\}}|v|_1^2$. If the ℓ^1-norm of the unit normal vectors outward to the facets of $A^K(A)$ is uniformly bounded away from infinity over $A \in \mathcal{A}$, it suffices to require $\Delta_{n,r} = o_{\mathrm{p}}\{(d\log p)^{-2}\}$. (iii) For the case of s-sparsely convex sets, to make $\hat{\rho}_n\{\mathcal{A}^{\mathrm{sp}}(s)\} = o_{\mathrm{p}}(1)$, we need to require $\Delta_{n,r} = o_{\mathrm{p}}\{(s^3\log p)^{-2}\}$.

As we have discussed in Remark 3, the validity of our proposed parametric bootstrap only requires the estimated long-run covariance matrix $\hat{\Xi}_n$ satisfying $|\hat{\Xi}_n - \Xi|_\infty = o_{\mathrm{p}}(\delta_n)$ for some $\delta_n \to 0$ as $n \to \infty$, where δ_n will be different for different selections of $\mathcal{A}$. There are various estimation methods for long-run covariance matrices, including the kernel-type estimators (Andrews, 1991) and utilizing moving block bootstraps (Lahiri, 2003). See also den Haan and Levin (1997) and Kiefer, Vogelsang and Bunzel (2000). Since the data sequence $\{X_t\}_{t=1}^n$ may be non-stationary, we suggest to adopt the kernel-type estimator for its long-run covariance matrix, that is

$$\hat{\Xi}_n = \sum_{j=-n+1}^{n-1} \mathcal{K}\left(\frac{j}{b_n}\right)\hat{H}_j\,, \tag{10}$$

where $\hat{H}_j = n^{-1}\sum_{t=j+1}^n (X_t - \bar{X})(X_{t-j} - \bar{X})^\top$ if $j \geq 0$ and $\hat{H}_j = n^{-1}\sum_{t=-j+1}^n (X_{t+j} - \bar{X})(X_t - \bar{X})^\top$ otherwise, with $\bar{X} = n^{-1}\sum_{t=1}^n X_t$. Here $\mathcal{K}(\cdot)$ is a symmetric kernel function that is continuous at 0 with $\mathcal{K}(0) = 1$, and b_n is the bandwidth diverging with n. Among a variety of kernel functions that guarantee the positive definiteness of the long-run covariance matrix estimators, Andrews (1991) derived an optimal kernel, i.e., the quadratic spectral kernel

$$\mathcal{K}_{\mathrm{QS}}(x) = \frac{25}{12\pi^2x^2}\left\{\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right\}, \tag{11}$$

by minimizing the asymptotic truncated mean square error of the estimator. To study the property of $\hat{\Xi}_n$ given in (10), we need the next condition imposed on the kernel $\mathcal{K}(\cdot)$.

Condition 7 (Kernel regularity). The kernel function $\mathcal{K}(\cdot):\mathbb{R}\to[-1,1]$ is continuously differentiable with bounded derivatives on $\mathbb{R}$ and satisfies (i) $\mathcal{K}(0)=1$, (ii) $\mathcal{K}(x)=\mathcal{K}(-x)$ for any $x\in\mathbb{R}$, and (iii) $|\mathcal{K}(x)|\lesssim|x|^{-\vartheta}$ as $|x|\to\infty$ for some constant $\vartheta>1$.

Theorem 11 (Bounds on $\Delta_{n,r}$). *Assume $p\geq n^{\kappa}$ for some universal constant $\kappa>0$. Let Condition 7 hold and $b_n\asymp n^{\rho}$.*

(i) *For α-mixing sequence $\{X_t\}$ with Conditions 1 and 2 being satisfied, if $0<\rho<(\vartheta-1)/(3\vartheta-2)$, there exist two constants $c_1>0$ depending only on (ρ,ϑ) and $c_2>0$ depending only on $(\gamma_1,\gamma_2,\vartheta)$ such that*

$$\Delta_{n,r}=O_{\mathrm{p}}\{B_n^2n^{-c_1}(\log p)^{c_2}\}+O(B_n^2n^{-\rho})\,.$$

(ii) *For m-dependent sequence $\{X_t\}$ with Condition 1 being satisfied, if $0<\rho<(\vartheta-1)/(3\vartheta-2)$, there exist two constants $c_1>0$ depending only on (ρ,ϑ) and $c_2>0$ depending only on (γ_1,ϑ) such that*

$$\Delta_{n,r}=O_{\mathrm{p}}\{B_n^2n^{-c_1}(\log p)^{c_2}\}+O(B_n^2m^2n^{-\rho})\,.$$

(iii) *For the sequence $\{X_t\}$ satisfying the model (4), assume Condition 1 is satisfied and $\Phi_{\psi_\nu,0}<\infty$ and $\Psi_{2,\alpha}<\infty$ for some $\alpha,\nu\in(0,\infty)$, if $0<\rho<\min\{(2\alpha+2\vartheta-3)/(2\vartheta-2),1/(2\alpha)\}$ with $\alpha>(3-2\vartheta)/2$, there exist two constants $c_1>0$ depending only on (ρ,ϑ,α) and $c_2>0$ depending only on $(\alpha,\vartheta,\gamma_1,\nu)$ such that*

$$\Delta_{n,r}=O_{\mathrm{p}}\{(B_n\Phi_{\psi_\nu,0}+B_n^2+\Phi^2_{\psi_\nu,0})n^{-c_1}(\log p)^{c_2}\}+O(n^{-\rho}\Psi_{2,0}\Psi_{2,\alpha}\varpi_n)$$

with $\varpi_n=(\log n)I(\alpha=1)+n^{1-\alpha}I(\alpha\neq1)$.

The proof of Theorem 11 is given in Section S14 of the supplementary material.

Remark 4. Write $H_j=n^{-1}\sum_{t=j+1}^{n}\mathbb{E}(X_tX_{t-j}^{\top})$ if $j\geq0$ and $H_j=n^{-1}\sum_{t=-j+1}^{n}\mathbb{E}(X_{t+j}X_t^{\top})$ if $j<0$. Let $\Xi^*=\sum_{j=-n+1}^{n-1}\mathcal{K}(j/b_n)H_j$. The terms $O(B_n^2n^{-\rho})$, $O(B_n^2m^2n^{-\rho})$ and $O(n^{-\rho}\Psi_{2,0}\Psi_{2,\alpha}\varpi_n)$ in (i), (ii) and (iii) of Theorem 11 are, respectively, bounds on the bias $|\Xi^*-\Xi|_\infty$ in the three cases. Thus the bias terms do not depend on p (at least directly). On the other hand, the terms $O_{\mathrm{p}}\{B_n^2n^{-c_1}(\log p)^{c_2}\}$, $O_{\mathrm{p}}\{B_n^2n^{-c_1}(\log p)^{c_2}\}$ and $O_{\mathrm{p}}\{(B_n\Phi_{\psi_\nu,0}+B_n^2+\Phi^2_{\psi_\nu,0})n^{-c_1}(\log p)^{c_2}\}$ are bounds on $|\hat{\Xi}_n-\Xi^*|_\infty$ in the three cases, respectively.

Now, combining Remark 3 and Theorem 11, we see that our proposed parametric bootstrap procedure is asymptotically valid even if the dimension p grows sub-exponentially fast in the sample size n. To implement the proposed parametric bootstrap, we need to solve two problems: (i) How to select bandwidth b_n in practice? and (ii) How to generate $\hat{G}\sim N(0,\hat{\Xi}_n)$ efficiently when p is large? For Problem (i), due to the positive definiteness of $\hat{\Xi}_n$ defined as (10) with the quadratic spectral kernel $\mathcal{K}_{\mathrm{QS}}(\cdot)$ defined as (11), we can use this kernel in our parametric bootstrap procedure. For $\mathcal{K}_{\mathrm{QS}}(\cdot)$, Andrews (1991) suggested Algorithm 1 to select b_n. For Problem (ii), to generate a random vector $\hat{G}\sim N(0,\hat{\Xi}_n)$, the standard approach consists of three steps: (a) perform the Cholesky decomposition for the $p\times p$ matrix $\hat{\Xi}_n=L^{\top}L$; (b) generate independent standard normal random variables $Z_1,\ldots,Z_p$ and let $Z=(Z_1,\ldots,Z_p)^{\top}$; (c) perform the transformation $\hat{G}=L^{\top}Z$. However, the computation complexity of the standard approach is $O(np^2+p^3)$ and it also requires a large storage space for $\{X_t\}_{t=1}^n$ and the

Algorithm 1 Data-driven procedure of bandwidth selection for $\mathcal{K}_{\mathrm{QS}}(\cdot)$

Step 1. For each $j \in [p]$, fit an AR(1) model to the j-th coordinate marginal sequence $\{X_{t,j}\}_{t=1}^{n}$. Denote by $\hat{\rho}_j$ and $\hat{\sigma}_j^2$, respectively, the estimated autoregressive coefficient and innovation variance.

Step 2. Select $b_n = 1.3221(\hat{a}n)^{1/5}$ with $\hat{a} = \{\sum_{j=1}^{p} 4\hat{\rho}_j^2\hat{\sigma}_j^4(1-\hat{\rho}_j)^{-8}\}/\{\sum_{j=1}^{p} \hat{\sigma}_j^4(1-\hat{\rho}_j)^{-4}\}$.

estimated matrix $\hat{\Xi}_n$. To circumvent the high computing cost with large p, we propose Algorithm 2 below which involves generating a random vector from an n-dimensional normal distribution instead. It is easy to check that such obtained $\hat{G} \sim N(0, \hat{\Xi}_n)$ conditionally on $\mathcal{X}_n$. Algorithm 2 was initially introduced in Chang, Yao and Zhou (2017).

4. Applications

In this section, we discuss several statistical applications of the Gaussian and parametric bootstrap approximation results for high-dimensional dependent data developed in Sections 2 and 3.

