High-dimensional bootstrap and asymptotic expansion

Yuta Koike [*][†]

March 16, 2026

Abstract

The recent seminal work of Chernozhukov, Chetverikov and Kato has shown that bootstrap approximation for the maximum of a sum of independent random vectors is justified even when the dimension is much larger than the sample size. In this context, numerical experiments suggest that third-moment matching bootstrap approximations would outperform normal approximation even without studentization, but the existing theoretical results cannot explain this phenomenon. In this paper, we develop an asymptotic expansion formula for the bootstrap coverage probability and show that it can give an explanation for the above phenomenon. In particular, we find the following interesting blessing of dimensionality phenomenon: The third-moment matching wild bootstrap is second-order accurate in high dimensions even without studentization if the covariance matrix has identical diagonal entries and bounded eigenvalues. We also show that a double wild bootstrap method is second-order accurate regardless of the covariance structure. The validity of these results is established under the assumption that the underlying distributions admit Stein kernels.

Keywords: Cornish–Fisher expansion; coverage probability; double bootstrap; Edgeworth expansion; second-order accuracy; Stein kernel.

1 Introduction

Let $X_1, \dots, X_n$ be independent centered random vectors in $\mathbb{R}^d$ with finite variance. Set

$$S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i.$$

The aim of this paper is to investigate the accuracy of bootstrap approximation for the maximum statistic

$$T_n := \max_{1 \le j \le d} S_{n,j},$$

when both n and d tend to infinity. The seminal work of Chernozhukov, Chetverikov & Kato [18] has established a Gaussian-type approximation for this statistic under very mild assumptions when the dimension d is possibly much larger than the sample size n. To be precise, let Z be a centered Gaussian vector in $\mathbb{R}^d$ with the same covariance matrix as S_n, say Σ. The Gaussian analog of T_n is given by

$$Z^\vee := \max_{1 \le j \le d} Z_j.$$

*Graduate School of Mathematical Sciences, University of Tokyo
†CREST, Japan Science and Technology Agency

Under mild moment assumptions, Chernozhukov, Chetverikov & Kato [18] have shown that

$$\sup_{t\in\mathbb{R}}|P(T_n \le t) - P(Z^\vee \le t)| = O\left(\left(\frac{\log^a(dn)}{n}\right)^b\right) \tag{1.1}$$

holds with $a = 7$ and $b = 1/8$. This result implies that given a significance level $\alpha \in (0, 1)$, the probability $P(T_n \ge c^G_{1-\alpha})$ is approximately equal to α as long as $\log d = o(n^{1/7})$, where $c^G_{1-\alpha}$ is the $(1-\alpha)$-quantile of $Z^\vee$. Therefore, we can use $c^G_{1-\alpha}$ as a critical value to construct asymptotically $(1-\alpha)$-level simultaneous confidence intervals or α-level tests for a high-dimensional vector of parameters; see [6, 23] for details. In practice, $c^G_{1-\alpha}$ is not computable because Σ is generally unknown, so we need to replace it by an estimate. In [18], this is implemented by the Gaussian wild (or multiplier) bootstrap: Let $w_1, \dots, w_n$ be i.i.d. standard normal variables independent of the data $X_1, \dots, X_n$. Define the Gaussian wild bootstrap version of S_n as follows:

$$S^*_n := \frac{1}{\sqrt{n}}\sum_{i=1}^n w_i(X_i - \bar{X}), \quad \text{where } \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i. \tag{1.2}$$

We may naturally expect that $c^G_{1-\alpha}$ would be well-approximated by the $(1-\alpha)$-quantile of the conditional law of $T^*_n := \max_{1\le j\le d} S^*_{n,j}$ given the data, say $\hat{c}_{1-\alpha}$. This is formally justified by [18]: They essentially prove

$$P(T_n \ge \hat{c}_{1-\alpha}) = \alpha + O\left(\left(\frac{\log^a(dn)}{n}\right)^b\right) \tag{1.3}$$

with $a = 7$ and $b = 1/8$. Subsequently, Chernozhukov *et al.* [20] have improved the convergence rates of (1.1) and (1.3) to $b = 1/6$. They also proved the left hand side of (1.1) can be replaced by $\sup_{A\in\mathcal{R}}|P(S_n \in A) - P(Z \in A)|$, where $\mathcal{R} := \{\prod_{j=1}^d [a_j, b_j] : a_j \le b_j,\ j = 1, \dots, d\}$ is the class of rectangles in $\mathbb{R}^d$.

It is easy to see that the conditional law of S^*_n given the data is $N(0, \widehat{\Sigma}_n)$, where $\widehat{\Sigma}_n$ is the sample covariance matrix: $\widehat{\Sigma}_n := n^{-1}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^\top$. Hence, the Gaussian wild bootstrap is essentially a feasible version of the normal approximation for T_n. Then, it is natural to ask whether the approximation accuracy can be improved by more sophisticated bootstrap methods such as the empirical and non-Gaussian wild bootstraps. In the fixed-dimensional setting, it is well-known that standard bootstrap methods improve the approximation accuracy in the coverage probabilities over the normal approximation only when the statistic of interest is asymptotically pivotal (cf. [38, Chapter 3]). Even worse, they can be harmful for the univariate sample mean without studentization; see [47, Section 3]. Despite these facts, numerical experiments suggest that third-moment matching bootstrap methods would outperform normal approximation (cf. [22, 26]). To appreciate this, we depict in Fig. 1 the P–P plot for the rejection rate $P(T_n \ge \hat{c}_{1-\alpha})$ against the nominal significance level α when $n = 200$ and $d = 400$, where $\hat{c}_{1-\alpha}$ is computed by either the Gaussian wild bootstrap or a wild bootstrap with third-moment matching. We can clearly see that the latter performance is much better than the former.

Deng & Zhang [26] tried to explain this phenomenon by showing that convergence rates of third-moment matching bootstrap approximations have a better dimension dependence, i.e. they achieve $a = 5$ and $b = 1/6$ in (1.3). Later, however, it was shown in [39] that the same convergence rate is achieved by normal approximation, i.e. (1.1) holds with $a = 5$ and $b = 1/6$. Chernozhukov *et al.* [22] have further improved

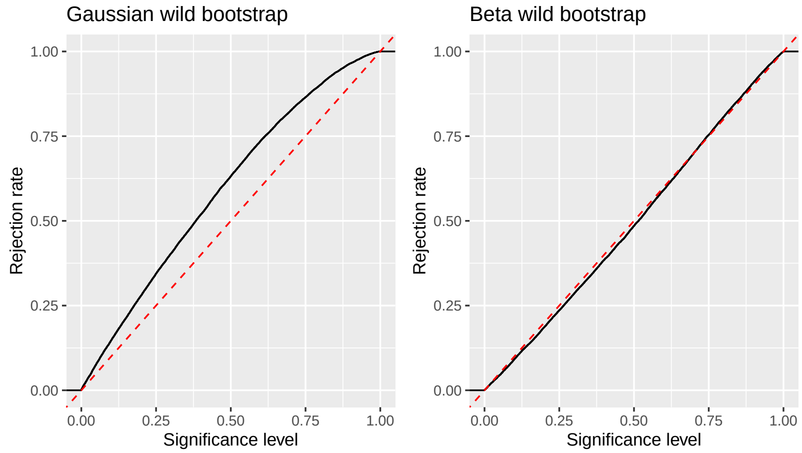

Figure 1: P–P plots for the rejection rate $P(T_n \geq \hat{c}_{1-\alpha})$ against the nominal significance level α when $n = 200$ and $d = 400$. The rejection rate is evaluated based on 20,000 Monte Carlo iterations. The critical value $\hat{c}_{1-\alpha}$ is computed by the Gaussian wild bootstrap for the left panel and the wild bootstrap with w_1 generated from the standardized beta distribution with parameters α, β given by (2.8) with $\nu = 0.1$ for the right panel, respectively. The number of bootstrap replications is 499. $X_1, \dots, X_n$ are generated from a Gaussian copula model with gamma marginals as in the simulation study of Section 3. The parameter matrix is $R = (0.2^{|j-k|})_{1 \leq j,k \leq d}$.

the convergence rate to $a = 5$ and $b = 1/4$ for both normal and bootstrap approximations. Moreover, if we require Σ to be invertible, it is possible to achieve the Berry–Esseen rate $n^{-1/2}$ up to a log factor even in the high-dimensional setting. Results in this direction first appeared in Fang & Koike [29], where the following result is obtained when $X_1, \dots, X_n$ are log-concave:

$$\sup_{A \in \mathcal{R}} |P(S_n \in A) - P(Z \in A)| = O\left(\sqrt{\frac{\log^3 d}{n}} \log n\right). \tag{1.4}$$

This rate is known to be optimal up to the $\log n$ factor in terms of both n and d; see Proposition 1.1 in [29] and also Corollary 2.1. This type of result has been further investigated in [24, 42, 48]. In particular, Chernozhukov *et al.* [24] have obtained the above nearly optimal rate when $\max_{i,j} |X_{ij}|$ is bounded. Here, the boundedness condition can be replaced with the sub-exponential condition by a simple truncation argument; see Appendix A. Further, in some situations, the rate $n^{-1/2}$ is (nearly) attainable even when Σ is (asymptotically) degenerate; see [31, 32, 49]. Nevertheless, all of these improvements are valid for normal approximation and thus do not explain the superior performance of third-moment matching bootstrap approximations.

In this paper, we aim to explain the superior performance of bootstrap approximation in high dimensions using Edgeworth expansion and related techniques. Specifically, we develop an asymptotic expansion formula for $P(T_n \geq \hat{c}_{1-\alpha})$ in Theorem 2.3. One main implication of this formula is that when $d \geq n$ and Σ has identical diagonal entries and bounded eigenvalues, the wild bootstrap with third-moment matching is second-order accurate in the sense that

$$P(T_n \geq \hat{c}_{1-\alpha}) = \alpha + O\left(\frac{\log^a(dn)}{n}\right)$$

for some constant $a > 0$ (see Corollary 2.2). This shows that high dimensionality can be beneficial to the accuracy of the third-moment matching wild bootstrap, revealing the blessing of dimensionality in this context. By contrast, the Gaussian wild bootstrap does not benefit from the high dimensionality in this situation, which clearly explains the performance difference in Fig. 1. At the same time, our asymptotic expansion formula also shows that this is not always the case: The structure of Σ strongly determines whether the above improvement occurs. In particular, when Σ is an equicorrelation matrix, the third-moment matching wild bootstrap could be inferior to the Gaussian wild bootstrap; see Corollary 2.4 and the simulation results in Section 3. For this reason, we also develop an alternative approximation to the quantiles of T_n that is second-order accurate regardless of the structure of Σ. A classical solution to this problem is bootstrapping the studentized version of S_n, but this is impossible in high dimensions since the sample covariance matrix $\widehat{\Sigma}_n$ is degenerate whenever $d \geq n$. Instead, we achieve this by Beran [7]'s double bootstrap method, another classical technique to improve the approximation accuracy for non-pivotal statistics; see Section 2.3.

Despite the fact that Edgeworth expansion is a standard tool to analyze the performance of bootstrap in the classical setting (cf. [38]), this approach has not been investigated for the above problem so far. One main reason is the lack of valid Edgeworth expansion for T_n in the high-dimensional setting. While asymptotic expansions for statistics of high-dimensional data have been actively studied in multivariate statistics (see [35] for an overview), results developed there seem inapplicable to our problem. One main reason is that T_n may not have any limit distribution as $n, d \to \infty$ even after properly scaled. In fact, this is one of the motivations for the development of Chernozhukov–Chetverikov–Kato's theory. In view of (1.4), we are concerned with Edgeworth expansion of $P(S_n \in A)$ over $A \in \mathcal{R}$. In the fixed-dimensional setting, a valid Edgeworth expansion of $P(S_n \in A)$ is conventionally derived from an asymptotic expansion of the characteristic function of S_n via Fourier analysis (see e.g. [8]). Such an argument makes the dimension dependence of the error bound extremely complicated, so it is rarely given explicitly. One exceptional work is Anderson *et al.* [2], but their proof technique seems to inherently require the condition $d \ll n$ and is thus inapplicable to our setting. In fact, in the high-dimensional setting, the geometry of the set A plays a key role to obtain an improved dimension dependence of error bounds, and it is unclear how to incorporate such information into Fourier analytic arguments. We also mention the recent work by Zhilova [60] who establishes explicit, computable error bounds for $\sup_{A \in \mathcal{A}} |P(S_n \in A) - P(S_n' \in A)|$ where S_n' is another sum of independent random vectors and $\mathcal{A}$ is either the class of balls or half-spaces. However, apart from other technical issues, these error bounds contain $1/\sqrt{n}$ terms and cannot be used for second-order analysis.