4.1. Testing high-dimensional mean vector

Given data $\{X_t\}_{t=1}^{n_1}$ with $\mathbb{E}(X_t) = \theta_x \in \mathbb{R}^p$ for any $t \in [n_1]$, it is of general interest in testing the hypothesis

$$H_0 : \theta_x = 0 \quad \text{versus} \quad H_1 : \theta_x \neq 0\,. \tag{12}$$

If there is another group of data $\{Y_t\}_{t=1}^{n_2}$ with $\mathbb{E}(Y_t) = \theta_y \in \mathbb{R}^p$ for any $t \in [n_2]$, we are also interested in the hypothesis testing problem

$$H_0 : \theta_x = \theta_y \quad \text{versus} \quad H_1 : \theta_x \neq \theta_y\,. \tag{13}$$

Hypotheses (12) and (13) are called, respectively, one-sample and two-sample mean testing problems in the literature. Lots of statistical inference problems in practice can be formulated as (12) and (13). Generally, the ℓ^2-type and ℓ^∞-type statistics are used to test the hypotheses (12) and (13) in the high-dimensional settings. With independent data, we refer to Chen and Qin (2010) and Cai, Liu and Xia (2014) for the uses of ℓ^2-type statistic and ℓ^∞-type statistic in these testing problems, respectively. It has been well known that the ℓ^2-type statistics are powerful for detecting relatively dense signals while the ℓ^∞-type statistics are preferable for detecting relatively sparse signals. In practice, we usually have less knowledge on whether the signals are dense or sparse. Let $\Gamma = \{\mathrm{Cov}(X_t)\}^{-1} = (\gamma_{i,j})_{p\times p}$. When Γ is known, to combine the advantages of the ℓ^2-type and ℓ^∞-type statistics, Zhang (2015) considered the following test statistic for (12) with independent data $\{X_t\}_{t=1}^{n_1}$:

$$T_n(s) = \max_{1\le j_1<\cdots<j_s\le p} \sum_{k=1}^{s} \frac{n_1 \bar{Z}_{j_k}^2}{\gamma_{j_k,j_k}}\,, \tag{14}$$

Algorithm 2 Generating $\hat{G} \sim N(0, \hat{\Xi}_n)$ for $\hat{\Xi}_n$ given in (10) with $\mathcal{K}_{\mathrm{QS}}(\cdot)$

Step 1. Obtain the bandwidth b_n by Algorithm 1. Define $\Theta = (\theta_{i,j})_{n\times n}$ with $\theta_{i,j} = \mathcal{K}_{\mathrm{QS}}\{(i-j)/b_n\}$.

Step 2. Generate $Z = (Z_1, \ldots, Z_n)^\top \sim N(0, \Theta)$ independent of $\{X_t\}_{t=1}^{n}$. Define $\hat{G} = n^{-1/2}\sum_{t=1}^{n} Z_t X_t$.

where $\bar{Z} = \Gamma\bar{X} := (\bar{Z}_1,\ldots,\bar{Z}_p)^\top$ with $\bar{X} = n_1^{-1}\sum_{t=1}^{n_1} X_t$. When Γ is unknown, Zhang (2015) proposed a feasible analogue for $T_n(s)$ by replacing Γ by its estimator $\hat{\Gamma}$. To simplify our presentation, we assume Γ is known in the rest of this subsection.

For dependent data, testing for white noise or serial correlation is a fundamental problem in statistical inference, as many testing problems in linear modelling can be transformed into a white noise test. Let $\{\varepsilon_t\}$ be a d-dimensional weakly stationary time series with mean zero. Denote by $\Sigma(k) = \mathrm{Cov}(\varepsilon_{t+k},\varepsilon_t)$ the autocovariance of ε_t at lag k. Given a prescribed integer K, the white noise hypothesis of $\{\varepsilon_t\}$ can be formulated as

$$H_0 : \Sigma(1) = \cdots = \Sigma(K) = 0 \quad \text{versus} \quad H_1 : H_0 \text{ is not true}. \tag{15}$$

Let $n_1 = n - K$ and $X_t = \{\mathrm{vec}(\varepsilon_{t+1}\varepsilon_t^\top),\ldots,\mathrm{vec}(\varepsilon_{t+K}\varepsilon_t^\top)\}^\top$, where $\mathrm{vec}(A)$ denotes a row vector that collecting all the elements in A. Then the white noise hypothesis (15) can be covered by the hypothesis (12) with $p = d^2K$. Chang, Yao and Zhou (2017) proposed a bootstrap test based on the ℓ^∞-type statistic for the white noise hypothesis (15) under the β-mixing assumption of $\{\varepsilon_t\}$. To enhance the power performance of Chang, Yao and Zhou (2017), we can use the test statistic $T_n(s)$ given in (14).

Notice that the distribution function of $T_n(s)$ can be written in terms of probability of the random vector $n_1^{1/2}\bar{Z}$ over a class of convex subsets of the form $\{w \in \mathbb{R}^p : \sum_{j\in\Theta_s}\gamma_{j,j}^{-1}w_j^2 \le t\}$ with $\Theta_s = \{w \in \mathbb{R}^p : |w|_0 = s\}$ which is a subset of the class of all s-sparsely convex sets in $\mathbb{R}^p$. Let $\hat{G} = (\hat{G}_1,\ldots,\hat{G}_p)^\top \sim N(0,\hat{\Xi}_n)$ with $\hat{\Xi}_n$ being the kernel-type estimator of the long-run covariance matrix $\mathrm{Cov}(n_1^{1/2}\bar{Z})$. Using the results developed in Sections 2 and 3, the null-distribution of $T_n(s)$ can be approximated by that of

$$\hat{T}_n = \max_{1\le j_1<\cdots<j_s\le p}\sum_{k=1}^{s}\frac{\hat{G}_{j_k}^2}{\gamma_{j_k,j_k}}$$

under both the α-mixing assumption and physical dependency assumption, where s can diverge with n at some polynomial rate. For any $\delta \in (0,1)$, let q_δ be the upper δ-quantile of the distribution of $\hat{T}_n$. Given the significant level δ, we reject the null hypothesis of the white noise hypothesis (15) if the test statistic $T_n(s)$ specified in (14) is larger than q_δ. Our procedure allows arbitrary dependency among the components of X_t.

4.2. Change point detection

Consider the problem of change point detection for high-dimensional distributions in a location family $X_t = \theta\cdot I(t > m) + \xi_t$, where $\theta \in \mathbb{R}^p$ is the location-shift parameter and $\{\xi_t\}$ is a sequence of stationary time series noise with mean zero. If $\theta = 0$ or $m \ge n$, there is no change point in $\{X_t\}_{t=1}^n$. Yu and Chen (2022) proposed a procedure to test whether there exists change point in the data based on the U-statistic

$$U_n = (U_{n,1},\ldots,U_{n,p})^\top = \binom{n}{2}^{-1}\sum_{1\le i<j\le n} h(X_i,X_j),$$

where $h : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^p$ is an anti-symmetric kernel $h(x,y) = -h(y,x)$. The anti-symmetry of the kernel h plays a key role in testing for the change point in terms of noise cancellations so that after proper normalization the distribution of U_n can be approximated by that of a Gaussian analogue. Specifically, under the null hypothesis that there is no change point and assuming independent and identically distributed noise $\{\xi_t\}$ with distribution F, Yu and Chen (2022) showed that $\mathbb{E}(U_n) = 0$ and the probability of U_n on max-rectangles can be well approximated by that of $N(0, 4\Gamma/3)$, where $\Gamma = \mathrm{Cov}\{g(X_1)\}$ and $g(x) = \mathbb{E}\{h(x, X_2)\}$ is the $\mathcal{L}^2(F)$ projection of h onto a linear subspace. On the other hand, under the alternative hypothesis when the change point location m is known, U_n is a two-sample Mann-Whitney test statistic (see e.g., Chapter 12 in van der Vaart (1998)), and the signal distortion under certain nonlinear kernels can be controlled such that the between-sample change point signal is magnitude preserving. To practically calibrate the distribution of $\max_{j\in[p]} |n^{1/2}U_{n,j}|_\infty$, Yu and Chen (2022) proposed a jackknife multiple bootstrap, which is powerful against alternatives with strong signals.

However, validity of the jackknife multiple bootstrap with a general nonlinear kernel heavily relies on the independent and identically distributed assumption of the noise sequence $\{\xi_t\}_{t=1}^n$ (Chen and Kato, 2020). For time series data, with the linear kernel $h(x, y) = x - y$ we may write

$$W_n = (W_{n,1}, \dots, W_{n,p})^\top = 2n^{-1/2}(n-1)^{-1}\sum_{t=1}^{n}(n-2t+1)X_t\,,$$

which can be viewed as one-pass CUSUM test statistic (Yu and Chen, 2021). Thus we can enhance the power performance of the change point test of Yu and Chen (2022) in the setting of linear kernel by using the test statistic

$$T_n(s) = \max_{1\le j_1<\cdots<j_s\le p}\sum_{k=1}^{s} W_{n,j_k}^2\,,$$

which allows s to diverge with n at some polynomial rate. Let $\hat{G} = (\hat{G}_1, \dots, \hat{G}_p)^\top \sim N(0, \hat{\Xi}_n)$ with $\hat{\Xi}_n$ being the kernel-type estimator of the long-run covariance matrix $\mathrm{Cov}(W_n)$ of the weighted sequence $\{2(n-1)^{-1}(n-2t+1)X_t\}_{t=1}^n$. Based on the results developed in Sections 2 and 3, the null-distribution of $T_n(s)$ can be calibrated by that of

$$\hat{T}_n = \max_{1\le j_1<\cdots<j_s\le p}\sum_{k=1}^{s} \hat{G}_{j_k}^2\,,$$

under both the α-mixing assumption and physical dependence assumption. For any $\delta \in (0,1)$, let q_δ be the upper δ-quantile of the distribution of $\hat{T}_n$. Given the significant level δ, we reject the null hypothesis that there is no change point if the test statistic $T_n(s) = \max_{1\le j_1<\cdots<j_s\le p}\sum_{k=1}^{s} W_{n,j_k}^2 > q_\delta$. Our procedure does not need to impose any specific structure assumption on the dependency among different components of X_t.

4.3. Confidence regions for the instantaneous covariance matrix and its inverse

Given d-dimensional dependent (and possibly non-stationary) data $\{Y_t\}_{t=1}^n$ with mean zero and instantaneous covariance Σ, i.e., $\mathbb{E}(Y_t) = 0$ and $\mathrm{Cov}(Y_t) = \Sigma$ for any $t \in [n]$, the instantaneous covariance matrix Σ and the precision matrix $\Omega = \Sigma^{-1} = (\omega_{i,j})_{d\times d}$ quantify the dependence among the d components of Y_t. Confidence regions for Σ and Ω can quantify the uncertainty in their estimates. For a given index set $\mathcal{S} \subset [d]^2$, denote by $\Sigma_{\mathcal{S}}$ and $\Omega_{\mathcal{S}}$ the vectors consisting, respectively, the entries of Σ and Ω with their indices in $\mathcal{S}$. We are interested in constructing a class of confidence regions $\{\mathcal{C}_{\mathcal{S},\delta}\}_{0<\delta<1}$ for $\Sigma_{\mathcal{S}}$ such that

$$\sup_{0<\delta<1} |\mathbb{P}(\Sigma_{\mathcal{S}} \in \mathcal{C}_{\mathcal{S},\delta}) - \delta| \to 0 \quad \text{as } n, d \to \infty\,. \tag{16}$$

We can also consider the confidence regions $\{C_{\mathcal{S},\delta}\}_{0<\delta<1}$ for $\Omega_{\mathcal{S}}$ such that

$$\sup_{0<\delta<1} |\mathbb{P}(\Omega_{\mathcal{S}} \in C_{\mathcal{S},\delta}) - \delta| \to 0 \quad \text{as } n, d \to \infty\,.$$