To circumvent the above issue, we develop valid asymptotic expansions using Stein's method. The use of Stein's method for asymptotic expansion was initiated by Barbour [4] who derived an asymptotic expansion of $\mathrm{E}[h(S_n)]$ when $d = 1$ and h is a smooth function. To drop the smoothness of the test function h, the so-called Cramér's condition is usually assumed in the Fourier analytic approach, but it is unknown how to (directly) incorporate Cramér's condition into Stein's method based arguments. Instead, we assume that the underlying random vectors have Stein kernels, motivated by the recent development of this approach by Fang & Liu [33] in the univariate case (see Lemma 2.1 ibidem). Apart from the technical difficulty, Cramér's condition is violated whenever the underlying statistic has a singular covariance matrix. This is unsuitable for application to bootstrap statistics in high-dimensions, so Stein kernels will be a more appropriate tool for our problem (see Remark 2.4).

In addition to the above development, we also establish two new inequalities for high-dimensional normal distributions in order to overcome further difficulties that arise specifically when proving the validity of asymptotic expansions in high dimensions. The first inequality is an anti-concentration inequality on rectangles for the higher-order terms of the Edgeworth expansion. Such a bound is necessary when controlling the remainder term appearing in an application of the so-called smoothing inequality, which is a standard initial step in justifying multivariate Edgeworth expansions (see e.g. [8, page 91]). Existing bounds, such as [8, Corollary 3.2], involve constants that grow polynomially with the dimension and therefore do not work in ultra-high-dimensional regimes where the dimension is far larger than the sample size. To address this limitation, in Lemma 4.2 we derive a new anti-concentration inequality for the higher-order terms of the Edgeworth expansion that depends only poly-logarithmically on the dimension. See also Remark 2.3. The second inequality is necessary for establishing the validity of the Cornish–Fisher expansion used in justifying the asymptotic expansion formula for $P(T_n \geq \hat{c}_{1-\alpha})$. In the conventional derivation of the Cornish–Fisher expansion, one assumes that the distribution function of the statistic admits an absolutely continuous limit distribution function G. The Cornish–Fisher expansion is then obtained by transforming the Edgeworth expansion through the inverse of G and applying a Taylor expansion (see [35, Section 5.2]). In this situation, since G does not depend on n, the validity of the Cornish–Fisher expansion follows almost automatically from that of the Edgeworth expansion. However, as noted earlier, in high-dimensional settings the maximum statistic T_n generally does not possess a non-degenerate limit distribution. Consequently, one must instead transform the Edgeworth expansion using the distribution function F_Z of $Z^\vee$, which depends on n, thereby requiring a sufficiently sharp *non-asymptotic* bound for the inverse of F_Z. To resolve this issue, we derive in Lemma 5.1 a new isoperimetric-type inequality for Gaussian maxima that yields precisely the control needed; see also Remark 2.6.

The remainder of the paper is organized as follows. Section 2 collects the main results of this paper: In Section 2.1, we establish valid Edgeworth expansions for S_n and S_n^* in high dimensions. Then, we develop an asymptotic expansion formula for $P(T_n \geq \hat{c}_{1-\alpha})$ and discuss its implications in Section 2.2. Finally, we show in Section 2.3 that a double wild bootstrap method is second-order accurate. Section 3 contains a small simulation study. Most proofs are collected in Sections 4 and 5. The appendix contains additional proofs and auxiliary results.

Notation For a vector $x \in \mathbb{R}^d$, we set $|x| := \sqrt{\sum_{j=1}^d x_j^2}$ and $x^\vee := \max_{1\leq j\leq d} x_j$. We denote by $\mathbf{1}_d = (1,\dots,1)^\top \in \mathbb{R}^d$ the all-ones vector in $\mathbb{R}^d$. For $r \in \mathbb{N}$, $(\mathbb{R}^d)^{\otimes r}$ denotes the set of real-valued d-dimensional r-arrays $V = (V_{j_1,\dots,j_r})_{1\leq j_1,\dots,j_r\leq d}$. In particular, $(\mathbb{R}^d)^{\otimes 1} = \mathbb{R}^d$ and $(\mathbb{R}^d)^{\otimes 2}$ is the set of $d \times d$ matrices. For $U \in (\mathbb{R}^d)^{\otimes q}$ and $V \in (\mathbb{R}^d)^{\otimes r}$, we set $U \otimes V := (U_{i_1,\dots,i_q} V_{j_1,\dots,j_r})_{1\leq i_1,\dots,i_q,j_1,\dots,j_r\leq d} \in (\mathbb{R}^d)^{\otimes(q+r)}$. We write $U^{\otimes 2} = U \otimes U$ for short. When $q = r$, we also set $\langle U, V\rangle := \sum_{j_1,\dots,j_r=1}^d U_{j_1,\dots,j_r} V_{j_1,\dots,j_r}$. In particular, when $q = r = 1$, $\langle U, V\rangle$ is the Euclidean inner product of U and V, which we also write $U \cdot V$. In addition, we set $\|V\|_1 := \sum_{j_1,\dots,j_r=1}^d |V_{j_1,\dots,j_r}|$ and $\|V\|_\infty := \max_{1\leq j_1,\dots,j_r\leq d} |V_{j_1,\dots,j_r}|$. Further, for $x \in \mathbb{R}^d$, we define

$x^{\otimes r} := (x_{j_1} \cdots x_{j_r})_{1 \le j_1,\dots,j_r \le d} \in (\mathbb{R}^d)^{\otimes r}$. Finally, we set

$$\overline{X^r} := \frac{1}{n} \sum_{i=1}^{n} X_i^{\otimes r}.$$

Given an r-times differentiable function $h : \mathbb{R}^d \to \mathbb{R}$, we set $\nabla^r h(x) := (\partial_{j_1,\dots,j_r} h(x))_{1 \le j_1,\dots,j_r \le d} \in (\mathbb{R}^d)^{\otimes r}$ for $x \in \mathbb{R}^d$, where $\partial_{j_1,\dots,j_r} = \frac{\partial^r}{\partial x_{j_1} \cdots \partial x_{j_r}}$. For $m \in \mathbb{N} \cup \{\infty\}$, $C_b^m(\mathbb{R}^d)$ denotes the set of bounded C^m functions with bounded derivatives.

For an invertible matrix V, ϕ_V denotes the density of $N(0, V)$. We write $\phi_d = \phi_{I_d}$ for short, where I_d is the $d \times d$ identity matrix. Further, we write $\phi = \phi_1$ for short. Φ denotes the standard normal distribution function. Also, for a distribution function $F : \mathbb{R} \to [0, 1]$, its (generalized) inverse is defined as $F^{-1}(p) = \inf\{t \in \mathbb{R} : F(t) \ge p\},\ p \in (0, 1)$. We refer to Appendix A.1 in [11] for useful properties of inverse distribution functions.

For a random vector ξ and $p \in (1, \infty)$, we set $\|\xi\|_p := (\mathrm{E}[|\xi|^p])^{1/p}$ (recall that $|\cdot|$ is the Euclidean norm). Further, for $\alpha > 0$, we set $\|\xi\|_{\psi_\alpha} := \inf\{t > 0 : \mathrm{E}[\exp\{(|\xi|/t)^\alpha\}] \le 2\}$. For two random vectors ξ and η, we write $\xi \stackrel{d}{=} \eta$ if ξ has the same law as η.

We assume $d \ge 3$ whenever we consider an expression containing $\log d$. A similar convention is applied to n.

2 Main results

Throughout the paper, we assume that S_n has an invertible covariance matrix Σ and denote by σ_* the square root of the minimum eigenvalue of Σ. We also set $\overline{\sigma} = \max_{j=1,\dots,d} \sqrt{\Sigma_{jj}}$ and $\underline{\sigma} = \min_{j=1,\dots,d} \sqrt{\Sigma_{jj}}$. Further, $w_1, \dots, w_n$ denote i.i.d. random variables independent of $X_1, \dots, X_n$. They are used to define the wild bootstrap statistic S_n^* in (1.2). We always assume $\mathrm{E}[w_1] = 0$ and $\mathrm{E}[w_1^2] = 1$. Also, P^* and E^* denote the conditional probability and expectation given the data $X_1, \dots, X_n$, respectively. For $p \in (0, 1)$, $\hat{c}_p$ denotes the conditional p-quantile of T_n^* given the data, i.e. $\hat{c}_p := \inf\{t \in \mathbb{R} : P^*(T_n^* \le t) \ge p\}$.

2.1 *Valid Edgeworth expansion in high dimensions*

We begin by introducing appropriate (second-order) Edgeworth expansions for S_n and S_n^*. The former is standard. That is, our Edgeworth expansion for S_n is defined as

$$p_n(z) = \phi_\Sigma(z) - \frac{1}{6\sqrt{n}} \langle \mathrm{E}[\overline{X^3}], \nabla^3 \phi_\Sigma(z) \rangle, \quad z \in \mathbb{R}^d.$$

The situation is different for the latter. In the low-dimensional setting, a natural bootstrap version of $p_n(z)$ would be obtained by replacing Σ and $\mathrm{E}[\overline{X^3}]$ with their sample counterparts $\widehat{\Sigma}_n$ and $\overline{X^3}$, respectively. However, when $d \ge n$, $\widehat{\Sigma}_n$ is always degenerate, so $\phi_{\widehat{\Sigma}_n}$ is not well-defined. For this reason, we consider an Edgeworth expansion "around ϕ_Σ". Formally, our Edgeworth expansion for S_n^* is defined as

$$\hat{p}_{n,\gamma}(z) = \phi_\Sigma(z) + \frac{1}{2} \langle \overline{X^2} - \Sigma, \nabla^2 \phi_\Sigma(z) \rangle - \frac{\gamma}{6\sqrt{n}} \langle \overline{X^3}, \nabla^3 \phi_\Sigma(z) \rangle, \quad z \in \mathbb{R}^d,$$

where $\gamma \in \mathbb{R}$ is a constant determined by the construction of S_n^*. Typically, $\gamma = 1$ for third-moment matching bootstrap methods.

Next, we formally define the notion of Stein kernel.

Definition 2.1 (Stein kernel)**.** Let ξ be a random vector in $\mathbb{R}^d$ with $\mathrm{E}[\|\xi\|_\infty] < \infty$. A measurable function $\tau : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ is called a *Stein kernel* for (the law of) ξ if $\mathrm{E}[\|\tau(\xi)\|_\infty] < \infty$ and

$$\mathrm{E}[(\xi - \mathrm{E}[\xi]) \cdot \nabla h(\xi)] = \mathrm{E}[\langle \tau(\xi), \nabla^2 h(\xi) \rangle] \tag{2.1}$$

for any $h \in C_b^2(\mathbb{R}^d)$.

The concept of Stein kernel was originally introduced in Stein [58, Lecture VI] for the univariate case. Although its partial multivariate extension dates back to [14], general treatments have started in more recent studies of [44, 56], stemming from the discovery of a connection to Malliavin calculus due to Nourdin & Peccati [54] (the so-called *Malliavin–Stein method*). We refer to [51] for the recent development.

Remark 2.1 (Alternative definition)**.** Our definition of Stein kernel is taken from [44]. In the literature, the definition of Stein kernel often requires (2.1) to hold with ∇h on both sides replaced by any bounded C^1 function $h : \mathbb{R}^d \to \mathbb{R}^d$ with bounded derivatives. Except for the case $d = 1$, this requirement is slightly stronger than ours. Nevertheless, as far as the author knows, this stronger requirement has so far been met by all known constructions of Stein kernels, including all the examples of this paper.