Given observations $\{Y_t\}_{t=1}^n$, we can estimate Σ as $\hat{\Sigma} = n^{-1}\sum_{t=1}^n Y_t Y_t^\top$ and estimate Ω by fitting the node-wise regressions $Y_{j,t} = \sum_{k\neq j}\beta_{j,k}Y_{k,t} + \epsilon_{j,t}$ for each $j \in [d]$. For high-dimensional scenario, we need to use the regularization method to estimate the parameters in the node-wise regressions. More specifically, write $\beta_j = (\beta_{j,1},\ldots,\beta_{j,j-1},-1,\beta_{j,j+1},\ldots,\beta_{j,d})^\top$ which can be estimated as

$$\hat{\beta}_j \equiv (\hat{\beta}_{j,1},\ldots,\hat{\beta}_{j,j-1},-1,\hat{\beta}_{j,j+1},\hat{\beta}_{j,d})^\top = \arg\min_{\gamma\in\Theta_j}\left\{\frac{1}{n}\sum_{t=1}^n(\gamma^\top Y_t)^2 + 2\lambda_j|\gamma|_1\right\},$$

where $\Theta_j = \{\gamma = (\gamma_1,\ldots,\gamma_d)^\top \in \mathbb{R}^d : \gamma_j = -1\}$ and $\lambda_j > 0$ is the tuning parameter. Let $V = \mathrm{Cov}(\epsilon_t) := (v_{i,j})_{d\times d}$ with $\epsilon_t = (\epsilon_{1,t},\ldots,\epsilon_{d,t})^\top$, which can be estimated as $\hat{V} = (\hat{v}_{i,j})_{d\times d}$ with

$$\hat{v}_{i,j} = \left\{-\frac{1}{n}\sum_{t=1}^n(\hat{\epsilon}_{i,t}\hat{\epsilon}_{j,t} + \hat{\beta}_{i,j}\hat{\epsilon}_{j,t}^2 + \hat{\beta}_{j,i}\hat{\epsilon}_{i,t}^2)\right\}I(i\neq j) + \left(\frac{1}{n}\sum_{t=1}^n\hat{\epsilon}_{i,t}\hat{\epsilon}_{j,t}\right)I(i=j)$$

and $\hat{\epsilon}_{j,t} = -\hat{\beta}_j^\top Y_t$ for $j \in [d]$ and $t \in [n]$. Then $\hat{\Omega} = (\hat{\omega}_{i,j})_{d\times d}$ with $\hat{\omega}_{i,j} = \hat{v}_{i,j}/(\hat{v}_{i,i}\hat{v}_{j,j})$ provides an estimate of Ω. Proposition 1 of Chang et al. (2018) shows that

$$\hat{\omega}_{i,j} - \omega_{i,j} = -\frac{1}{n}\sum_{t=1}^n v_{i,i}^{-1}v_{j,j}^{-1}(\epsilon_{i,t}\epsilon_{j,t} - v_{i,j}) + o_{\mathrm{p}}\{(n\log d)^{-1/2}\},$$

where the remainder term $o_{\mathrm{p}}\{(n\log d)^{-1/2}\}$ holds uniformly over $(i,j) \in [d]^2$. Let $p = |\mathcal{S}|$. We can see the leading terms of $\hat{\Sigma}_{\mathcal{S}} - \Sigma_{\mathcal{S}}$ and $\hat{\Omega}_{\mathcal{S}} - \Omega_{\mathcal{S}}$ can both be formulated as a general form $n^{-1}\sum_{t=1}^n X_t$ for some p-dimensional dependent sequence $\{X_t\}_{t=1}^n$. Write $\Xi = \mathrm{Cov}(n^{-1/2}\sum_{t=1}^n X_t)$ and denote by $\hat{\Xi}_n$ the estimate of Ξ given in Section 3. Let $\hat{G}^{(\mathcal{S})} = \{\hat{G}_1^{(\mathcal{S})},\ldots,\hat{G}_p^{(\mathcal{S})}\}^\top \sim N(0,\hat{\Xi}_n)$ and define

$$f_s\{\hat{G}^{(\mathcal{S})}\} = \max_{1\le j_1<\cdots<j_s\le p}\sum_{k=1}^s a_k\{\hat{G}_{j_k}^{(\mathcal{S})}\}^2,$$

where $a_1,\ldots,a_s > 0$ denote the prescribed weights. For any $\delta \in (0,1)$, let $q_{\mathcal{S},\delta}$ be the upper δ-quantile of the distribution of $f_s\{\hat{G}^{(\mathcal{S})}\}$, which can be determined by Monte Carlo simulation. Write $\hat{\Sigma}_{\mathcal{S}} = \{\hat{\sigma}_1^{(\mathcal{S})},\ldots,\hat{\sigma}_p^{(\mathcal{S})}\}^\top$. We can select the confidence region $C_{\mathcal{S},\delta}$ for $\Sigma_{\mathcal{S}}$ as follows:

$$C_{\mathcal{S},\delta} = \left\{\xi = (\xi_1,\ldots,\xi_p)^\top : \max_{1\le j_1<\cdots<j_s\le p}\sum_{k=1}^s a_k\{\hat{\sigma}_{j_k}^{(\mathcal{S})} - \xi_{j_k}\}^2 \le q_{\mathcal{S},\delta}\right\}. \tag{17}$$

Based on the results developed in Sections 2 and 3, we know such defined $C_{\mathcal{S},\delta}$ satisfies (16). Analogously, we can also obtain the confidence region $C_{\mathcal{S},\delta}$ for $\Omega_{\mathcal{S}}$ in the same manner. Chang et al. (2018) used this idea to construct the confidence region for $\Omega_{\mathcal{S}}$ with selecting $s = 1$ in (17), and established its validity under the β-mixing assumption. Using the results in Sections 2 and 3, we can show the validity of the confidence region defined as (17) in more general α-mixing setting and physical dependence setting with diverging s. Such defined $C_{\mathcal{S},\delta}$ can be applied for testing the structures and doing support recovering of Σ and Ω, respectively. See Chang et al. (2018) for the usefulness of these confidence regions. As we have discussed in Section 4.1, involving a larger s in (17) can enhance the power performance in finite-samples in comparison to the ℓ^∞-type statistic that with $s = 1$.

5. Proof of main results in Section 2.1

In this section, we provide proofs of the high-dimensional CLTs on hyper-rectangles in Section 2.1 under α-mixing (Theorem 1), dependency graph (Theorem 2), and physical dependence (Theorem 3) frameworks. Such quantitative CLTs on hyper-rectangles are the backbone for deriving CLTs on simple convex sets (Section 2.2) and sparsely convex sets (Section 2.3).

5.1. Proof of Theorem 1

To prove Theorem 1, we need the following lemma which is proved in Lemma C.5 by Chen and Kato (2019). The proof is also implicit in the proof of Lemma C.1 in Chen (2018), where a conditional version is given.

Lemma 1. *Let Y and W be centered Gaussian random vectors in $\mathbb{R}^p$ with covariance matrices $\Sigma_y = (\sigma^y_{j,k})_{j,k\in[p]}$ and $\Sigma_w = (\sigma^w_{j,k})_{j,k\in[p]}$, respectively. If $\min_{j\in[p]}\sigma^y_{j,j} \vee \min_{j\in[p]}\sigma^w_{j,j} \geq c$ for some universal constant $c > 0$, it then holds that $\sup_{u\in\mathbb{R}^p} |\mathbb{P}(Y \leq u) - \mathbb{P}(W \leq u)| \leq C|\Sigma_y - \Sigma_w|_\infty^{1/3}(\log p)^{2/3}$ for some constant $C > 0$ only depending on c.*

Without loss of generality, we let $\log p = o(n^{2/21})$, $\log p = o\{n^{\gamma_2/(3+6\gamma_2)}\}$ and $B_n^2(\log p)^{1/\gamma_2} = o(n^{1/3})$, since otherwise we can make the assertions trivial. Let $Q = o(n)$ be a positive integer that will diverge with n. We first decompose the sequence $[n]$ to $L+1$ blocks with $L = \lfloor n/Q\rfloor$: $\mathcal{G}_\ell = \{(\ell-1)Q+1,\ldots,\ell Q\}$ for $\ell \in [L]$ and $\mathcal{G}_{L+1} = \{LQ+1,\ldots,n\}$. Additionally, let b and h be two nonnegative integers such that $Q = b+h$ and $h = o(b)$. We decompose each $\mathcal{G}_\ell$ for $\ell \in [L]$ to a "large" block with length b and a "small" block with length h. Specifically, $\mathcal{I}_\ell = \{(\ell-1)Q+1,\ldots,(\ell-1)Q+b\}$ and $\mathcal{J}_\ell = \{(\ell-1)Q+b+1,\ldots,\ell Q\}$ for $\ell \in [L]$, and $\mathcal{J}_{L+1} = \mathcal{G}_{L+1}$. Define $\tilde{X}_\ell = b^{-1/2}\sum_{t\in\mathcal{I}_\ell} X_t$ and $\check{X}_\ell = h^{-1/2}\sum_{t\in\mathcal{J}_\ell} X_t$ for $\ell \in [L]$, and $\check{X}_{L+1} = (n-LQ)^{-1/2}\sum_{t\in\mathcal{J}_{L+1}} X_t$. Let $\{Y_t\}_{t=1}^n$ be a sequence of independent normal random vectors with mean zero, where the covariance of Y_t ($t \in \mathcal{I}_\ell$) is $\mathbb{E}(\tilde{X}_\ell\tilde{X}_\ell^\top)$ for each $\ell \in [L]$. Define $\tilde{Y}_\ell = b^{-1/2}\sum_{t\in\mathcal{I}_\ell} Y_t$ for $\ell \in [L]$. Let $S^{(1)}_{n,x} = L^{-1/2}\sum_{\ell=1}^L \tilde{X}_\ell$ and $S^{(1)}_{n,y} = L^{-1/2}\sum_{\ell=1}^L \tilde{Y}_\ell$. Write $\tilde{\Xi} = \mathrm{Cov}\{S^{(1)}_{n,y}\}$. It holds that $\tilde{\Xi} = L^{-1}\sum_{\ell=1}^L \mathbb{E}(\tilde{Y}_\ell\tilde{Y}_\ell^\top) = L^{-1}\sum_{\ell=1}^L \mathbb{E}(\tilde{X}_\ell\tilde{X}_\ell^\top)$. Define

$$\varrho_n^{(1)} := \sup_{u\in\mathbb{R}^p,\nu\in[0,1]} |\mathbb{P}\{\sqrt{\nu}S^{(1)}_{n,x} + \sqrt{1-\nu}S^{(1)}_{n,y} \leq u\} - \mathbb{P}\{S^{(1)}_{n,y} \leq u\}|, \tag{18}$$

$$\varrho_n^{(2)} := \sup_{u\in\mathbb{R}^p,\nu\in[0,1]} |\mathbb{P}\{\sqrt{\nu}S_{n,x} + \sqrt{1-\nu}S^{(1)}_{n,y} \leq u\} - \mathbb{P}\{S^{(1)}_{n,y} \leq u\}|. \tag{19}$$

Write $\tilde{X}_\ell = (\tilde{X}_{\ell,1},\ldots,\tilde{X}_{\ell,p})^\top$. We first present the following lemmas. The proof of Lemma 2 is almost identical to the proof of Lemma L1 in Chang, Jiang and Shao (2023) with $m = 1$ but using the condition $\mathbb{E}(|X_{t,j}|^4) \lesssim B_n^4$ in the steps based on Davydov's inequality (Davydov, 1968, Corollary 2). We omit details here. The proofs of Lemmas 3 and 4 are given in Sections S1.3 and S1.4 of the supplementary material, respectively.