The validity of Edgeworth expansion for S_n is ensured if the summands have Stein kernels:

Theorem 2.1 (Edgeworth expansion for S_n)**.** *Suppose that X_i has a Stein kernel τ_i^X for every $i = 1, \dots, n$. Suppose also that there exists a constant $b > 0$ such that*

$$\max_{1 \le i \le n} \max_{1 \le j \le d} \|X_{ij}\|_{\psi_1} \le b \tag{2.2}$$

and

$$\max_{1 \le i \le n} \max_{1 \le j,k \le d} \|\tau_{i,jk}^X(X_i)\|_{\psi_{1/2}} \le b^2. \tag{2.3}$$

Further, assume $\log^3 d \le n$. Then,

$$\sup_{A \in \mathcal{R}} \left| P(S_n \in A) - \int_A p_n(z) dz \right| \le C \frac{b^5}{\sigma_*^5} \frac{\log^3 d}{n} \log n. \tag{2.4}$$

Remark 2.2. Here and below, we do not intend to optimize the dependence of bounds on b and σ_*.

Remark 2.3 (Proof strategy)**.** One main tool to prove (2.4) and its bootstrap counterpart (2.7) is an *explicit* decomposition formula of $\mathrm{E}[h(S_n)] - \int_{\mathbb{R}^d} h(z) p_n(z) dz$ for a suitable test function $h : \mathbb{R}^d \to \mathbb{R}$ using Stein kernel (see Lemma 4.3). In the univariate case, such a decomposition has recently been derived in [33] for *any* bounded h (see Eqs.(2.10) and (2.11) ibidem). Although their argument can be extended more or less naturally to the multivariate case as long as h is sufficiently smooth, it would generally fail if h is just bounded. This is because when $d \ge 2$, the Stein equation associated with h is a second-order partial differential equation, so its solution is less smooth than in the univariate case (see also [37, Remark 2.7] for a related discussion). For this reason, we prove the decomposition formula with h replaced by a smoothed version $h_t(x) = \mathrm{E}[h(\sqrt{1-t}x + \sqrt{t}Z)]$ with $t \in (0, 1]$. For $h = 1_A$ with $A \in \mathcal{R}$, this formula along with

Lemma 4.4 gives an error bound for $\mathrm{E}[h_t(W)] - \int_{\mathbb{R}^d} h_t(z)p_n(z)dz$ depending (poly-)logarithmically on d and t

Therefore, in contrast to the univariate case, we need an additional argument to make the smoothing error negligible. Conventionally, this is accomplished by the so-called *smoothing inequality* and a suitable estimate for $\sup_A \int_{(\partial A)^t} |p_n(z)|dz$ (called an *anti-concentration inequality*), where the supremum is taken over $\mathcal{R}$ in our case. See e.g. [8, Lemma 11.4] for the former and [8, Corollary 3.2] for the latter, respectively. However, to the author's knowledge, all the existing bounds for $\sup_A \int_{(\partial A)^t} |p_n(z)|dz$ depend polynomially on d and are inadequate for our purpose. To resolve this issue, we establish a novel bound for $\sup_A |\int_{(\partial A)^t} p_n(z)dz|$ which depends only poly-logarithmically on d (see Lemma 4.2). This together with a modified smoothing inequality (Lemma 4.1) shows that the above smoothing error is negligible.

Below we give a few examples satisfying (2.3).

Example 2.1 (Log-concave distribution)**.** Log-concave distributions are prominent examples having Stein kernels with the desired properties. The following result follows from [1, Proposition A.5], [34, Theorem 2.3 and Proposition 3.2] and Lemma D.6.

Lemma 2.1 (Stein kernel of log-concave distribution)**.** *Suppose that a random vector X in $\mathbb{R}^d$ has a log-concave density. Then X has a Stein kernel τ and*

$$\max_{1\leq j\leq d}\|X_j\|_{\psi_1} \leq C \max_{1\leq j\leq d}\sqrt{\mathrm{E}[X_j^2]}, \qquad \max_{1\leq j,k\leq d}\|\tau_{jk}(X)\|_{\psi_{1/2}} \leq C \max_{1\leq j\leq d}\mathrm{E}[X_j^2]$$

for some universal constant $C > 0$.

Example 2.2 (Gaussian copula model)**.** Let R be a $d\times d$ positive semidefinite symmetric matrix with unit diagonals. Also, for every $j = 1,\dots,d$, let μ_j be a non-degenerate probability distribution on $\mathbb{R}$ (i.e. μ_j is not the unit mass at a point), and denote by F_j its distribution function. The Gaussian copula model $U = (U_1,\dots,U_d)^\top$ with parameter matrix R and marginal distributions $\mu_1,\dots,\mu_d$ is defined as $U_j = F_j^{-1}(\Phi(Z_j))$ for $j = 1,\dots,d$, where $Z \sim N(0,R)$.

Proposition 2.1 (Stein kernel of Gaussian copula model)**.** *Suppose that there exists a constant $\kappa > 0$ such that for every $j = 1,\dots,d$ and any Borel set $B \subset \mathbb{R}$,*

$$\liminf_{h\downarrow 0}\frac{\mu_j(B^h) - \mu_j(B)}{h} \geq \kappa \min\{\mu_j(B), 1-\mu_j(B)\}, \tag{2.5}$$

where $B^h := \{t\in\mathbb{R} : |t-s| < h \text{ for some } s\in B\}$. Then $X := U - \mathrm{E}[U]$ has a Stein kernel τ and

$$\max_{1\leq j\leq d}\|X_j\|_{\psi_1} \leq C\kappa^{-1}, \qquad \max_{1\leq j,k\leq d}\|\tau_{jk}(X)\|_{\psi_1} \leq C\kappa^{-2}$$

for some universal constant $C > 0$.

The maximal constant κ satisfying (2.5) is called the *Cheeger (isoperimetric) constant* of μ_j. We refer to [10, Theorem 1.3] for a useful equivalent formulation in the univariate case. When μ_j is log-concave, then (2.5) is satisfied with $\kappa = 1/\sqrt{3\,\mathrm{Var}[X_j]}$ by Proposition 4.1 in [9]. Since the gamma distribution with shape parameter ≥ 1 is log-concave, Proposition 2.1 shows that the simulated model in the introduction satisfies the assumptions of Theorem 2.1. We can actually show that any gamma distribution has a positive Cheeger constant; see Proposition 4.1.

Example 2.3 (Affine transformation)**.** Let ξ be a random vector in $\mathbb{R}^r$ with a Stein kernel τ. Then, a straightforward computation shows that for any $a \in \mathbb{R}^d$ and $d \times r$ matrix V, $V\xi + a$ has a Stein kernel given by $x \mapsto \mathrm{E}[V\tau(\xi)V^\top \mid V\xi + a = x]$.

Example 2.4 (Multiplicative perturbation)**.** Let X be a random vector in $\mathbb{R}^d$ and ϵ a centered random variable independent of X and having a Stein kernel τ. Then ϵX has a Stein kernel $x \mapsto \mathrm{E}[\tau(\epsilon)X^{\otimes 2} \mid \epsilon X = x]$, provided that $\mathrm{E}[\|\epsilon X\|_\infty] + \mathrm{E}[\|\tau(\epsilon)X^{\otimes 2}\|_\infty] < \infty$. This easily follows by applying Example 2.3 conditional on X. This type of random vector arises in high-dimensional regression; see [18, Section 4]. Note that univariate Stein kernels can be written down explicitly and their properties are well-investigated in the literature; see [28, 46] for example.

Example 2.5 (General continuous distribution)**.** Let X be a centered random vector in $\mathbb{R}^d$ with density f. If $\mathrm{E}[|X|^2] < \infty$ and the support of f is a (possibly unbounded) rectangle, we can construct a Stein kernel for X by modifying the construction given in [51, Remark 4.13] as follows (see also the proof of [3, Theorem 4]). For a vector $x \in \mathbb{R}^d$ and $j \in \{1, \dots, d\}$, we set $x_{1:j} := (x_1, \dots, x_j)^\top$ and $x_{j:d} := (x_j, \dots, x_d)^\top$. Also, define a function $g_j : \mathbb{R}^j \to \mathbb{R}$ as $g_j(u) = \int_{\mathbb{R}^{d-j}} f(u, v)dv$. Thanks to the fact that f is supported by a rectangle, we can define a function $\tau_{1j} : \mathbb{R}^d \to \mathbb{R}$ such that

$$\begin{aligned} \tau_{1j}(x)f(x) = g_{j-1}(x_{1:(j-1)}) &\int_{\mathbb{R}^{j-1}\times[x_j,\infty)} u_1 f(u, x_{(j+1):d})du \\ &- \int_{x_j}^{\infty} g_j(x_{1:(j-1)}, u)du \int_{\mathbb{R}^j} u_1 f(u, x_{(j+1):d})du \end{aligned}$$

a.e. x with respect to the Lebesgue measure, where we set $g_0 \equiv 1$ by convention. By construction, it is straightforward to verify that $\sum_{j=1}^d \partial_j(\tau_{1j}f)(x) = -x_1 f(x)$ a.e. x. Moreover, the assumptions $\mathrm{E}[|X|^2] < \infty$ and $\mathrm{E}[X] = 0$ respectively imply $\mathrm{E}[|\tau_{1j}(X)|] < \infty$ and $\tau_{1j}(x)f(x) \to 0$ as $|x_j| \to \infty$ a.e. $(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_d)$. Hence, integration by parts gives $\sum_{j=1}^d \mathrm{E}[\tau_{1j}(X)\partial_j h(X)] = \mathrm{E}[X_1 h(X)]$ for any $h \in C_b^1(\mathbb{R}^d)$. Rotating the indices, we can analogously construct functions $\tau_{ij} : \mathbb{R}^d \to \mathbb{R}$ $(i, j = 1, \dots, d)$ satisfying $\mathrm{E}[|\tau_{ij}(X)|] < \infty$ and $\sum_{j=1}^d \mathrm{E}[\tau_{ij}(X)\partial_j h(X)] = \mathrm{E}[X_i h(X)]$ for any $h \in C_b^1(\mathbb{R}^d)$. Consequently, $(\tau_{ij})_{1\leq i,j\leq d}$ is a Stein kernel for X.

Let us give a sufficient condition to bound $\|\tau_{ij}(X)\|_{\psi_{1/2}}$ (we will actually bound $\|\tau_{ij}(X)\|_{\psi_1}$). For a non-empty subset $I \subset \{1, \dots, d\}$, we denote by f_I the density of $(X_i)_{i\in I}$. Suppose that there exists a constant $K \geq 1$ such that for any disjoint partition $\{1, \dots, d\} = I_1 \cup I_2 \cup I_3$,

$$\frac{1}{K} \leq \frac{f_{I_1}((X_i)_{i\in I_1}) f_{I_2}((X_i)_{i\in I_2}) f_{I_3}((X_i)_{i\in I_3})}{f(X)} \leq K \quad \text{a.s.} \tag{2.6}$$

Also, suppose that the Cheeger constant of the law of X_j is bounded below by $1/K$ for all $j \in \{1, \dots, d\}$ (cf. Example 2.2). Then we have $|\tau_{ij}(X)| \leq 2(K + |X_i| + \mathrm{E}[|X_i|])K^3$ a.s. Hence, there exists a universal constant C such that $\|\tau_{ij}(X)\|_{\psi_1} \leq CK^4$.

Condition (2.6) can be interpreted as a weak dependence condition between the components of X. Though this condition seems somewhat restrictive, it is satisfied e.g. when the copula of X belongs to the Farlie–Gumbel–Morgenstern family with appropriate parameter values (see [53, Example 3.31]).

Remark 2.4 (Relation to classical conditions)**.** (a) In the univariate case, if a non-degenerate distribution has a Stein kernel, then it has a non-zero absolutely continuous part; see [28, Theorem 1.1]. In particular, it must satisfy Cramér's condition.
(b) In the multivariate case, a non-degenerate distribution may not satisfy Cramér's condition even when it has a Stein kernel: A simple example is a multivariate normal distribution with singular covariance matrix. This example is indeed important in the high-dimensional setting when analyzing the Gaussian wild bootstrap.
(c) If a probability distribution has a Stein kernel, its support is convex. This follows from the fact that any one-dimensional marginal has a Stein kernel and thus is supported by an interval according to [28, Theorem 1.1].