Lemma 2. *Assume Conditions 1–2 hold. Then $|\tilde{\Xi} - \Xi|_\infty \lesssim B_n^2(hb^{-1} + bn^{-1})$.*

Lemma 3. *Assume Conditions 1–3 hold. Let $\gamma = \gamma_2/(2\gamma_2+1)$ and $h = C\{\log(pn)\}^{1/\gamma_2}$ for some sufficiently large $C > 0$. If $B_n^2 h \ll b \ll nB_n^{-2}$, then $\varrho_n^{(1)} \lesssim B_n L^{-1/6}(\log p)^{7/6}$ provided that $(\log p)^{3+\gamma} = o(b^{3\gamma/2}L^\gamma)$ and $\log p = o(L^{2/5})$.*

Lemma 4. *Let $p \geq n^{\kappa}$ for some universal constant $\kappa > 0$. Assume Conditions 1–3 hold. Let $\gamma = \gamma_2/(2\gamma_2+1)$ and $h = C(\log p)^{1/\gamma_2}$ for some sufficiently large $C > 0$. If b satisfies $\min\{nB_n^{-2}, n^{1/2}\} \gg b \gg \max\{n^{1/4}(\log p)^{(3-\gamma_2)/(4\gamma_2)}, B_n^2 h\}$, then $\varrho_n^{(2)} \lesssim B_n L^{-1/6}(\log p)^{7/6}$ provided that $\log p = o(L^{2/5})$ and $(\log p)^{3+\gamma} = o(b^{3\gamma/2}L^{\gamma})$.*

Now we begin to prove Theorem 1. Let $G \sim N(0,\Xi)$ be independent of $\mathcal{X}_n = \{X_1,\ldots,X_n\}$, where $\Xi = \mathrm{Cov}(n^{-1/2}\sum_{t=1}^n X_t)$. Recall $\varrho_n = \sup_{u\in\mathbb{R}^p, v\in[0,1]} |\mathbb{P}(\sqrt{v}S_{n,x} + \sqrt{1-v}G \leq u) - \mathbb{P}(G \leq u)|$ and note $S_{n,x}$ is independent of G and $S_{n,y}^{(1)}$. For any $u \in \mathbb{R}^p$, we have

$$
\begin{aligned}
&|\mathbb{P}(\sqrt{v}S_{n,x} + \sqrt{1-v}G \leq u) - \mathbb{P}\{\sqrt{v}S_{n,x} + \sqrt{1-v}S_{n,y}^{(1)} \leq u\}| \\
&\quad = \left| \int \mathbb{P}(\sqrt{1-v}G \leq u - \sqrt{v}v)\,\mathrm{d}F_{S_{n,x}}(v) - \int \mathbb{P}\{\sqrt{1-v}S_{n,y}^{(1)} \leq u - \sqrt{v}v\}\,\mathrm{d}F_{S_{n,x}}(v) \right| \\
&\quad \leq \int |\mathbb{P}(\sqrt{1-v}G \leq u - \sqrt{v}v) - \mathbb{P}\{\sqrt{1-v}S_{n,y}^{(1)} \leq u - \sqrt{v}v\}|\,\mathrm{d}F_{S_{n,x}}(v) \\
&\quad \leq \sup_{u\in\mathbb{R}^p} |\mathbb{P}(G \leq u) - \mathbb{P}\{S_{n,y}^{(1)} \leq u\}|\,,
\end{aligned}
$$

where $F_{S_{n,x}}(\cdot)$ denotes the distribution function of $S_{n,x}$. By triangle inequality, it holds that

$$
\begin{aligned}
\varrho_n \leq &\sup_{u\in\mathbb{R}^p} |\mathbb{P}(\sqrt{v}S_{n,x} + \sqrt{1-v}G \leq u) - \mathbb{P}\{\sqrt{v}S_{n,x} + \sqrt{1-v}S_{n,y}^{(1)} \leq u\}| \\
&+ \sup_{u\in\mathbb{R}^p} |\mathbb{P}\{\sqrt{v}S_{n,x} + \sqrt{1-v}S_{n,y}^{(1)} \leq u\} - \mathbb{P}\{S_{n,y}^{(1)} \leq u\}| \\
&+ \sup_{u\in\mathbb{R}^p} |\mathbb{P}\{S_{n,y}^{(1)} \leq u\} - \mathbb{P}(G \leq u)| \\
\leq\ &\varrho_n^{(2)} + 2\sup_{u\in\mathbb{R}^p} |\mathbb{P}(G \leq u) - \mathbb{P}\{S_{n,y}^{(1)} \leq u\}|\,.
\end{aligned}
\tag{20}
$$

Notice that $S_{n,y}^{(1)} \sim N(0,\tilde{\Xi})$ and $G \sim N(0,\Xi)$. By Lemmas 1 and 2 with $b = o(n^{1/2})$,

$$
\sup_{u\in\mathbb{R}^p} |\mathbb{P}\{S_{n,y}^{(1)} \leq u\} - \mathbb{P}(G \leq u)| \lesssim B_n^{2/3}h^{1/3}b^{-1/3}(\log p)^{2/3}\,.
$$

Notice that $\log p = o(n^{2/21})$, $\log p = o\{n^{\gamma_2/(3+6\gamma_2)}\}$ and $B_n^2(\log p)^{1/\gamma_2} = o(n^{1/3})$. Letting $h \asymp (\log p)^{1/\gamma_2}$ and $b \asymp n^{1/3}$, if $(\log p)^{3-\gamma_2} = o(n^{\gamma_2/3})$, together with Lemma 4, then $\varrho_n \lesssim B_n^{2/3}n^{-1/9}(\log p)^{(1+2\gamma_2)/(3\gamma_2)} + B_n n^{-1/9}(\log p)^{7/6}$. We construct Theorem 1. $\square$

5.2. Proof of Theorem 2

Without loss of generality, we assume $(D_n D_n^*)^2(\log p)^7 = o(n)$, since otherwise we can make the assertions hold trivially. Let $\{Y_t\}_{t=1}^n$, independent of $\mathcal{X}_n = \{X_1,\ldots,X_n\}$, be a centered Gaussian sequence such that $\mathrm{Cov}(Y_t,Y_s) = \mathrm{Cov}(X_t,X_s)$ for all $t,s \in [n]$. Define $S_{n,y} = n^{-1/2}\sum_{t=1}^n Y_t$. Then $S_{n,y} =^d G \sim N(0,\Xi)$. Let $\mathcal{W}_n = \{W_1,\ldots,W_n\}$ be an independent copy of $\mathcal{Y}_n = \{Y_1,\ldots,Y_n\}$ which is also independent of $\mathcal{X}_n$. Analogously, we can define $S_{n,w}$ based on $\mathcal{W}_n$. Write $X_t = (X_{t,1},\ldots,X_{t,p})^\top$, $Y_t = (Y_{t,1},\ldots,Y_{t,p})^\top$

and $W_t = (W_{t,1}, \ldots, W_{t,p})^\top$. Recall $\mathcal{N}_t = \{s \in V_n : (t,s) \in E_n\}$ for any $t \in [n]$. For $\phi \geq 1$, define

$$M_{n,x}(\phi) = \max_{t\in[n]} \mathbb{E}\bigg\{ \max_{j\in[p], s\in\cup_{\ell\in\mathcal{N}_t}\mathcal{N}_\ell} |X_{s,j}|^3 \times I\bigg(\max_{j\in[p], s\in\cup_{\ell\in\mathcal{N}_t}\mathcal{N}_\ell} |X_{s,j}| > \frac{\sqrt{n}}{8(D_n^4 D_n^*)^{1/3}\phi\log p}\bigg)\bigg\}, \tag{21}$$

$$\tilde{M}_{n,x}(\phi) = \max_{t\in[n]} \max_{s\in(\cup_{\ell\in\mathcal{N}_t}\mathcal{N}_\ell)\setminus\mathcal{N}_t} \mathbb{E}\bigg\{ \max_{j\in[p], s'\in\cup_{\ell\in\{s\}\cup\mathcal{N}_t}\mathcal{N}_\ell} |X_{s',j}|^3 \times I\bigg(\max_{j\in[p], s'\in\cup_{\ell\in\{s\}\cup\mathcal{N}_t}\mathcal{N}_\ell} |X_{s',j}| > \frac{\sqrt{n}}{8(D_n^4 D_n^*)^{1/3}\phi\log p}\bigg)\bigg\}.$$

Similarly, we define $M_{n,y}(\phi)$ and $\tilde{M}_{n,y}(\phi)$ in the same manner with X_s replaced by Y_s. Set $M_n(\phi) = M_{n,x}(\phi) + M_{n,y}(\phi)$ and $\tilde{M}_n(\phi) = \tilde{M}_{n,x}(\phi) + \tilde{M}_{n,y}(\phi)$.

Let $\beta = \phi\log p$. For a given $u = (u_1, \ldots, u_p)^\top \in \mathbb{R}^p$, define $F_\beta(v) = \beta^{-1}\log[\sum_{j=1}^p \exp\{\beta(v_j - u_j)\}]$ for any $v = (v_1, \ldots, v_p)^\top \in \mathbb{R}^p$. Such defined function $F_\beta(v)$ satisfies the property $0 \leq F_\beta(v) - \max_{j\in[p]}(v_j - u_j) \leq \beta^{-1}\log p = \phi^{-1}$ for any $v \in \mathbb{R}^p$. Select a thrice continuously differentiable function $g_0 : \mathbb{R} \to [0,1]$ whose derivatives up to the third order are all bounded such that $g_0(t) = 1$ for $t \leq 0$ and $g_0(t) = 0$ for $t \geq 1$. Define $g(t) := g_0(\phi t)$ for any $t \in \mathbb{R}$, and $q(v) := g\{F_\beta(v)\}$ for any $v \in \mathbb{R}^p$. Define

$$\mathcal{T}_n := q(\sqrt{\nu}S_{n,x} + \sqrt{1-\nu}S_{n,y}) - q(S_{n,w}).$$

Using the same arguments stated in Section S1.3 of the supplementary material, we have

$$\varrho_n \lesssim \phi^{-1}(\log p)^{1/2} + \sup_{u\in\mathbb{R}^p, \nu\in[0,1]} |\mathbb{E}(\mathcal{T}_n)|.$$

To specify the convergence rate of ϱ_n, we only need to bound $\sup_{u\in\mathbb{R}^p,\nu\in[0,1]} |\mathbb{E}(\mathcal{T}_n)|$. Define $Z(\omega) = \sum_{t=1}^n Z_t(\omega)$ for any $\omega \in [0,1]$, where

$$Z_t(\omega) = n^{-1/2}\{\sqrt{\omega}(\sqrt{\nu}X_t + \sqrt{1-\nu}Y_t) + \sqrt{1-\omega}W_t\}.$$

Then $Z(1) = \sqrt{\nu}S_{n,x} + \sqrt{1-\nu}S_{n,y}$ and $Z(0) = S_{n,w}$. For notational simplicity, we omit the dependence of $Z(\omega)$ and $Z_t(\omega)$ on ω in the rest of this proof. Let $\delta_t = \sum_{s\in\mathcal{N}_t} Z_s = (\delta_{t,1}, \ldots, \delta_{t,p})^\top$, $Z^{(-t)} = Z - \delta_t$, and

$$\dot{Z}_t = n^{-1/2}\{\omega^{-1/2}(\sqrt{\nu}X_t + \sqrt{1-\nu}Y_t) - (1-\omega)^{-1/2}W_t\} = (\dot{Z}_{t,1}, \ldots, \dot{Z}_{t,p})^\top.$$

By Taylor expansion, we have $2\mathbb{E}(\mathcal{T}_n) = \mathrm{I} + \mathrm{II} + \mathrm{III}$, where

$$\mathrm{I} = \sum_{j=1}^p \sum_{t=1}^n \int_0^1 \mathbb{E}[\partial_j q\{Z^{(-t)}\}\dot{Z}_{t,j}]\,\mathrm{d}\omega, \quad \mathrm{II} = \sum_{j,k=1}^p \sum_{t=1}^n \int_0^1 \mathbb{E}[\partial_{jk} q\{Z^{(-t)}\}\delta_{t,k}\dot{Z}_{t,j}]\,\mathrm{d}\omega,$$

$$\mathrm{III} = \sum_{j,k,l=1}^p \sum_{t=1}^n \int_0^1\int_0^1 (1-\tau)\mathbb{E}[\partial_{jkl} q\{Z^{(-t)} + \tau\delta_t\}\delta_{t,k}\delta_{t,l}\dot{Z}_{t,j}]\,\mathrm{d}\tau\mathrm{d}\omega.$$

Since $Z^{(-t)}$ and $\dot{Z}_{t,j}$ are independent, we have $\mathrm{I} = 0$. The following two lemmas give the upper bounds for II and III, respectively, whose proofs are given in Sections S2.1 and S2.2 of the supplementary material, respectively.

Lemma 5. $|\mathrm{III}| \lesssim D_n^2\phi^3 n^{-1/2}(\log p)^2\{M_n(\phi) + B_n^3\phi^{-1}(\log p)^{1/2} + B_n^3\varrho_n\}$.

Lemma 6. $|\mathrm{II}| \lesssim D_n D_n^*\phi^3 n^{-1/2}(\log p)^2\{\tilde{M}_n(\phi) + B_n^3\phi^{-1}(\log p)^{1/2} + B_n^3\varrho_n\}$.

Hence, by Lemmas 5 and 6, we have

$$\varrho_n \lesssim \phi^{-1}(\log p)^{1/2} + D_n\phi^3 n^{-1/2}(\log p)^2\{D_n M_n(\phi) + D_n^*\tilde{M}_n(\phi) + B_n^3 D_n^*\phi^{-1}(\log p)^{1/2} + B_n^3 D_n^*\varrho_n\}\,.$$

Taking $\phi = C'n^{1/6}B_n^{-1}(D_nD_n^*)^{-1/3}(\log p)^{-2/3}$ for some sufficiently small $C' > 0$, then

$$\varrho_n \lesssim B_n^{-3}M_n\{C'n^{1/6}B_n^{-1}(D_nD_n^*)^{-1/3}(\log p)^{-2/3}\} + B_n^{-3}\tilde{M}_n\{C'n^{1/6}B_n^{-1}(D_nD_n^*)^{-1/3}(\log p)^{-2/3}\} + n^{-1/6}B_n(D_nD_n^*)^{1/3}(\log p)^{7/6}\,.$$

Write

$$X^{(t)} = \max_{j\in[p],\, s\in\cup_{\ell\in\mathcal{N}_t}\mathcal{N}_\ell} |X_{s,j}| \quad \text{and} \quad X^{(t),s} = \max_{j\in[p],\, s'\in\cup_{\ell\in\{s\}\cup\mathcal{N}_t}\mathcal{N}_\ell} |X_{s',j}|\,.$$

By Condition 1, $\mathbb{P}\{X^{(t)} > u\} \le 2pD_n^*\exp(-u^{\gamma_1}B_n^{-\gamma_1})$ and $\mathbb{P}\{X^{(t),s} > u\} \le 4pD_n^*\exp(-u^{\gamma_1}B_n^{-\gamma_1})$ for any $u > 0$. Notice that $D_n \le D_n^* \le D_n^2$ and $\mathbb{E}\{|\xi|^3 I(|\xi| > v)\} = v^3\mathbb{P}(|\xi| > v) + 3\int_v^\infty u^2\mathbb{P}(|\xi| > u)\,\mathrm{d}u$. Due to $D_n^3(\log p)^{1+3/\gamma_1} = o(n)$ and $p \ge n^\kappa$ for some $\kappa > 0$, we have

$$\mathbb{E}[\{X^{(t)}\}^3 I\{X^{(t)} > B_n n^{1/3}(8C')^{-1}\cdot D_n^{-1}(\log p)^{-1/3}\}] \lesssim B_n^3\exp\{-Cn^{\gamma_1/3}D_n^{-\gamma_1}(\log p)^{-\gamma_1/3}\} \lesssim B_n^3 D_n^{1/2}n^{-1/6}(\log p)^{7/6}\,,$$

which implies

$$M_{n,x}\{C'n^{1/6}B_n^{-1}(D_nD_n^*)^{-1/3}(\log p)^{-2/3}\} \lesssim n^{-1/6}B_n^3 D_n^{1/2}(\log p)^{7/6}\,.$$

Recall Y_t is a normal random vector. Since $\max_{t\in[n],j\in[p]}\mathbb{E}(|X_{t,j}|^2) \lesssim B_n^2$ and $\mathrm{Cov}(Y_t, Y_s) = \mathrm{Cov}(X_t, X_s)$ for all $t, s \in [n]$, then $\max_{t\in[n],j\in[p]}\mathbb{P}(|Y_{t,j}| > u) \le 2\exp(-Cu^2B_n^{-2})$ for any $u > 0$. Analogously,

$$M_{n,y}\{C'n^{1/6}B_n^{-1}(D_nD_n^*)^{-1/3}(\log p)^{-2/3}\} \lesssim B_n^3 D_n^{1/2}n^{-1/6}(\log p)^{7/6}$$

provided that $D_n^3(\log p)^{5/2} = o(n)$. Notice that $(D_nD_n^*)^2(\log p)^7 = o(n)$ and $\gamma_1 \ge 1$. Thus,

$$M_n\{C'n^{1/6}B_n^{-1}(D_nD_n^*)^{-1/3}(\log p)^{-2/3}\} \lesssim B_n^3 D_n^{1/2}n^{-1/6}(\log p)^{7/6}\,.$$

By the same arguments, we can also show

$$\mathbb{E}[\{X^{t,(s)}\}^3 I\{X^{t,(s)} > (8C')^{-1}B_n n^{1/3}D_n^{-1}(\log p)^{-1/3}\}] \lesssim B_n^3 D_n^{1/2}n^{-1/6}(\log p)^{7/6}\,,$$

which implies

$$\tilde{M}_n\{C'n^{1/6}B_n^{-1}(D_nD_n^*)^{-1/3}(\log p)^{-2/3}\} \lesssim B_n^3 D_n^{1/2}n^{-1/6}(\log p)^{7/6}\,.$$

Hence, $\varrho_n \lesssim n^{-1/6}B_n(D_nD_n^*)^{1/3}(\log p)^{7/6}$. We complete the proof of Theorem 2. □

5.3. Proof of Theorem 3

Let $X_t^{(m)} = \mathbb{E}(X_t \mid \varepsilon_t, \ldots, \varepsilon_{t-m})$ for any $m \geq 1$. Then $\{X_t^{(m)}\}_{t=1}^n$ is an m-dependent sequence with mean zero. Let $\Xi^{(m)} = \mathrm{Cov}\{S_{n,x}^{(m)}\}$ with $S_{n,x}^{(m)} = n^{-1/2}\sum_{t=1}^n X_t^{(m)}$. Recall $S_{n,x} = n^{-1/2}\sum_{t=1}^n X_t$. Write $S_{n,x} = (S_{n,x,1}, \ldots, S_{n,x,p})^\top$ and $S_{n,x}^{(m)} = \{S_{n,x,1}^{(m)}, \ldots, S_{n,x,p}^{(m)}\}^\top$.

Lemma 7. *Let $\{X_t\}$ be a sequence of centered random vectors generated from the model (4) such that $\Phi_{\psi_\nu,\alpha} < \infty$ for some $\alpha, \nu \in (0,\infty)$. Then there exists a universal constant $C > 0$ depending only on ν such that* $\max_{j\in[p]} \mathbb{P}\{|S_{n,x,j}^{(m)} - S_{n,x,j}| > u\} \leq C\exp\{-(4e)^{-1}(1+2\nu)(um^\alpha\Phi_{\psi_\nu,\alpha}^{-1})^{2/(1+2\nu)}\}$ *for any $u > 0$.*

Lemma 8. *Let $q \geq 2$. For each $j \in [p]$, it holds that $\|S_{n,x,j}\|_q \leq (q-1)^{1/2}\Theta_{0,q,j}$, $\|S_{n,x,j}^{(m)}\|_q \leq (q-1)^{1/2}\Theta_{0,q,j}$ and $\|S_{n,x,j} - S_{n,x,j}^{(m)}\|_q \leq (q-1)^{1/2}\Theta_{m+1,q,j}$.*

The proof of Lemma 7 essentially follows from the arguments in proving Lemma C.3 of Zhang and Wu (2017) with the necessary modification using the uniform functional dependence measure to non-stationarity of the sequence $\{X_t\}$. Details are omitted. The proof of Lemma 8 is given in Section S3.1 of the supplementary material.