As a first application, we derive the exact convergence rate of the coverage error of (infeasible) Gaussian approximation in the spherical case. This will serve as a benchmark. Recall that c_p^G denotes the p-quantile of $Z^\vee$.

Corollary 2.1 (Coverage error of Gaussian approximation in the spherical case)**.** *Under the assumptions of Theorem 2.1, suppose additionally that $\Sigma = \sigma I_d$ for some $\sigma > 0$. Suppose also that $d \to \infty$, $\log^3 d = o(n/\log^2 n)$ and $b/\sigma = O(1)$ as $n \to \infty$. Further, suppose that $(nd)^{-1}\sum_{i=1}^n\sum_{j=1}^d \mathrm{E}[(X_{ij}/\sigma)^3] \to \gamma_X$ as $n \to \infty$. Then, for any $\alpha \in (0,1)$,*

$$\sqrt{\frac{n}{\log^3 d}}\left(P(T_n \geq c_{1-\alpha}^G) - \alpha\right) \to -\frac{\sqrt{2}}{3}\gamma_X \log(1-\alpha) \quad (n \to \infty).$$

We turn to Edgeworth expansion for S_n^*. Its validity is ensured if the weight variables have Stein kernels:

Theorem 2.2 (Edgeworth expansion for S_n^*)**.** *Assume* (2.2)*. Also, suppose that w_1 satisfies either of the following conditions:*

(i) w_1 has a Stein kernel τ^* and there exists a constant $b_w \geq 1$ such that $|w_1| \leq b_w$ and $|\tau^*(w_1)| \leq b_w^2$.
(ii) $w_1 \sim N(0,1)$. We set $b_w = 1$ in this case.

Further, assume $\log^3 d \leq n$. Set $\gamma := \mathrm{E}[w_1^3]$. Then we have

$$\sup_{A\in\mathcal{R}}\left|P^*(S_n^* \in A) - \int_A \hat{p}_{n,\gamma}(z)dz\right| \leq C\frac{b_w^5 b^5}{\sigma_*^5}\frac{\log^3(dn)}{n}\log n \tag{2.7}$$

with probability at least $1 - 1/n$.

We can construct a random variable w_1 satisfying Condition (i) and $\mathrm{E}[w_1^3] = 1$ as follows: Let η be a random variable following the beta distribution with parameters $\alpha, \beta > 0$. Then $w := (\eta - \mathrm{E}[\eta])/\sqrt{\mathrm{Var}[\eta]}$ satisfies (i) by [46, Example 4.9] and Example 2.3. Also, we have

$$\mathrm{E}[w_1^3] = \frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}} = \frac{2(1-2\mu)\sqrt{1+\nu}}{(2+\nu)\sqrt{\mu(1-\mu)}},$$

where $\mu = \alpha/(\alpha+\beta)$ and $\nu = \alpha+\beta$. From this expression, given a positive constant $\nu > 0$, we have $\mathrm{E}[w_1^3] = 1$ if we set

$$\alpha = \nu\frac{c-(2+\nu)\sqrt{c}}{2c}, \quad \beta = \nu\frac{c+(2+\nu)\sqrt{c}}{2c} \quad \text{with } c = \nu^2+20\nu+20. \tag{2.8}$$

A drawback of Theorem 2.2 is that two-point distributions do not admit Stein kernels (cf. [28, Theorem 1]). In particular, it does not cover Mammen's wild bootstrap (cf. Eq.(3.1)) examined in the simulation study of [26]. However, the above standardized beta distribution becomes closer to Mammen's two-point distribution as ν is closer to 0, and their numerical difference virtually vanishes. Our simulation study shows that the beta wild bootstrap with $\nu = 0.1$ performs very similarly to Mammen's one.

2.2 Asymptotic expansion of coverage probability

Next, we develop an asymptotic expansion of the bootstrap coverage probability $P(T_n \geq \hat{c}_{1-\alpha})$ and investigate its implications. Before proceeding, we introduce some notation used throughout this section. For $t \in \mathbb{R}$, we set $A(t) := (-\infty, t]^d$. We denote by f_Σ the density of $Z^\vee$, where $Z \sim N(0,\Sigma)$. Note that f_Σ is a C^∞ function since Σ is invertible. Also, we set $\varsigma_d := \sqrt{\mathrm{Var}[Z^\vee]\log d}$. By Lemma D.3, ς_d is bounded from below by a positive constant depending only on $\overline{\sigma}$ and $\underline{\sigma}$. By Lemma D.1, ς_d is generally bounded by $\overline{\sigma}\sqrt{\log d}$, but we often have $\varsigma_d = O(1)$ as $d \to \infty$, known as a *superconcentration* phenomenon (cf. [16]). For example, this is the case when $\Sigma_{jj} = 1$ for all j and there exists a constant $C > 0$ such that $\Sigma_{jk} \leq C/\log(2+|j-k|)$ for all j,k. This follows from [16, Theorem 9.12].

Theorem 2.3 (Asymptotic expansion of bootstrap coverage probability)**.** *Under the assumptions of Theorems 2.1 and 2.2, let $\lambda > 0$ be a constant such that $b/\sigma_* \leq \lambda$. Then, for any $\varepsilon \in (0,1/2)$, there exist positive constants c and C depending only on λ, ε and b_w such that if*

$$\frac{\varsigma_d^3}{\sigma_*^3}\frac{\log^3(dn)}{n}\log n \leq c, \tag{2.9}$$

then

$$\sup_{\varepsilon<\alpha<1-\varepsilon}\left|P(T_n \geq \hat{c}_{1-\alpha}) - \left(\alpha - (1-\gamma)Q_n(c_{1-\alpha}^G) - \mathrm{E}[R_n(\alpha)]\right)\right| \leq C\frac{\varsigma_d^3}{\sigma_*^3}\frac{\log^3(dn)}{n}\log n,$$

where

$$Q_n(t) := \int_{A(t)}\{p_n(z) - \phi_\Sigma(z)\}dz = -\frac{1}{6\sqrt{n}}\langle \mathrm{E}[\overline{X^3}], \int_{A(t)}\nabla^3\phi_\Sigma(z)dz\rangle, \qquad t \in \mathbb{R},$$

and

$$R_n(\alpha) := \frac{1}{\sqrt{n}}\frac{\left\langle \overline{X^3}\otimes\mathbf{1}_d, \Psi_\alpha^{\otimes 2}\right\rangle}{2f_\Sigma(c_{1-\alpha}^G)}, \qquad \Psi_\alpha := \int_{A(c_{1-\alpha}^G)}\nabla^2\phi_\Sigma(z)dz.$$

Remark 2.5 (Univariate case)**.** When $d=1$ and $\Sigma=1$, the above asymptotic expansion formula reduces to

$$\begin{cases}\alpha - \frac{\mathrm{E}[\overline{X^3}]}{6\sqrt{n}}\{2(c_{1-\alpha}^G)^2+1\}\phi(c_{1-\alpha}^G) & \text{if } \gamma = 0,\\ \alpha - \frac{\mathrm{E}[\overline{X^3}]}{2\sqrt{n}}(c_{1-\alpha}^G)^2\phi(c_{1-\alpha}^G) & \text{if } \gamma = 1.\end{cases}$$

These recover the asymptotic expansion formulae for normal and empirical bootstrap coverage probabilities, respectively; see e.g. [47, Eqs.(2)–(3)] (note that $c_{1-\alpha}^G = \Phi^{-1}(1-\alpha) = -\Phi^{-1}(\alpha)$ when $d=1$).

Remark 2.6 (Proof strategy)**.** A basic strategy to prove Theorem 2.3 is the same as the classical one (cf. Section 3.5.2 in [38]): We replace the "complicated" random variable $\hat{c}_{1-\alpha}$ by its asymptotic expansion (known as *Cornish–Fisher expansion*). After this replacement, it turns out that the proof amounts to Edgeworth expansion for the maximum component of a sum of independent random vectors with (approximate) Stein kernels, which can be handled by a similar strategy to the proof of Theorem 2.1.

A major difficulty specific to our setting arises when we derive an error bound for Cornish–Fisher expansion. This is due to the fact that the "centering" distribution function $F_Z(t) = P(Z^\vee \leq t)$ for our Edgeworth expansions depends on d and may not converge to a non-degenerate distribution function as $d \to \infty$. In particular, bounding derivatives of F_Z^{-1} is non-trivial in our setting. To resolve this issue, we develop a novel isoperimetric-type inequality for $Z^\vee$ suitable for our purpose (see Lemma 5.1). We remark that a Gaussian-type isoperimetric inequality for $Z^\vee$ can be derived from Gaussian isoperimetry, but the dimension dependence is less sharp than ours; see Remark 5.1 for details.

Now we discuss implications of Theorem 2.3. An easy consequence is that any wild bootstrap approximation is second-order accurate when $\mathrm{E}[\overline{X^3}] = 0$ as long as w_1 satisfies the assumptions in Theorem 2.2. However, simulation results suggest that the choice of w_1 would affect the performance even when $\mathrm{E}[\overline{X^3}] = 0$, so there is still room to investigate.

The following corollary gives a more interesting implication:

Corollary 2.2. *Under the assumptions of Theorem 2.3, suppose additionally that* $\mathrm{E}[w_1^3] = 1$, $\varepsilon \geq 2e^{-d/2}$, $\underline{\sigma} = \overline{\sigma} =: \sigma$ *and the maximum eigenvalue of* Σ *is bounded by* $K\sigma^2$ *with some constant* $K > 0$*. Then there exist a constant* $C > 0$ *depending only on* λ, ε, K *and* b_w *such that*

$$\sup_{\varepsilon<\alpha<1-\varepsilon} |P(T_n \geq \hat{c}_{1-\alpha}) - \alpha| \leq C\left(\frac{\varsigma_d^3}{\sigma_*^3}\frac{\log^3(dn)}{n}\log n + \frac{\varsigma_d}{\sigma}\sqrt{\frac{\log^3 d}{dn}}\right). \tag{2.10}$$

Observe that the second term on the right hand side of (2.10) is divided by $\sqrt{d}$. Hence, Corollary 2.2 implies that the third-moment matching wild bootstrap is second-order accurate if $d \geq n$ and Σ has identical diagonal entries and bounded eigenvalues with respect to d. This seems to be a new result on the blessing of dimensionality, although excessively high dimensionality is harmful due to the first term of the bound. We also remark that such an improvement generally does not occur without third-moment matching:

Corollary 2.3. *Under the assumptions of Corollary 2.1, if* w_i *satisfies the conditions in Theorem 2.2 with* $b_w = O(1)$*, then for any* $\alpha \in (0, 1)$*,*

$$\sqrt{\frac{n}{\log^3 d}}\left(P(T_n \geq \hat{c}_{1-\alpha}) - \alpha\right) \to -(1 - \mathrm{E}[w_1^3])\frac{\sqrt{2}}{3}\gamma_X \log(1-\alpha) \quad (n \to \infty).$$

Remark 2.7. The additional assumptions on Σ in Corollary 2.2 are necessary to ensure that T_n is truly high-dimensional:

(a) The assumption $\underline{\sigma} = \overline{\sigma}$ ensures that no single component solely affects variations of T_n. Formally, this assumption is mainly used to prove $P(Z_j > c_{1-\alpha}^G) = O(d^{-1}\sqrt{\log d})$ as $d \to \infty$ for every j (corresponding to the second inequality of (5.28)).