5.3.1. *Proof of Part* (i) *of Theorem* 3

Recall $\Xi = \mathrm{Cov}(S_{n,x})$. We will apply the large-and-small-blocks technique stated in Appendix 5.1 to derive the upper bound of ϱ_n. Without loss of generality, we assume $\Phi_{\psi_\nu,0} = o\{n^{\alpha/(3+9\alpha)}\}$ and $\Psi_{2,\alpha}\Psi_{2,0} = o\{n^{\alpha/(1+3\alpha)}\}$, since otherwise the assertions hold trivially. Let $Q = o(n)$ be a positive integer that will diverge with n. We first decompose the sequence $\{X_t^{(m)}\}_{t=1}^n$ to $L+1$ blocks with $L = \lfloor n/Q \rfloor$: $\mathcal{G}_\ell = \{(\ell-1)Q+1, \ldots, \ell Q\}$ for $\ell \in [L]$ and $\mathcal{G}_{L+1} = \{LQ+1, \ldots, n\}$. Let $b \gg m$ be a nonnegative integer such that $Q = b+m$. We decompose each $\mathcal{G}_\ell$ for $\ell \in [L]$ to a "large" block with length b and a "small" block with length m. Specifically, $\mathcal{I}_\ell = \{(\ell-1)Q+1, \ldots, (\ell-1)Q+b\}$ and $\mathcal{J}_\ell = \{(\ell-1)Q+b+1, \ldots, \ell Q\}$ for $\ell \in [L]$, and $\mathcal{J}_{L+1} = \mathcal{G}_{L+1}$. Define $\tilde{X}_\ell^{(m)} = b^{-1/2}\sum_{t\in\mathcal{I}_\ell} X_t^{(m)}$ and $\check{X}_\ell^{(m)} = m^{-1/2}\sum_{t\in\mathcal{J}_\ell} X_t^{(m)}$ for $\ell \in [L]$, and $\check{X}_{L+1}^{(m)} = (n-LQ)^{-1/2}\sum_{t\in\mathcal{J}_{L+1}} X_t^{(m)}$. Since $\{X_t^{(m)}\}_{t=1}^n$ is an m-dependent sequence, we know $\{\tilde{X}_\ell^{(m)}\}_{\ell=1}^L$ is an independent sequence. Let $\{Y_t^{(m)}\}_{t=1}^n$ be a sequence of independent normal random vectors with mean zero, where the covariance of $Y_t^{(m)}$ ($t \in \mathcal{I}_\ell$) is $\mathbb{E}[\tilde{X}_\ell^{(m)}\{\tilde{X}_\ell^{(m)}\}^\top]$ for each $\ell \in [L]$. Define $\tilde{Y}_\ell^{(m)} = b^{-1/2}\sum_{t\in\mathcal{I}_\ell} Y_t^{(m)}$ for $\ell \in [L]$. Let $\tilde{S}_{n,x}^{(m)} = L^{-1/2}\sum_{\ell=1}^L \tilde{X}_\ell^{(m)}$ and $\tilde{S}_{n,y}^{(m)} = L^{-1/2}\sum_{\ell=1}^L \tilde{Y}_\ell^{(m)}$. Write $\tilde{\Xi} = \mathrm{Cov}\{\tilde{S}_{n,y}^{(m)}\}$. Then $\tilde{\Xi} = L^{-1}\sum_{\ell=1}^L \mathbb{E}[\tilde{Y}_\ell^{(m)}\{\tilde{Y}_\ell^{(m)}\}^\top] = L^{-1}\sum_{\ell=1}^L \mathbb{E}[\tilde{X}_\ell^{(m)}\{\tilde{X}_\ell^{(m)}\}^\top] = \mathrm{Cov}\{\tilde{S}_{n,x}^{(m)}\}$. Define

$$\varrho_{n,1}^{(m)} := \sup_{u\in\mathbb{R}^p, \nu\in[0,1]} |\mathbb{P}\{\sqrt{\nu}\tilde{S}_{n,x}^{(m)} + \sqrt{1-\nu}\tilde{S}_{n,y}^{(m)} \leq u\} - \mathbb{P}\{\tilde{S}_{n,y}^{(m)} \leq u\}|,$$

$$\varrho_{n,2}^{(m)} := \sup_{u\in\mathbb{R}^p, \nu\in[0,1]} |\mathbb{P}\{\sqrt{\nu}S_{n,x}^{(m)} + \sqrt{1-\nu}\tilde{S}_{n,y}^{(m)} \leq u\} - \mathbb{P}\{\tilde{S}_{n,y}^{(m)} \leq u\}|.$$

Write $\tilde{X}_\ell^{(m)} = \{\tilde{X}_{\ell,1}^{(m)}, \ldots, \tilde{X}_{\ell,p}^{(m)}\}^\top$. We first present the following lemmas whose proofs are given in Sections S3.2–S3.6 of the supplementary material, respectively.

Lemma 9. *If $\Phi_{\psi_\nu,0} < \infty$ for some $\nu \in (0,\infty)$, then* $\max_{\ell\in[L]}\max_{j\in[p]} \mathbb{P}\{|\tilde{X}_{\ell,j}^{(m)}| > u\} \lesssim \exp\{-(4e)^{-1}(1+2\nu)(u\Phi_{\psi_\nu,0}^{-1})^{2/(1+2\nu)}\}$ *for any $u > 0$.*

Lemma 10. *If $\Phi_{\psi_\nu,0} < \infty$ for some $\nu \in (0,\infty)$, we have* $\max_{\ell\in[L]}\max_{j\in[p]}\mathbb{E}\{|\tilde{X}_{\ell,j}^{(m)}|^q\} \lesssim \Phi_{\psi_\nu,0}^q$ *for any positive integer $q \geq 1$.*

Lemma 11. *If $\Phi_{\psi_\nu,0}, \Psi_{2,\alpha} < \infty$ for some $\nu,\alpha \in (0,\infty)$, it holds that* $|\tilde{\Xi}-\Xi|_\infty \lesssim \Phi_{\psi_\nu,0}^2(mb^{-1}+bn^{-1}) + m^{-\alpha}\Psi_{2,\alpha}\Psi_{2,0}$.

Lemma 12. *Let $p \geq n^\kappa$ for some universal constant $\kappa > 0$. Assume $\Phi_{\psi_\nu,0}, \Psi_{2,\alpha} < \infty$ for some $\nu,\alpha \in (0,\infty)$, and $\min_{j\in[p]} V_{n,j} \geq C$ for some universal constant $C > 0$, where $V_{n,j}$ is defined in (3). If $\Phi_{\psi_\nu,0}^2 m \ll b \ll \Phi_{\psi_\nu,0}^{-2} n$ and $m^\alpha \gg \Psi_{2,\alpha}\Psi_{2,0}$, then $\varrho_{n,1}^{(m)} \lesssim L^{-1/6}\Phi_{\psi_\nu,0}(\log p)^{7/6}$ provided that $(\log p)^{5+6\nu} = o(L^2)$.*

Lemma 13. *Let $p \geq n^\kappa$ for some universal constant $\kappa > 0$. Assume $\Phi_{\psi_\nu,0}, \Psi_{2,\alpha} < \infty$ for some $\nu,\alpha \in (0,\infty)$, and $\min_{j\in[p]} V_{n,j} \geq C$ for some universal constant $C > 0$, where $V_{n,j}$ is defined in (3). If $\max\{m\Phi_{\psi_\nu,0}^2, n^{1/4}m^{3/4}(\log p)^{(6\nu-1)/4}\} \ll b \ll \min\{n\Phi_{\psi_\nu,0}^{-2}, (mn)^{1/2}\}$ and $m^\alpha \gg \Psi_{2,\alpha}\Psi_{2,0}$, then $\varrho_{n,2}^{(m)} \lesssim L^{-1/6}\Phi_{\psi_\nu,0}(\log p)^{7/6}$ provided that $(\log p)^{5+6\nu} = o(L^2)$.*

Recall $G \sim N(0,\Xi)$ and $\tilde{S}_{n,y}^{(m)} \sim N(0,\tilde{\Xi})$. By Lemma 1,

$$\sup_{u\in\mathbb{R}^p}|\mathbb{P}(G \leq u) - \mathbb{P}\{\tilde{S}_{n,y}^{(m)} \leq u\}| \lesssim |\tilde{\Xi}-\Xi|_\infty^{1/3}(\log p)^{2/3}.$$

Let $D_n = Cm^{-\alpha}\Phi_{\psi_\nu,\alpha}(\log p)^{(1+2\nu)/2}$ for some sufficiently large constant $C > 0$. Then $D_n(\log p)^{1/2} \lesssim m^{-\alpha}\Phi_{\psi_\nu,\alpha}(\log p)^{1+\nu}$. Define the event $\mathcal{E} = \{|S_{n,x} - S_{n,x}^{(m)}|_\infty \leq D_n\}$.

If $m^\alpha \gg \Psi_{2,\alpha}\Psi_{2,0}$ and $\max\{m\Phi_{\psi_\nu,0}^2, n^{1/4}m^{3/4}(\log p)^{(6\nu-1)/4}\} \ll b \ll \min\{n\Phi_{\psi_\nu,0}^{-2}, (mn)^{1/2}\}$, then

$$\begin{aligned}
\varrho_n \leq{}& \sup_{u\in\mathbb{R}^p,\nu\in[0,1]}|\mathbb{P}\{\sqrt{\nu}S_{n,x} + \sqrt{1-\nu}\tilde{S}_{n,y}^{(m)} \leq u\} - \mathbb{P}\{\tilde{S}_{n,y}^{(m)} \leq u\}| \\
&+ 2\sup_{u\in\mathbb{R}^p}|\mathbb{P}(G \leq u) - \mathbb{P}\{\tilde{S}_{n,y}^{(m)} \leq u\}| \\
\lesssim{}& \varrho_{n,2}^{(m)} + \mathbb{P}(\mathcal{E}^c) + \sup_{u\in\mathbb{R}^p,\nu\in[0,1]}|\mathbb{P}\{\tilde{S}_{n,y}^{(m)} \leq u - \sqrt{\nu}D_n\} - \mathbb{P}\{\tilde{S}_{n,y}^{(m)} \leq u\}| \qquad (22)\\
&+ \sup_{u\in\mathbb{R}^p,\nu\in[0,1]}|\mathbb{P}\{\tilde{S}_{n,y}^{(m)} \leq u + \sqrt{\nu}D_n\} - \mathbb{P}\{\tilde{S}_{n,y}^{(m)} \leq u\}| + |\tilde{\Xi}-\Xi|_\infty^{1/3}(\log p)^{2/3} \\
\lesssim{}& L^{-1/6}\Phi_{\psi_\nu,0}(\log p)^{7/6} + D_n(\log p)^{1/2} + \mathbb{P}(\mathcal{E}^c) + |\tilde{\Xi}-\Xi|_\infty^{1/3}(\log p)^{2/3},
\end{aligned}$$

provided that $(\log p)^{5+6\nu} = o(L^2)$, where the first step is identical to (20), and the last step is based on Lemma 13 and Nazarov's inequality. Lemma 7 implies

$$\mathbb{P}(\mathcal{E}^c) \lesssim p\exp\{-C(D_n m^\alpha \Phi_{\psi_\nu,\alpha}^{-1})^{2/(1+2\nu)}\} \lesssim L^{-1/6}\Phi_{\psi_\nu,0}(\log p)^{7/6}.$$