(b) The boundedness of the maximum eigenvalue of Σ rules out situations where the components of X_i have a common factor. In such a situation, the common factor solely affects variations of T_n. In fact, we obtain the following corollary when this is the case:

Corollary 2.4. *Under the assumptions of Theorem 2.3, suppose additionally that $X_1, \dots, X_n$ are i.i.d. and each has the same law as $\sqrt{\rho}U\mathbf{1}_d + \sqrt{1-\rho}V$, where $\rho \in (0,1)$ is a constant, U is a random variable with mean 0 and variance 1, and V is a random vector in $\mathbb{R}^d$ independent of U with* $\mathrm{E}[V] = 0$ *and* $\mathrm{Cov}[V] = I_d$. *Moreover, suppose that ρ and U do not depend on d, n. Then, for any $\alpha \in (0,1)$,*

$$\begin{aligned} P(T_n \geq \hat{c}_{1-\alpha}) - \alpha = \frac{\mathrm{E}[U^3]}{\sqrt{n}} \left(\frac{\gamma-1}{6}(z_\alpha^2 - 1) + \frac{z_\alpha^2}{2} \right) \phi(z_\alpha) \\ + (1-\rho)^{\frac{3}{2}} \Upsilon_{n,\gamma}(\alpha) + o(n^{-1/2}) \end{aligned} \tag{2.11}$$

as $d, n \to \infty$, provided that $b = O(1)$ and $\log^9 d = o(n/\log^2 n)$. Here, we set $z_\alpha := \Phi^{-1}(\alpha)$ and

$$\Upsilon_{n,\gamma}(\alpha) := -\frac{1-\gamma}{6\sqrt{n}} \langle \mathrm{E}[V^{\otimes 3}], \int_{A(c^G_{1-\alpha})} \nabla^3 \phi_\Sigma(z) dz \rangle + \frac{1}{2\sqrt{n}} \frac{\langle \mathrm{E}[V^{\otimes 3}] \otimes \mathbf{1}_d, \Psi_\alpha^{\otimes 2} \rangle}{f_\Sigma(c^G_{1-\alpha})}.$$

Moreover, if $\gamma = 1$, then $\Upsilon_{n,\gamma}(\alpha) = o(n^{-1/2})$ as $d, n \to \infty$.

If $\mathrm{E}[V^{\otimes 3}] = 0$ in Corollary 2.4, the asymptotic expansion formula (2.11) has the same form as in the univariate case (cf. Remark 2.5). Therefore, in such a situation the third-moment matching wild bootstrap underperforms the Gaussian wild bootstrap if $|z_\alpha| > 1$ according to [47, Section 3]. We expect that a similar phenomenon would occur when ρ is close to 1; see Section 3 for numerical evidence. On the other hand, Corollary 2.4 also shows that the coverage error of the third-moment matching wild bootstrap is always of order $O(n^{-1/2})$ regardless of the value of $\mathrm{E}[V^{\otimes 3}]$. The simulation results in Section 3 suggest that the Gaussian wild bootstrap would not enjoy this property because its performance in Design (I) significantly worsens when the value of ρ is close to 0.

2.3 Double wild bootstrap

As mentioned in the introduction, the lack of second-order accuracy in standard bootstrap methods is due to the fact that T_n is not asymptotically pivotal. If we knew the distribution function of T_n, say F_n, then $F_n(T_n)$ would give an (exactly) pivotal statistic. Beran [7] suggested estimating F_n by the bootstrap distribution function $\hat{F}_n(t) = P^*(T_n^* \leq t)$ and use $\hat{F}_n(T_n)$ to construct critical values. This method is called *bootstrap prepivoting*. Note that $\hat{F}_n$ can be computed by simulating the conditional law of T_n^* given the data. To estimate the law of $\hat{F}_n(T_n)$, we use the following nested double wild bootstrap procedure following [7]: Let $v_1, \dots, v_n$ be i.i.d. variables independent of everything else and such that $\mathrm{E}[v_1] = 0$ and $\mathrm{E}[v_1^2] = 1$. We define the wild bootstrap statistic of S_n^* as

$$S_n^{**} = \frac{1}{\sqrt{n}} \sum_{i=1}^n v_i (X_i^* - \bar{X}^*), \quad \text{where } X_i^* = w_i(X_i - \bar{X}),\ \bar{X}^* = \frac{1}{n}\sum_{i=1}^n X_i^*.$$

Then define $\hat{F}_n^*(t) = P^{**}(T_n^{**} \leq t)$ for $t \in \mathbb{R}$, where $T_n^{**} := \max_{1\leq j\leq d} S_{n,j}^{**}$ and P^{**} is the conditional probability given $X_1, \dots, X_n, w_1, \dots, w_n$. We regard $\hat{F}_n^*(T_n^*)$ as a bootstrap version of $\hat{F}_n(T_n)$ and estimate the law of $\hat{F}_n(T_n)$ by the conditional law of $\hat{F}_n^*(T_n^*)$. Formally, given a significance level $\alpha \in (0,1)$, let $\hat{\beta}_\alpha$ be the conditional $(1-\alpha)$-quantile of $\hat{F}_n^*(T_n^*)$ given the data. We expect that $P(\hat{F}_n(T_n) \geq \hat{\beta}_\alpha) = P(T_n \geq \hat{c}_{\hat{\beta}_\alpha})$ would be close to α. This is formally justified by the following theorem:

Theorem 2.4 (Second-order accuracy of double bootstrap coverage probability)**.** *Suppose that the assumptions of Theorem 2.3 are satisfied. Suppose also that* v_1 *has a Stein kernel* τ^{**} *and there exists a constant* $b_v \geq 1$ *such that* $|v_1| \leq b_v$ *and* $|\tau^{**}(v_1)| \leq b_v^2$*. Further, assume* $\mathrm{E}[w_1^3] = \mathrm{E}[v_1^3] = 1$*. Then, for any* $\varepsilon \in (0,1/4)$*, there exists a constant* $C > 0$ *depending only on* $\lambda, \varepsilon, b_w$ *and* b_v *such that*

$$\sup_{2\varepsilon<\alpha<1-2\varepsilon} \left| P\left(T_n \geq \hat{c}_{\hat{\beta}_\alpha}\right) - \alpha \right| \leq C \frac{\varsigma_d^4}{\sigma_*^4} \frac{\log^3(dn)}{n} \log n. \tag{2.12}$$

Remark 2.8 (Proof strategy)**.** A key fact to prove Theorem 2.4 is that the coverage probability for the second-level bootstrap enjoys the same asymptotic expansion formula as $P(T_n \geq \hat{c}_{1-\alpha})$ with high probability; see Lemma 5.5. Equivalently, the distribution functions of $\hat{F}_n(T_n)$ and $\hat{F}_n^*(T_n^*)$ share the same asymptotic expansion formula with high probability. This enables us to prove (2.12) by a standard argument to prove the accuracy of bootstrap approximation.

Remark 2.9 (p-value)**.** One can easily check that $T_n \geq \hat{c}_{\hat{\beta}_\alpha}$ is equivalent to $P^*(\hat{p}_n^* \leq \hat{p}_n) \leq \alpha$, where $\hat{p}_n := 1 - \hat{F}_n(T_n)$ and $\hat{p}_n^* := 1 - \hat{F}_n^*(T_n^*)$ are the p-values of the first- and second-level bootstraps, respectively. Hence the p-value of the double bootstrap method is $P^*(\hat{p}_n^* \leq \hat{p}_n)$.

Remark 2.10 (Computational cost)**.** It is well-known that the double bootstrap often entails a prohibitively high computational cost, as it requires nested resampling of the data (cf. [38, Section 3.11.1]). Specifically, if B_1 and B_2 denote the numbers of the first- and second-level bootstrap replications, respectively, a total of $1 + B_1 + B_1 B_2$ test statistics must be computed. For this reason, several authors have investigated ways to reduce its computational burden. For instance, in the context of constructing confidence intervals, Lee & Young [45] showed that the coverage error due to the Monte Carlo approximation in the double bootstrap is of order $o(B_1^{-1/2}) + O(B_2^{-1})$, suggesting that B_2 may be chosen much smaller than B_1. Moreover, for a fixed confidence level, Nankervis [52] proposed stopping rules for the second-level bootstrap to avoid unnecessary computation that does not affect the resulting confidence interval. As an alternative approach, Davidson & MacKinnon [25] proposed the *fast double bootstrap (FDB)*, which requires only a single second-level bootstrap sample for each first-level one. While the FDB dramatically reduces the computational burden relative to the standard double bootstrap, it does not necessarily improve the order of accuracy in approximating coverage probabilities from a theoretical point of view (see Chang & Hall [15]).

3 Simulation study

This section conducts a small Monte Carlo study to supplement our theoretical findings. We adopt a similar simulation design to [26]: We set $d = 400$ and generate the data from a Gaussian copula model, i.e. $X_1, \dots, X_n$ are i.i.d. with the same law as $U - \mathrm{E}[U]$, where U is defined as in Example 2.2. The marginal distributions μ_j are the gamma distribution with shape parameter 1 and unit scale. As the parameter matrix R, we consider two designs: (I) $R = \rho \mathbf{1}_d^{\otimes 2} + (1-\rho) I_d$ and (II) $R = (\rho^{|j-k|})_{1 \leq j,k \leq d}$. Here, the parameter ρ

is varied as $\rho \in \{0.2, 0.8\}$. We also vary the sample size as $n \in \{200, 400\}$. We compute the rejection rate $P(T_n \geq \hat{c})$ at the 10% significance level based on 20,000 Monte Carlo iterations, where $\hat{c}$ is an estimated 90% quantile of the corresponding statistic using various bootstrap methods. In addition, to assess the performance when the skewness of the data is zero, we also consider the case that $X_i \stackrel{d}{=} U - U'$, where U' is an independent copy of U. To keep the marginal kurtosis at the same level, we change the shape parameter of the gamma distribution to 0.5 in this case. We remark that these models satisfy (2.2) and (2.3) thanks to Proposition 2.1.

For the bootstrap methods, we consider the empirical bootstrap (EB), wild bootstrap and double wild bootstrap (DB) methods. For the wild bootstrap, we consider the following 4 types of weight variables:

GB w_1 is a standard normal variable.
MB w_1 follows Mammen's two point distribution [50]:

$$P\left(w_1 = \frac{\sqrt{5}+1}{2}\right) = 1 - P\left(w_1 = -\frac{\sqrt{5}-1}{2}\right) = \frac{\sqrt{5}-1}{2\sqrt{5}}. \tag{3.1}$$

RB w_1 is a Rademacher variable: $P(w_1 = \pm 1) = 1/2$.
BB w_1 follows the standardized beta distribution with parameters given by (2.8) with $\nu = 0.1$.

The double wild bootstrap is implemented with both w_1 and v_1 generated from the standardized beta distribution with parameters given by (2.8) with $\nu = 0.1$. Note that our theoretical results are applicable only to GB, BB and DB. We include EB, MB and RB in our assessment because they are commonly used in the literature. The number of bootstrap replications is set to 499 for the first-level bootstrap and 99 for the second-level bootstrap in DB (see [38, Appendix II.8] and [45] for discussions about the number of bootstrap replications).

We summarize the simulation results in Tables 1 and 2. First, Table 1 reports empirical rejection rates at the 10% level when the laws of X_i are asymmetric. We find that the difference in performance between GB and BB is largely in line with our Corollaries 2.2 and 2.4: First, in Design (I), GB outperforms BB when $\rho = 0.8$, while BB slightly outperforms GB when $\rho = 0.2$. This is consistent with Corollary 2.4: When ρ is close to 1, the first term on the right hand side of (2.11) dominates the coverage error, and this term has the same form as the univariate case. According to computations in [47, Section 3], this implies that the coverage error of GB would be smaller than that of BB. On the other hand, when ρ is closer to 0, the second term on the right hand side of (2.11) would be non-negligible for GB, so the above argument would be invalid in this case. Next, in Design (II), BB performs much better than GB, which is consistent with Corollary 2.2: Since Σ has no large eigenvalues in this design, BB is second-order accurate by Corollary 2.2.

Turning to the performance of DB, it uniformly outperforms the other methods when $n = 400$. It tends to over-reject when $n = 200$, but it still outperforms the others in Design (I). In Design (II), BB outperforms DB when $n = 200$. This would be because BB is also second-order accurate in this design *thanks to the high dimensionality* by Corollary 2.2. Since our theory suggests that an increase in the sample size n would improve the performance of DB more than the others, it is expected that DB performs much better when $n = 400$.