Due to $b \ll (mn)^{1/2}$, by Lemma 11, $|\tilde{\Xi}-\Xi|_\infty \lesssim \Phi_{\psi_\nu,0}^2 mb^{-1} + m^{-\alpha}\Psi_{2,\alpha}\Psi_{2,0}$. By (22) and recalling $L \asymp nb^{-1}$, we have

$$\begin{aligned}
\varrho_n \lesssim{}& b^{1/6}n^{-1/6}\Phi_{\psi_\nu,0}(\log p)^{7/6} + m^{-\alpha}\Phi_{\psi_\nu,\alpha}(\log p)^{1+\nu} \\
&+ \Phi_{\psi_\nu,0}^{2/3}m^{1/3}b^{-1/3}(\log p)^{2/3} + \Psi_{2,\alpha}^{1/3}\Psi_{2,0}^{1/3}m^{-\alpha/3}(\log p)^{2/3}.
\end{aligned}$$

With selecting $b \asymp m^{2/3}n^{1/3}$, if $\Psi_{2,\alpha}^{1/\alpha}\Psi_{2,0}^{1/\alpha} \ll m \ll \min\{n\Phi_{\psi_\nu,0}^{-6}, n(\log p)^{3(1-6\nu)}, n(\log p)^{-3(5+6\nu)/4}\}$, then

$$\varrho_n \lesssim m^{1/9}n^{-1/9}\Phi_{\psi_\nu,0}(\log p)^{7/6} + m^{-\alpha}\Phi_{\psi_\nu,\alpha}(\log p)^{1+\nu} + \Psi_{2,\alpha}^{1/3}\Psi_{2,0}^{1/3}m^{-\alpha/3}(\log p)^{2/3}\,.$$

Recall $\Phi_{\psi_\nu,0} = o\{n^{\alpha/(3+9\alpha)}\}$ and $\Psi_{2,\alpha}\Psi_{2,0} = o\{n^{\alpha/(1+3\alpha)}\}$. Letting $m \asymp n^{1/(1+3\alpha)}$, we have

$$\varrho_n \lesssim n^{-\alpha/(3+9\alpha)}(\log p)^{2/3}\{\Phi_{\psi_\nu,0}(\log p)^{1/2} + \Psi_{2,\alpha}^{1/3}\Psi_{2,0}^{1/3}\} + n^{-\alpha/(1+3\alpha)}\Phi_{\psi_\nu,\alpha}(\log p)^{1+\nu}$$

provided that $(\log p)^{6\nu-1} = o\{n^{\alpha/(1+3\alpha)}\}$ and $(\log p)^{(5+6\nu)/4} = o\{n^{\alpha/(1+3\alpha)}\}$. We have Part (i) of Theorem 3. □

5.3.2. *Proof of Part* (ii) *of Theorem* 3

Without loss of generality, we assume $\Phi_{\psi_\nu,\alpha}^{4+2\alpha} = o(n^\alpha)$, since otherwise the assertions hold trivially. Let $\{Y_t^{(m)}\}_{t=1}^n$ be a sequence of centered Gaussian random vectors in $\mathbb{R}^p$ such that $\mathrm{Cov}\{Y_t^{(m)}, Y_s^{(m)}\} = \mathrm{Cov}\{X_t^{(m)}, X_s^{(m)}\}$ for all $t, s \in [n]$. Set $S_{n,y}^{(m)} = n^{-1/2}\sum_{t=1}^n Y_t^{(m)}$. Recall $S_{n,x} = n^{-1/2}\sum_{t=1}^n X_t$. Lemma 7 implies $\max_{j\in[p]}\mathbb{E}\{|S_{n,x,j} - S_{n,x,j}^{(m)}|^2\} \lesssim \Phi_{\psi_\nu,\alpha}^2 m^{-2\alpha}$. Since $\mathrm{Var}\{S_{n,x,j}^{(m)}\} \geq \mathrm{Var}(S_{n,x,j}) + \mathbb{E}\{|S_{n,x,j} - S_{n,x,j}^{(m)}|^2\} - 2\{\mathrm{Var}(S_{n,x,j})\}^{1/2}[\mathbb{E}\{|S_{n,x,j} - S_{n,x,j}^{(m)}|^2\}]^{1/2}$, if we select $m \geq C\Phi_{\psi_\nu,\alpha}^{1/\alpha}$ for some sufficiently large $C > 0$, we know that $\mathrm{Var}\{S_{n,x,j}^{(m)}\}$ is uniformly bounded away from zero. By Hölder's inequality, Jensen's inequality and Condition 1, $\|X_{t,j}^{(m)}\|_{\psi_{\gamma_1}} \leq \|X_{t,j}\|_{\psi_{\gamma_1}} \leq B_n$ for any $t \in [n]$ and $j \in [p]$. Thus Corollary 1 yields that

$$\begin{aligned}\varrho_n\{S_{n,x}^{(m)}, S_{n,y}^{(m)}\} &:= \sup_{u\in\mathbb{R}^p, \nu\in[0,1]} |\mathbb{P}\{\sqrt{\nu}S_{n,x}^{(m)} + \sqrt{1-\nu}S_{n,y}^{(m)} \leq u\} - \mathbb{P}\{S_{n,y}^{(m)} \leq u\}| \\ &\lesssim n^{-1/6}B_n m^{2/3}(\log p)^{7/6}\,.\end{aligned}$$

Recall $G \sim N(0,\Xi)$ with $\Xi = \mathrm{Cov}(S_{n,x})$. Let $\tilde{D}_n = Cm^{-\alpha}\Phi_{\psi_\nu,\alpha}(\log p)^{(1+2\nu)/2}$ for some sufficiently large constant $C > 0$. Then $\tilde{D}_n(\log p)^{1/2} \lesssim m^{-\alpha}\Phi_{\psi_\nu,\alpha}(\log p)^{1+\nu}$. Consider the event $\mathcal{E} = \{|S_{n,x} - S_{n,x}^{(m)}|_\infty \leq \tilde{D}_n\}$. Write $\Xi^{(m)} = \mathrm{Cov}\{S_{n,x}^{(m)}\}$. By Lemma 1, $\sup_{u\in\mathbb{R}^p}|\mathbb{P}(G \leq u) - \mathbb{P}\{S_{n,y}^{(m)} \leq u\}| \lesssim |\Xi^{(m)} - \Xi|_\infty^{1/3}(\log p)^{2/3}$. Then

$$\begin{aligned}\varrho_n &\leq \sup_{u\in\mathbb{R}^p,\nu\in[0,1]} |\mathbb{P}\{\sqrt{\nu}S_{n,x} + \sqrt{1-\nu}S_{n,y}^{(m)} \leq u\} - \mathbb{P}\{S_{n,y}^{(m)} \leq u\}| \\ &\quad + 2\sup_{u\in\mathbb{R}^p}|\mathbb{P}(G \leq u) - \mathbb{P}\{S_{n,y}^{(m)} \leq u\}| \\ &\leq \varrho_n\{S_{n,x}^{(m)}, S_{n,y}^{(m)}\} + \mathbb{P}(\mathcal{E}^c) + \sup_{u\in\mathbb{R}^p,\nu\in[0,1]}|\mathbb{P}\{S_{n,y}^{(m)} \leq u - \sqrt{\nu}\tilde{D}_n\} - \mathbb{P}\{S_{n,y}^{(m)} \leq u\}| \\ &\quad + \sup_{u\in\mathbb{R}^p,\nu\in[0,1]}|\mathbb{P}\{S_{n,y}^{(m)} \leq u + \sqrt{\nu}\tilde{D}_n\} - \mathbb{P}\{S_{n,y}^{(m)} \leq u\}| + |\Xi^{(m)} - \Xi|_\infty^{1/3}(\log p)^{2/3} \\ &\lesssim n^{-1/6}B_n m^{2/3}(\log p)^{7/6} + \tilde{D}_n(\log p)^{1/2} + \mathbb{P}(\mathcal{E}^c) + |\Xi^{(m)} - \Xi|_\infty^{1/3}(\log p)^{2/3}\,,\end{aligned}$$

where the first step is identical to (20), and the last step is based on Nazarov's inequality. By Lemma 7, $\mathbb{P}(\mathcal{E}^c) \lesssim p\exp\{-C(\tilde{D}_n m^\alpha \Phi_{\psi_\nu,\alpha}^{-1})^{2/(1+2\nu)}\} \lesssim n^{-1/6}B_n m^{2/3}(\log p)^{7/6}$. It follows from Cauchy-Schwarz inequality and Lemma 8 that

$$
\begin{aligned}
&|\mathbb{E}(S_{n,x,j}S_{n,x,k}) - \mathbb{E}\{S^{(m)}_{n,x,j}S^{(m)}_{n,x,k}\}| \\
&\quad \le |\mathbb{E}[S_{n,x,j}\{S_{n,x,k} - S^{(m)}_{n,x,k}\}]| + |\mathbb{E}\{S^{(m)}_{n,x,k}\{S_{n,x,j} - S^{(m)}_{n,x,j}\}]| \\
&\quad \le \|S_{n,x,j}\|_2\|S_{n,x,k} - S^{(m)}_{n,x,k}\|_2 + \|S^{(m)}_{n,x,k}\|_2\|S_{n,x,j} - S^{(m)}_{n,x,j}\|_2 \\
&\quad \le \Theta_{0,2,j}\Theta_{m+1,2,k} + \Theta_{0,2,k}\Theta_{m+1,2,j}\,.
\end{aligned}
$$

Then $|\Xi^{(m)} - \Xi|_\infty \le 2(\max_{j\in[p]}\Theta_{0,2,j})(\max_{j\in[p]}\Theta_{m+1,2,j}) \lesssim m^{-\alpha}\Psi_{2,\alpha}\Psi_{2,0}$. Thus,

$$
\varrho_n \lesssim n^{-1/6}B_n m^{2/3}(\log p)^{7/6} + m^{-\alpha}\Phi_{\psi_\nu,\alpha}(\log p)^{1+\nu} + m^{-\alpha/3}\Psi_{2,\alpha}^{1/3}\Psi_{2,0}^{1/3}\cdot(\log p)^{2/3}\,.
$$

Letting $m \asymp n^{1/(4+2\alpha)}$, we have

$$
\varrho_n \lesssim n^{-\alpha/(12+6\alpha)}\{B_n(\log p)^{7/6} + (\log p)^{2/3}\Psi_{2,\alpha}^{1/3}\Psi_{2,0}^{1/3}\} + n^{-\alpha/(4+2\alpha)}\Phi_{\psi_\nu,\alpha}(\log p)^{1+\nu}\,.
$$

We have Part (ii) of Theorem 3. □

Acknowledgments

The authors would like to thank the editor, the associate editor and two reviewers for their constructive suggestions which led to the improvements of the paper.

Funding

Chang and Wu were supported in part by the National Natural Science Foundation of China (grant nos. 71991472, 72125008 and 11871401). Chang was also supported by the Center of Statistical Research at Southwestern University of Finance and Economics. Chen was supported by in part by the National Science Foundation (grant no. 1752614).

Supplementary Material

Additional technical proofs for "Central limit theorems for high dimensional dependent data" (DOI: 10.3150/23-BEJ1614SUPP; .pdf). Technical proofs of Theorems 4–11, and Proposition 1.