Next, Table 2 reports empirical rejection rates at the 10% level when the laws of X_i are symmetric. Recall that our Theorem 2.3 implies that both GB and BB are second-order accurate in this case. Reflecting this fact, GB clearly performs better than in the asymmetric case. The performance of BB is improved in Design (I) but not in Design (II). The latter would be due to the same reasoning as above, i.e. BB is second-order accurate in Design (II) even when the skewness is not zero by Corollary 2.2. By contrast, the performance of DB is not improved. This is not surprising because DB is already second-order accurate in the asymmetric case and the zero skewness condition would not contribute to its performance. When comparing GB and BB, BB still outperforms GB. This may be due to an effect of kurtosis, but we will need higher-order asymptotic expansions for the formal discussion and leave it to future work.

Finally, we briefly discuss the performances of EB, MB and RB. First, EB tends to under-reject and its performance is not pronounced compared to other methods. In fact, we can observe similar phenomena in the simulation results of [22, 26]. Formally, this does not contradict our theory because we have no valid Edgeworth expansion for EB in high-dimensions, although it is unclear whether this is an artifact of our proof strategy. Next, although MB is not covered by our theory, its performance is similar to BB. This is perhaps explained by the fact that their weights are very close numerically. Third, RB performs remarkably well in the symmetric case. This is already observed in the simulation study of [22] who explain this phenomenon by their Theorem 2.3. Another possible explanation is the match of higher moments, but we have no formal theoretical result for this so far.

Table 1: Rejection rate at the 10% level (Asymmetric case)

n	ρ	EB	GB	MB	RB	BB	DB
		(I) $R_{jk} = \rho + (1-\rho)1_{\{j=k\}}$					
200	0.2	0.061	0.124	0.080	0.155	0.078	0.114
	0.8	0.071	0.090	0.072	0.093	0.071	0.101
400	0.2	0.072	0.122	0.083	0.140	0.085	0.107
	0.8	0.080	0.095	0.081	0.096	0.080	0.102
		(II) $R_{jk} = \rho^{\lvert j-k \rvert}$					
200	0.2	0.065	0.146	0.092	0.195	0.091	0.117
	0.8	0.069	0.139	0.089	0.177	0.088	0.113
400	0.2	0.074	0.135	0.090	0.160	0.088	0.101
	0.8	0.079	0.135	0.092	0.155	0.091	0.104

4 Proofs for Section 2.1

We use the following notation in the remainder of the paper: For two random variables ξ and η, we write $\xi \lesssim \eta$ or $\eta \gtrsim \xi$ if there exists a *universal* constant $C > 0$ such that $\xi \leq C\eta$. Also, given real numbers $\theta_1, \dots, \theta_m$, we use $C_{\theta_1,\dots,\theta_m}$ to denote positive constants, which depend only on $\theta_1, \dots, \theta_m$ and may be different in different expressions.

Table 2: Rejection rate at the 10% level (Symmetric case)

n	ρ	EB	GB	MB	RB	BB	DB
		(I) $R_{jk} = \rho + (1-\rho)1_{\{j=k\}}$					
200	0.2	0.065	0.076	0.083	0.100	0.082	0.114
	0.8	0.089	0.092	0.091	0.096	0.091	0.105
400	0.2	0.081	0.088	0.090	0.099	0.092	0.105
	0.8	0.095	0.099	0.098	0.099	0.095	0.103
		(II) $R_{jk} = \rho^{\lvert j-k \rvert}$					
200	0.2	0.062	0.071	0.085	0.101	0.084	0.114
	0.8	0.067	0.076	0.088	0.100	0.086	0.109
400	0.2	0.079	0.084	0.092	0.103	0.092	0.106
	0.8	0.083	0.087	0.096	0.102	0.094	0.106

4.1 General error bounds via approximate Stein kernel

We first develop error bounds for high-dimensional Edgeworth expansion in a somewhat general form. For later use, we consider a situation where a Stein kernel exists in an approximate sense:

Definition 4.1 (Approximate Stein kernel)**.** Let ξ be a random vector in $\mathbb{R}^d$ with $\mathrm{E}[\|\xi\|_\infty] < \infty$. Given measurable functions $\tau : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ and $\beta : \mathbb{R}^d \to \mathbb{R}^d$, we call (τ, β) an *approximate Stein kernel* for (the law of) ξ if $\mathrm{E}[\|\tau(\xi)\|_\infty] + \mathrm{E}[\|\beta(\xi)\|_\infty] < \infty$ and

$$\mathrm{E}[(\xi - \mathrm{E}[\xi]) \cdot \nabla h(\xi)] = \mathrm{E}[\langle \tau(\xi), \nabla^2 h(\xi) \rangle] + \mathrm{E}[\beta(\xi) \cdot \nabla h(\xi)]$$

for any $h \in C_b^2(\mathbb{R}^d)$.

With this definition, we have the following results:

Theorem 4.1 (Error bound for Edgeworth expansion via approximate Stein kernel)**.** *Let $\xi_1, \dots, \xi_n$ be independent random vectors in $\mathbb{R}^d$ with mean 0 and finite variance. Set $W := \sum_{i=1}^n \xi_i$ and $\Sigma_W := \mathrm{Cov}[W]$. For each $i = 1, \dots, n$, suppose that ξ_i has an approximate Stein kernel (τ_i, β_i) such that $\mathrm{E}[\|\xi_i\|_\infty^5] + \mathrm{E}[\|\tau_i(\xi_i)\|_\infty^{5/2}] + \mathrm{E}[\|\beta_i(\xi_i)\|_\infty^{5/2}] < \infty$ for all $i = 1, \dots, n$. Set $T = \sum_{i=1}^n \tau_i(\xi_i)$, $\bar{T} = T - \Sigma$, $B = \sum_{i=1}^n \beta_i(\xi_i)$ and*

$$p_W(z) = \phi_\Sigma(z) + \frac{1}{2}\langle \Sigma_W - \Sigma, \nabla^2 \phi_\Sigma(z) \rangle - \frac{1}{6}\langle \mathrm{E}[W^{\otimes 3}], \nabla^3 \phi_\Sigma(z) \rangle, \quad z \in \mathbb{R}^d. \tag{4.1}$$

Then, there exists a universal constant $C > 0$ such that for any $t \in (0, 1/2]$,

$$\sup_{A \in \mathcal{R}} \left| P(W \in A) - \int_A p_W(z) dz \right|$$
$$\leq C|\log t| \frac{\log^2 d}{\sigma_*^4} \left(\mathrm{E}\,\|\bar{T}\|_\infty^2 + \mathrm{E}\left\| \sum_{i=1}^n \tau_i(\xi_i)^{\otimes 2} \right\|_\infty + \mathrm{E}\left\| \sum_{i=1}^n \xi_i^{\otimes 4} \right\|_\infty \right)$$

$$
\begin{aligned}
&+ C\frac{\log^{5/2} d}{\sigma_*^5}\left(\mathrm{E}\left\|\bar{T}\otimes\sum_{i=1}^n \xi_i^{\otimes 3}\right\|_\infty + \mathrm{E}\left\|\sum_{i=1}^n \xi_i^{\otimes 3}\otimes\tau_i(\xi_i)\right\|_\infty\right)\\
&+ C\frac{\log^2 d}{\sigma_*^4}\left(\mathrm{E}\left\|B\otimes\sum_{i=1}^n \xi_i^{\otimes 3}\right\|_\infty + \mathrm{E}\left\|\sum_{i=1}^n \xi_i^{\otimes 3}\otimes\beta_i(\xi_i)\right\|_\infty\right.\\
&\qquad\qquad \left. + \mathrm{E}\left\|\bar{T}\otimes\sum_{i=1}^n \beta_i(\xi_i)\otimes\xi_i\right\|_\infty + \mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)\otimes\xi_i\otimes\tau_i(\xi_i)\right\|_\infty\right)\\
&+ C\frac{\log d}{\sigma_*^2}\left(\mathrm{E}\left\|B\right\|_\infty^2 + \mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)^{\otimes 2}\right\|_\infty\right)\\
&+ C\frac{\log^{3/2} d}{\sigma_*^3}\left(\mathrm{E}\left\|B\otimes\sum_{i=1}^n \beta_i(\xi_i)\otimes\xi_i\right\|_\infty + \mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)^{\otimes 2}\otimes\xi_i\right\|_\infty\right)\\
&+ C\overline{\sigma}\sqrt{t}\left(\frac{\log d}{\underline{\sigma}} + \frac{\log^2 d}{\sigma_*^3}\|\Sigma_W-\Sigma\|_\infty + \frac{\log^{5/2} d}{\sigma_*^4}\left\|\sum_{i=1}^n \mathrm{E}[\xi_i^{\otimes 3}]\right\|_\infty\right). \qquad (4.2)
\end{aligned}
$$

Theorem 4.2 (Refined Gaussian comparison inequality)**.** *Let W be a centered Gaussian vector in $\mathbb{R}^d$ with covariance matrix Σ_W. Then, there exists a universal constant $C > 0$ such that for any $t \in (0, 1/2]$,*

$$
\begin{aligned}
&\sup_{A\in\mathcal{R}}\left|P(W\in A) - \int_A p_W(z)dz\right|\\
&\le C|\log t|\frac{\log^2 d}{\sigma_*^4}\|\Sigma_W-\Sigma\|_\infty^2 + C\overline{\sigma}\sqrt{t}\left(\frac{\log d}{\underline{\sigma}} + \frac{\log^2 d}{\sigma_*^3}\|\Sigma_W-\Sigma\|_\infty\right),
\end{aligned}
$$

where p_W is defined by (4.1)*; note that $\mathrm{E}[W^{\otimes 3}] = 0$ in the present case.*

The second result will be used for the Gaussian wild bootstrap.

The remainder of this subsection is devoted to the proofs of Theorems 4.1 and 4.2. As usual, the proof starts with a smoothing inequality. We will use the following version.

Lemma 4.1. *Let μ be a finite measure, ν a finite signed measure, and K a probability measure on $\mathbb{R}^d$. Let $\varepsilon > 0$ be a constant such that $\alpha := K([-\varepsilon,\varepsilon]^d) > 1/2$. Let $h : \mathbb{R}^d \to \mathbb{R}$ be a bounded measurable function. Then we have*

$$
\left|\int hd(\mu-\nu)\right| \le (2\alpha-1)^{-1}[\gamma^*(h;\varepsilon) + \tau^*(h;\varepsilon) + \alpha\tilde{\tau}^*(h;\varepsilon)],
$$

where

$$
\gamma^*(h;\varepsilon) = \sup_{y\in\mathbb{R}^d}\gamma(h_y;\varepsilon), \qquad \tau^*(h;\varepsilon) = \sup_{y\in\mathbb{R}^d}\tau(h_y;\varepsilon), \qquad \tilde{\tau}^*(h;\varepsilon) = \sup_{y\in\mathbb{R}^d}\tilde{\tau}(h_y;\varepsilon),
$$

with $h_y(x) = h(x+y)$,

$$
\begin{aligned}
\gamma(h;\varepsilon) &= \max\left\{\int M_h(x;\varepsilon)(\mu-\nu)*K(dx), -\int m_h(x;\varepsilon)(\mu-\nu)*K(dx)\right\},\\
\tau(h;\varepsilon) &= \max\left\{\int [M_h(x;\varepsilon)-h(x)]\nu(dx), \int [h(x)-m_h(x;\varepsilon)]\nu(dx)\right\},
\end{aligned}
$$

$$\tilde{\tau}(h;\varepsilon) = \sup_{y\in[-\varepsilon,\varepsilon]^d} \left| \int [h(x+y) - h(x)]\nu(dx) \right|,$$

$$M_h(x;\varepsilon) = \sup_{y:\|y-x\|_\infty \le \varepsilon} h(y), \qquad m_h(x;\varepsilon) = \inf_{y:\|y-x\|_\infty \le \varepsilon} h(y),$$

and $*$ *denotes the convolution of two finite signed measures.*

The proof of this lemma is a straightforward modification of [8, Lemma 11.4] and given in Appendix B.3, but its statement contains an important difference from the original one: The bound does not contain the positive part of the signed measure ν. This is important for bounding $\tau^*(h;\varepsilon)$ and $\tilde{\tau}^*(h;\varepsilon)$ in our setting. To bound these quantities, we will use the following anti-concentration inequality. For $A = \prod_{j=1}^d [a_j, b_j] \in \mathcal{R}$ and $u, v \in \mathbb{R}_+^d := [0,\infty)^d$, define $A^{u,v} = \prod_{j=1}^d [a_j - u_j, b_j + v_j]$.

Lemma 4.2. *Let* $r \in \mathbb{N}$. *Then*

$$\sup_{A\in\mathcal{R}} \sup_{\varepsilon>0} \sup_{u,v\in[0,\varepsilon]^d} \frac{1}{\varepsilon} \left\| \int_{A^{u,v}\setminus A} \nabla^r \phi_\Sigma(z)dz \right\|_1 \le C_r \frac{(\log d)^{(r+1)/2}}{\sigma_*^{r+1}}, \tag{4.3}$$

where $C_r > 0$ *is a constant depending only on* r.

Proof. See Appendix B.1. □

We will apply Lemma 4.1 with $K = N(0, t\Sigma)$. To bound the quantity $\gamma^*(h;\varepsilon)$, we introduce some notation and lemmas. Given a bounded measurable function $h : \mathbb{R}^d \to \mathbb{R}$ and $s \in [0,1]$, we define a function $h_s : \mathbb{R}^d \to \mathbb{R}$ as

$$h_s(x) = \mathrm{E}[h(\sqrt{1-s}x + \sqrt{s}Z)], \qquad x \in \mathbb{R}^d, \tag{4.4}$$

where $Z \sim N(0,\Sigma)$. When $s > 0$, $h_s(x)$ can be rewritten as

$$h_s(x) = s^{-d/2} \int_{\mathbb{R}^d} h(z)\phi_\Sigma\left(\frac{z - \sqrt{1-s}x}{\sqrt{s}}\right) dz.$$

By this expression, h_s is infinitely differentiable and

$$\nabla^r h_s(x) = \left(-\sqrt{\frac{1-s}{s}}\right)^r \int_{\mathbb{R}^d} h(\sqrt{1-s}x + \sqrt{s}z)\nabla^r \phi_\Sigma(z)dz \tag{4.5}$$

for any $r \in \mathbb{N}$. In particular, $h_s \in C_b^\infty(\mathbb{R}^d)$. We will use the following lemmas to bound $\gamma^*(h;\varepsilon)$.

Lemma 4.3. *Let* $h : \mathbb{R}^d \to \mathbb{R}$ *be a bounded measurable function and* $t \in (0,1]$. *Then, under the assumptions of Theorem 4.1,*

$$\begin{aligned}
&\mathrm{E}[h_t(W)] - \int_{\mathbb{R}^d} h_t(z)p_W(z)dz \\
&= \frac{1}{4}\int_t^1 \left(\int_s^1 m_u^4 \left(\mathrm{E}[\langle \bar{T}^{\otimes 2}, \nabla^4 h_u(W)\rangle] - \sum_{i=1}^n \mathrm{E}[\langle \tau_i(\xi_i)^{\otimes 2}, \nabla^4 h_u(W)\rangle] \right) du \right) ds \\
&\quad - \frac{1}{8}\int_t^1 \left(\int_s^1 m_u^4 \sum_{i=1}^n \mathrm{E}[\langle \tau_i(\xi_i) \otimes \xi_i^{\otimes 2}, \nabla^4 h_u(W)\rangle] du \right) ds
\end{aligned}$$

$$+\frac{1}{16}\int_t^1\left(\int_s^1 m_u\left(\int_u^1 m_v^5\sum_{i=1}^n \mathrm{E}[\langle\xi_i^{\otimes 3}\otimes(\xi_i+B-\beta_i(\xi_i)),\nabla^4 h_v(W)\rangle]dv\right)du\right)ds$$
$$+\frac{1}{16}\int_t^1\left(\int_s^1 m_u\left(\int_u^1 m_v^5\sum_{i=1}^n \mathrm{E}[\langle\xi_i^{\otimes 3}\otimes(\bar{T}-\tau_i(\xi_i)),\nabla^5 h_v(W)\rangle]dv\right)du\right)ds$$
$$+\frac{1}{4}\int_t^1 m_s\left(\int_s^1 m_u^2\left(\mathrm{E}[\langle B^{\otimes 2},\nabla^2 h_u(W)\rangle]-\sum_{i=1}^n \mathrm{E}[\langle\beta_i(\xi_i)^{\otimes 2},\nabla^2 h_u(W)\rangle]\right)du\right)ds$$
$$+\frac{1}{4}\int_t^1 m_s\left(\int_s^1 (m_u^3+m_u^4)\sum_{i=1}^n \mathrm{E}[\langle\beta_i(\xi_i)\otimes(\bar{T}-\tau_i(\xi_i)),\nabla^3 h_u(W)\rangle]du\right)ds$$
$$-\frac{1}{8}\int_t^1 m_s\left(\int_s^1 m_u^4\sum_{i=1}^n \mathrm{E}[\langle\beta_i(\xi_i)\otimes\xi_i^{\otimes 2},\nabla^3 h_u(W)\rangle]du\right)ds$$
$$+\frac{1}{8}\int_t^1 m_s\left(\int_s^1 m_u\left(\int_u^1 m_v^4\sum_{i=1}^n \mathrm{E}[\langle\beta_i(\xi_i)\otimes\xi_i\otimes(\bar{T}-\tau_i(\xi_i)),\nabla^4 h_v(W)\rangle]dv\right)du\right)ds$$
$$+\frac{1}{8}\int_t^1 m_s\left(\int_s^1 m_u\left(\int_u^1 m_v^4\sum_{i=1}^n \mathrm{E}[\langle\beta_i(\xi_i)\otimes\xi_i\otimes(\xi_i+B-\beta(\xi_i)),\nabla^3 h_v(W)\rangle]dv\right)du\right)ds, \quad (4.6)$$

where $m_s=1/\sqrt{1-s}$ *for* $s\in[0,1)$. *Also, under the assumptions of Theorem 4.2,*

$$\mathrm{E}[h_t(W)]-\int_{\mathbb{R}^d} h_t(z)p_W(z)dz=\frac{1}{4}\int_t^1\left(\int_s^1 m_u^4\,\mathrm{E}[\langle(\Sigma_W-\Sigma)^{\otimes 2},\nabla^4 h_u(W)\rangle]du\right)ds. \quad (4.7)$$

Proof. See Appendix B.2. □

Lemma 4.4. *Let* $h=1_A$ *with* $A\in\mathcal{R}$. *Then, for any* $s\in(0,1)$ *and* $r\in\mathbb{N}$,

$$\sup_{x\in\mathbb{R}^d}\|\nabla^r h_s(x)\|_1\le C_r\left(\frac{1-s}{\sigma_*^2 s}\log d\right)^{r/2}, \quad (4.8)$$

where $C_r>0$ *is a constant depending only on* r.

Proof. Observe that $\{z\in\mathbb{R}^d:\sqrt{1-s}x+\sqrt{s}z\in A\}\in\mathcal{R}$ for any $x\in\mathbb{R}^d$. Thus, the claim follows from (4.5) and Lemma D.4. □

Proof of Theorem 4.1. Without loss of generality, we may assume that W and Z are independent. We are going to apply Lemma 4.1 to μ,ν and K defined as

$$\mu(A)=P(\sqrt{1-t}W\in A),\quad \nu(A)=\int_{\mathbb{R}^d}1_A(\sqrt{1-t}z)p_W(z)dz,\quad K(A)=P(\sqrt{t}Z\in A).$$

They are chosen so that $\sqrt{1-t}W+\sqrt{t}Z\sim\mu*K$ and

$$\nu*K(A)=\int_{\mathbb{R}^d}\mathrm{E}[1_A(\sqrt{1-t}z+\sqrt{t}Z)]p_W(z)dz$$

for every $A\in\mathcal{R}$. Since

$$P(\|Z\|_\infty>\overline{\sigma}\sqrt{2\log(2d)})\le\sum_{j=1}^d P(|Z_j|>\sigma_j\sqrt{2\log(2d)})\le\frac{d}{2}e^{-\log(2d)}=\frac{1}{4},$$

we have $\alpha := K([-\varepsilon,\varepsilon]^d) \geq 3/4 > 1/2$ with $\varepsilon = \overline{\sigma}\sqrt{2t\log(2d)}$. Let $h = 1_A$ with $A = \prod_{j=1}^d [a_j, b_j] \in \mathcal{R}$. Then we have $M_h(x;\varepsilon) = 1_{A^\varepsilon}(x)$ and $m_h(x;\varepsilon) = 1_{A^{-\varepsilon}}(x)$, where we set $A^r := \prod_{j=1}^d [a_j - r, b_j + r]$ for any $r \in \mathbb{R}$ with interpreting $[a,b] = \emptyset$ if $a > b$. Hence

$$\gamma(h;\varepsilon) \leq \sup_{h=1_A, A\in\mathcal{R}} \left| \mathrm{E}[h_t(W)] - \int h_t(z) p_W(z) dz \right| \tag{4.9}$$

and

$$\int [M_h(x;\epsilon) - h(x)]\nu(dx) \leq \sup_{A\in\mathcal{R}; u,v\in[0,\varepsilon]^d} \left| \int 1_{A^{u,v}\setminus A}(\sqrt{1-t}z) p_W(z) dz \right|.$$

Further, for each $j = 1, \dots, d$, set

$$\begin{cases} I_j = [a_j + \varepsilon, b_j - \varepsilon],\ u_j = v_j = \varepsilon & \text{if } a_j + \varepsilon < b_j - \varepsilon, \\ I_j = \{(a_j + b_j)/2\},\ u_j = v_j = (b_j - a_j)/2 & \text{if } b_j \leq a_j + 2\varepsilon. \end{cases}$$

Then we have $\tilde{A} := \prod_{j=1}^d I_j \in \mathcal{R}$, $u := (u_1, \dots, u_d)^\top \in [0,\varepsilon]^d$, $v := (v_1, \dots, v_d)^\top \in [0,\varepsilon]^d$ and

$$\int [h(x) - m_h(x;\epsilon)]\nu(dx) = \int 1_{\tilde{A}^{u,v}\setminus \tilde{A}}(\sqrt{1-t}z) p_W(z) dz.$$

Consequently,

$$\tau(h;\varepsilon) \leq \sup_{A\in\mathcal{R}; u,v\in[0,\varepsilon]^d} \left| \int 1_{A^{u,v}\setminus A}(\sqrt{1-t}z) p_W(z) dz \right|. \tag{4.10}$$

Besides, for any $x \in \mathbb{R}^d$ and $y \in [-\varepsilon,\varepsilon]^d$, we have $h(x+y) - h(x) = 1_{(A+y)\setminus A}(x) - 1_{A\setminus(A+y)}(x)$. For each $j = 1, \dots, d$, set

$$\begin{cases} I_j = [a_j + y_j, b_j], u_j = 0, v_j = y_j & \text{if } y_j \geq 0, a_j + y_j < b_j, \\ I_j = [a_j, b_j + y_j], u_j = -y_j, v_j = 0 & \text{if } y_j < 0, a_j < b_j + y_j, \\ I_j = \{a_j + y_j\}, u_j = 0, v_j = b_j - a_j & \text{otherwise.} \end{cases}$$

Then we have $\tilde{A} := \prod_{j=1}^d I_j \in \mathcal{R}$, $u := (u_1, \dots, u_d)^\top \in [0,\varepsilon]^d$, $v := (v_1, \dots, v_d)^\top \in [0,\varepsilon]^d$ and

$$\int 1_{(A+y)\setminus A}(x)\nu(dx) = \int 1_{\tilde{A}^{u,v}\setminus \tilde{A}}(\sqrt{1-t}z) p_W(z) dz.$$

Also, observe that $A \setminus (A+y) = [(A+y) - y] \setminus (A+y)$ and $A + y \in \mathcal{R}$. Hence we conclude

$$\tilde{\tau}(h;\varepsilon) \leq 2 \sup_{A\in\mathcal{R}; u,v\in[0,\varepsilon]^d} \left| \int 1_{A^{u,v}\setminus A}(\sqrt{1-t}z) p_W(z) dz \right|. \tag{4.11}$$