References

Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. MR1106513 https://doi.org/10.2307/2938229

Baldi, P. and Rinott, Y. (1989). On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.* **17** 1646–1650. MR1048950

Bentkus, V. (2003). On the dependence of the Berry-Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. MR1965117 https://doi.org/10.1016/S0378-3758(02)00094-0

Berkes, I., Liu, W. and Wu, W.B. (2014). Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.* **42** 794–817. MR3178474 https://doi.org/10.1214/13-AOP850

Bhattacharya, R.N. and Holmes, S. (2010). An exposition of Götze's estimation of the rate of convergence in the multivariate central limit theorem. Available at arXiv:1003.4254.

Bhattacharya, R.N. and Rao, R.R. (2010). *Normal Approximation and Asymptotic Expansions*. *Classics in Applied Mathematics* **64**. Philadelphia, PA: SIAM. MR3396213 https://doi.org/10.1137/1.9780898719895.ch1

Boussama, F., Fuchs, F. and Stelzer, R. (2011). Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Process. Appl.* **121** 2331–2360. MR2822779 https://doi.org/10.1016/j.spa.2011.06.001

Bradley, R.C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. MR2178042 https://doi.org/10.1214/154957805100000104

Bradley, R.C. (2007). *Introduction to Strong Mixing Conditions. Vol. 2*. Heber City, UT: Kendrick Press. MR2325295

Cai, T.T., Liu, W. and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. MR3164870 https://doi.org/10.1111/rssb.12034

Chang, J., Chen, X. and Wu, M. (2024). Supplement to "Central limit theorems for high dimensional dependent data." https://doi.org/10.3150/23-BEJ1614SUPP

Chang, J., Hu, Q., Liu, C. and Tang, C.Y. (2023). Optimal covariance matrix estimation for high-dimensional noise in high-frequency data. *J. Econometrics*. To appear. https://doi.org/10.1016/j.jeconom.2022.06.010

Chang, J., Jiang, Q. and Shao, X. (2023). Testing the martingale difference hypothesis in high dimension. *J. Econometrics*. To appear. https://doi.org/10.1016/j.jeconom.2022.09.001

Chang, J., Qiu, Y., Yao, Q. and Zou, T. (2018). Confidence regions for entries of a large precision matrix. *J. Econometrics* **206** 57–82. MR3840783 https://doi.org/10.1016/j.jeconom.2018.03.020

Chang, J., Tang, C.Y. and Wu, Y. (2013). Marginal empirical likelihood and sure independence feature screening. *Ann. Statist.* **41** 2123–2148. MR3127860 https://doi.org/10.1214/13-AOS1139

Chang, J., Yao, Q. and Zhou, W. (2017). Testing for high-dimensional white noise using maximum cross-correlations. *Biometrika* **104** 111–127. MR3626482 https://doi.org/10.1093/biomet/asw066

Chang, J., Zheng, C., Zhou, W.-X. and Zhou, W. (2017a). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics* **73** 1300–1310. MR3744543 https://doi.org/10.1111/biom.12695

Chang, J., Zhou, W., Zhou, W.-X. and Wang, L. (2017b). Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering. *Biometrics* **73** 31–41. MR3632349 https://doi.org/10.1111/biom.12552

Chen, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Ann. Statist.* **46** 642–678. MR3782380 https://doi.org/10.1214/17-AOS1563

Chen, X. and Kato, K. (2019). Randomized incomplete U-statistics in high dimensions. *Ann. Statist.* **47** 3127–3156. MR4025737 https://doi.org/10.1214/18-AOS1773

Chen, X. and Kato, K. (2020). Jackknife multiplier bootstrap: Finite sample approximations to the U-process supremum with applications. *Probab. Theory Related Fields* **176** 1097–1163. MR4087490 https://doi.org/10.1007/s00440-019-00936-y

Chen, S.X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697 https://doi.org/10.1214/09-AOS716

Chen, L.H.Y. and Shao, Q.-M. (2004). Normal approximation under local dependence. *Ann. Probab.* **32** 1985–2028. MR2073183 https://doi.org/10.1214/009117904000000450

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. MR3350040 https://doi.org/10.1007/s00440-014-0565-9

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

Chernozhukov, V., Chetverikov, D. and Kato, K. (2019). Inference on causal and structural parameters using many moment inequalities. *Rev. Econ. Stud.* **86** 1867–1900. MR4009488 https://doi.org/10.1093/restud/rdy065

Chernozhuokov, V., Chetverikov, D. and Koike, Y. (2023). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab.* To appear.

Chernozhuokov, V., Chetverikov, D., Kato, K. and Koike, Y. (2022a). Improved central limit theorem and bootstrap approximations in high dimensions. *Ann. Statist.* **50** 2562–2586. MR4500619 https://doi.org/10.1214/22-aos2193

Das, D. and Lahiri, S. (2021). Central Limit Theorem in high dimensions: The optimal bound on dimension growth rate. *Trans. Amer. Math. Soc.* **374** 6991–7009. MR4315595 https://doi.org/10.1090/tran/8459

Davydov, J.A. (1968). The convergence of distributions which are generated by stationary random processes. *Theory Probab. Appl.* **13** 691–696.

den Haan, W.J. and Levin, A.T. (1997). A practitioner's guide to robust covariance matrix estimation. In *Robust Inference*. *Handbook of Statist.* **15** 299–342. Amsterdam: North-Holland. MR1492717 https://doi.org/10.1016/S0169-7161(97)15014-3

Deng, H. (2020). Slightly conservative bootstrap for maxima of sums. Available at arXiv:2007.15877.

Deng, H. and Zhang, C.-H. (2020). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Ann. Statist.* **48** 3643–3671. MR4185823 https://doi.org/10.1214/20-AOS1946

Doukhan, P., Massart, P. and Rio, E. (1994). The functional central limit theorem for strongly mixing processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **30** 63–82. MR1262892

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. *Springer Series in Statistics*. New York: Springer. MR1964455 https://doi.org/10.1007/b97702

Fang, X. and Koike, Y. (2021). High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.* **31** 1660–1686. MR4312842 https://doi.org/10.1214/20-aap1629

Hafner, C.M. and Preminger, A. (2009). On asymptotic theory for multivariate GARCH models. *J. Multivariate Anal.* **100** 2044–2054. MR2543085 https://doi.org/10.1016/j.jmva.2009.03.011

Hoeffding, W. and Robbins, H. (1948). The central limit theorem for dependent random variables. *Duke Math. J.* **15** 773–780. MR0026771

Hörmann, S. (2009). Berry-Esseen bounds for econometric time series. *ALEA Lat. Am. J. Probab. Math. Stat.* **6** 377–397. MR2557877

Jirak, M. (2016). Berry-Esseen theorems under weak dependence. *Ann. Probab.* **44** 2024–2063. MR3502600 https://doi.org/10.1214/15-AOP1017

Kiefer, N.M., Vogelsang, T.J. and Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica* **68** 695–714. MR1769382 https://doi.org/10.1111/1468-0262.00128

Koike, Y. (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* **4** 257–297. MR4273258 https://doi.org/10.1007/s42081-020-00096-7

Koike, Y. (2023). High-dimensional central limit theorems for homogeneous sums. *J. Theoret. Probab.* To appear. https://doi.org/10.1007/s10959-022-01156-2

Kuchibhotla, A.K. and Rinaldo, A. (2020). High-dimensional CLT for sums of non-degenerate random vectors: $n^{-1/2}$-rate. Available at arXiv:2009.13673.

Lahiri, S.N. (2003). *Resampling Methods for Dependent Data*. *Springer Series in Statistics*. New York: Springer. MR2001447 https://doi.org/10.1007/978-1-4757-3803-2

Lopes, M.E. (2022). Central limit theorem and bootstrap approximation in high dimensions: Near $1/\sqrt{n}$ rates via implicit smoothing. *Ann. Statist.* **50** 2492–2513. MR4505371 https://doi.org/10.1214/22-aos2184

Lopes, M.E., Lin, Z. and Müller, H.-G. (2020). Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional and multinomial data. *Ann. Statist.* **48** 1214–1229. MR4102694 https://doi.org/10.1214/19-AOS1844

Petrov, V.V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. *Oxford Studies in Probability* **4**. Oxford University Press, New York: The Clarendon Press. MR1353441

Raič, M. (2019). A multivariate Berry-Esseen theorem with explicit constants. *Bernoulli* **25** 2824–2853. MR4003566 https://doi.org/10.3150/18-BEJ1072

Rio, E. (2009). Moment inequalities for sums of dependent random variables under projective conditions. *J. Theoret. Probab.* **22** 146–163. MR2472010 https://doi.org/10.1007/s10959-008-0155-9

Rio, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes*. *Probability Theory and Stochastic Modelling* **80**. Berlin: Springer. MR3642873 https://doi.org/10.1007/978-3-662-54323-8

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* **42** 43–47. MR0074711 https://doi.org/10.1073/pnas.42.1.43

Saulis, L. and Statulevičius, V.A. (1991). *Limit Theorems for Large Deviations*. Netherlands: Springer.

Song, Y., Chen, X. and Kato, K. (2019). Approximating high-dimensional infinite-order U-statistics: Statistical and computational guarantees. *Electron. J. Stat.* **13** 4794–4848. MR4038726 https://doi.org/10.1214/19-EJS1643

Sunklodas, Ĭ. (1984). The rate of convergence in the central limit theorem for strongly mixing random variables. *Lith. Math. J.* **24** 182–190.

van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

Wainwright, M.J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint*. *Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge: Cambridge Univ. Press. MR3967104 https://doi.org/10.1017/9781108627771

Wong, K.C., Li, Z. and Tewari, A. (2020). Lasso guarantees for β-mixing heavy-tailed time series. *Ann. Statist.* **48** 1124–1142. MR4102690 https://doi.org/10.1214/19-AOS1840

Wu, W.B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215 https://doi.org/10.1073/pnas.0506715102

Wu, W.B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389 https://doi.org/10.1214/009117907000000060

Wu, W.B. and Shao, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41** 425–436. MR2052582 https://doi.org/10.1239/jap/1082999076

Wu, W.-B. and Wu, Y.N. (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electron. J. Stat.* **10** 352–379. MR3466186 https://doi.org/10.1214/16-EJS1108

Yu, M. and Chen, X. (2021). Finite sample change point inference and identification for high-dimensional mean vectors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 247–270. MR4250275 https://doi.org/10.1111/rssb.12406

Yu, M. and Chen, X. (2022). A robust bootstrap change point test for high-dimensional location parameter. *Electron. J. Stat.* **16** 1096–1152. MR4377134 https://doi.org/10.1214/21-ejs1915

Zhang, X. (2015). Testing high dimensional mean under sparsity. Available at arXiv:1509.08444.

Zhang, X. and Cheng, G. (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli* **24** 2640–2675. MR3779697 https://doi.org/10.3150/17-BEJ939

Zhang, D. and Wu, W.B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. MR3718156 https://doi.org/10.1214/16-AOS1512

Received June 2022 and revised March 2023