In addition, observe that $h_y = 1_{A-y}$ for any $y \in \mathbb{R}^d$. Hence, Eqs.(4.9)–(4.11) still hold if the left hand side is replaced by the one marked with an asterisk. As a result, Lemma 4.1 gives

$$\sup_{A\in\mathcal{R}} \left| P(\sqrt{1-t}W \in A) - \int_{\mathbb{R}^d} 1_A(\sqrt{1-t}z) p_W(z) dz \right|$$

$$\leq 2 \sup_{h=1_A, A\in\mathcal{R}} \left| \mathrm{E}[h_t(W)] - \int h_t(z) p_W(z) dz \right| + 6 \sup_{A\in\mathcal{R}; u,v\in[0,\varepsilon]^d} \left| \int 1_{A^{u,v}\setminus A}(\sqrt{1-t}z) p_W(z) dz \right|. \tag{4.12}$$

Note that $A/\sqrt{1-t} \in \mathcal{R}$ for any $A \in \mathcal{R}$. Thus, the left hand side of (4.12) is equal to

$$\sup_{A\in\mathcal{R}} \left| P(W \in A) - \int_A p_W(z) dz \right|.$$

Further, by Lemmas D.2 and 4.2

$$\begin{aligned}
&\sup_{A\in\mathcal{R}; u,v\in[0,\varepsilon]^d} \left| \int 1_{A^{u,v}\setminus A}(\sqrt{1-t}z) p_W(z) dz \right| \\
&\lesssim \frac{\varepsilon}{\sqrt{1-t}} \left(\frac{\sqrt{\log d}}{\underline{\sigma}} + \frac{\log^{3/2} d}{\sigma_*^3} \|\Sigma_W - \Sigma\|_\infty + \frac{\log^2 d}{\sigma_*^4} \left\| \sum_{i=1}^n \mathrm{E}[\xi_i^{\otimes 3}] \right\|_\infty \right) \\
&\lesssim \overline{\sigma}\sqrt{t} \left(\frac{\log d}{\underline{\sigma}} + \frac{\log^2 d}{\sigma_*^3} \|\Sigma_W - \Sigma\|_\infty + \frac{\log^{5/2} d}{\sigma_*^4} \left\| \sum_{i=1}^n \mathrm{E}[\xi_i^{\otimes 3}] \right\|_\infty \right).
\end{aligned} \tag{4.13}$$

Meanwhile, set $\kappa := (\log d)/\sigma_*^2$. Then, by Lemma 4.4, we have for any $u, v \in (0,1)$

$$|\mathrm{E}[\langle \bar{T}^{\otimes 2}, \nabla^4 h_u(W)\rangle]| \lesssim \mathrm{E}\, \|\bar{T}\|_\infty^2 \left(\frac{1-u}{u}\kappa \right)^2,$$

$$\left| \sum_{i=1}^n \mathrm{E}[\langle \tau_i(\xi_i)^{\otimes 2}, \nabla^4 h_u(W)\rangle] \right| \lesssim \mathrm{E} \left\| \sum_{i=1}^n \tau_i(\xi_i)^{\otimes 2} \right\|_\infty \left(\frac{1-u}{u}\kappa \right)^2,$$

$$\left| \sum_{i=1}^n \mathrm{E}[\langle \tau_i(\xi_i) \otimes \xi_i^{\otimes 2}, \nabla^4 h_u(W)\rangle] \right| \lesssim \mathrm{E} \left\| \sum_{i=1}^n \tau_i(\xi_i) \otimes \xi_i^{\otimes 2} \right\|_\infty \left(\frac{1-u}{u}\kappa \right)^2,$$

$$\left| \sum_{i=1}^n \mathrm{E}[\langle \xi_i^{\otimes 4}, \nabla^4 h_v(W)\rangle] \right| \lesssim \mathrm{E} \left\| \sum_{i=1}^n \xi_i^{\otimes 4} \right\|_\infty \left(\frac{1-v}{v}\kappa \right)^2,$$

$$\left| \sum_{i=1}^n \mathrm{E}[\langle \xi_i^{\otimes 3} \otimes B, \nabla^4 h_v(W)\rangle] \right| \lesssim \mathrm{E} \left\| B \otimes \sum_{i=1}^n \xi_i^{\otimes 3} \right\|_\infty \left(\frac{1-v}{v}\kappa \right)^2,$$

$$\left| \sum_{i=1}^n \mathrm{E}[\langle \xi_i^{\otimes 3} \otimes \beta_i(\xi_i), \nabla^4 h_v(W)\rangle] \right| \lesssim \mathrm{E} \left\| \sum_{i=1}^n \xi_i^{\otimes 3} \otimes \beta_i(\xi_i) \right\|_\infty \left(\frac{1-v}{v}\kappa \right)^2,$$

$$\left| \sum_{i=1}^n \mathrm{E}[\langle \xi_i^{\otimes 3} \otimes \bar{T}, \nabla^5 h_v(W)\rangle] \right| \lesssim \mathrm{E} \left\| \bar{T} \otimes \sum_{i=1}^n \xi_i^{\otimes 3} \right\|_\infty \left(\frac{1-v}{v}\kappa \right)^{5/2},$$

$$\left| \sum_{i=1}^n \mathrm{E}[\langle \xi_i^{\otimes 3} \otimes \tau_i(\xi_i), \nabla^5 h_v(W)\rangle] \right| \lesssim \mathrm{E} \left\| \sum_{i=1}^n \xi_i^{\otimes 3} \otimes \tau_i(\xi_i) \right\|_\infty \left(\frac{1-v}{v}\kappa \right)^{5/2},$$

$$|\mathrm{E}[\langle B^{\otimes 2}, \nabla^2 h_u(W)\rangle]| \lesssim \mathrm{E}\, \|B\|_\infty^2 \left(\frac{1-u}{u}\kappa \right),$$

$$\left| \sum_{i=1}^n \mathrm{E}[\langle \beta_i(\xi_i)^{\otimes 2}, \nabla^2 h_u(W)\rangle] \right| \lesssim \mathrm{E} \left\| \sum_{i=1}^n \beta_i(\xi_i)^{\otimes 2} \right\|_\infty \left(\frac{1-u}{u}\kappa \right),$$

$$\left|\sum_{i=1}^n \mathrm{E}[\langle \beta_i(\xi_i)\otimes \bar{T}, \nabla^3 h_u(W)\rangle]\right| \lesssim \mathrm{E}\left\|\bar{T}\otimes B\right\|_\infty \left(\frac{1-u}{u}\kappa\right)^{3/2},$$

$$\left|\sum_{i=1}^n \mathrm{E}[\langle \beta_i(\xi_i)\otimes \tau_i(\xi_i), \nabla^3 h_u(W)\rangle]\right| \lesssim \mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)\otimes \tau_i(\xi_i)\right\|_\infty \left(\frac{1-u}{u}\kappa\right)^{3/2},$$

$$\left|\sum_{i=1}^n \mathrm{E}[\langle \beta_i(\xi_i)\otimes \xi_i^{\otimes 2}, \nabla^3 h_u(W)\rangle]\right| \lesssim \mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)\otimes \xi_i^{\otimes 2}\right\|_\infty \left(\frac{1-u}{u}\kappa\right)^{3/2},$$

$$\left|\sum_{i=1}^n \mathrm{E}[\langle \beta_i(\xi_i)\otimes \xi_i\otimes \bar{T}, \nabla^4 h_v(W)\rangle]\right| \lesssim \mathrm{E}\left\|\bar{T}\otimes\sum_{i=1}^n \beta_i(\xi_i)\otimes \xi_i\right\|_\infty \left(\frac{1-u}{u}\kappa\right)^{2},$$

$$\left|\sum_{i=1}^n \mathrm{E}[\langle \beta_i(\xi_i)\otimes \xi_i\otimes \tau_i(\xi_i), \nabla^4 h_v(W)\rangle]\right| \lesssim \mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)\otimes \xi_i\otimes \tau_i(\xi_i)\right\|_\infty \left(\frac{1-v}{v}\kappa\right)^{2},$$

$$\left|\sum_{i=1}^n \mathrm{E}[\langle \beta_i(\xi_i)\otimes \xi_i\otimes B, \nabla^3 h_v(W)\rangle]\right| \lesssim \mathrm{E}\left\|B\otimes\sum_{i=1}^n \beta_i(\xi_i)\otimes \xi_i\right\|_\infty \left(\frac{1-v}{v}\kappa\right)^{3/2},$$

$$\left|\sum_{i=1}^n \mathrm{E}[\langle \beta_i(\xi_i)\otimes \xi_i\otimes \beta_i(\xi_i), \nabla^3 h_v(W)\rangle]\right| \lesssim \mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)^{\otimes 2}\otimes \xi_i\right\|_\infty \left(\frac{1-v}{v}\kappa\right)^{3/2}.$$

Also, by the AM–GM inequality,

$$2\,\mathrm{E}\left\|\sum_{i=1}^n \tau_i(\xi_i)\otimes\xi_i^{\otimes 2}\right\|_\infty \kappa^2 \le \mathrm{E}\left\|\sum_{i=1}^n \tau_i(\xi_i)^{\otimes 2}\right\|_\infty \kappa^2 + \mathrm{E}\left\|\sum_{i=1}^n \xi_i^{\otimes 4}\right\|_\infty \kappa^2,$$

$$2\,\mathrm{E}\left\|\bar{T}\otimes B\right\|_\infty \kappa^{3/2} \le \mathrm{E}\left\|\bar{T}\right\|_\infty^2 \kappa^2 + \mathrm{E}\left\|B\right\|_\infty^2 \kappa,$$

$$2\,\mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)\otimes\tau_i(\xi_i)\right\|_\infty \kappa^{3/2} \le \mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)^{\otimes 2}\right\|_\infty \kappa + \mathrm{E}\left\|\sum_{i=1}^n \tau_i(\xi_i)^{\otimes 2}\right\|_\infty \kappa^2,$$

$$2\,\mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)\otimes\xi_i^{\otimes 2}\right\|_\infty \kappa^{3/2} \le \mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)^{\otimes 2}\right\|_\infty \kappa + \mathrm{E}\left\|\sum_{i=1}^n \xi_i^{\otimes 4}\right\|_\infty \kappa^2.$$

Further, observe that

$$\frac{1}{w^a\sqrt{1-w}} \le \frac{\sqrt{2}}{w^a} + \frac{2^a}{\sqrt{1-w}} \quad \text{for any } w\in(0,1), a>0$$

and

$$\int_t^1 \frac{1}{s}ds = |\log t|, \quad \int_t^1 \frac{1}{\sqrt{s}}ds \le 2, \quad \int_t^1 \frac{1}{\sqrt{1-s}}ds \le 2, \quad \int_t^1 \frac{-\log s}{\sqrt{1-s}}ds \lesssim 1.$$

Combining these estimates with (4.12)–(4.13) and (4.6) gives the desired result. □

Proof of Theorem 4.2. The claim follows by replacing (4.6) with (4.7) in the proof of Theorem 4.1. □

4.2 Proof of Theorem 2.1

Below we will frequently use the following identity without reference: For any $x_1,\dots,x_n\in\mathbb{R}^d$ and $r\in\mathbb{N}$,

$$\left\|\sum_{i=1}^n x_i^{\otimes(2r)}\right\|_\infty = \max_{1\le j\le d}\sum_{i=1}^n x_{ij}^{2r}.$$

This follows from the AM-GM inequality.

Proof of Theorem 2.1. Observe that for every $i = 1, \dots, n$, $(\tau_i^X + (\tau_i^X)^\top)/2$ is a Stein kernel for X_i and $\max_{j,k} \|(\tau_{i,jk}^X(X_i) + \tau_{i,kj}^X(X_i))/2\|_{\psi_{1/2}} \lesssim b^2$. Therefore, we may assume $\tau_{i,jk}^X = \tau_{i,kj}^X$ for all $j, k \in \{1, \dots, d\}$ without loss of generality. This particularly implies $\mathrm{E}[\tau_i^X(X_i)] = \mathrm{E}[X_i^{\otimes 2}]$.

We apply Theorem 4.1 to $\xi_i = X_i/\sqrt{n}$. Observe that $\Sigma_W = \Sigma$ and that ξ_i has a Stein kernel τ_i satisfying $\tau_i(\xi_i) = \tau_i^X(X_i)/n$.

Let us bound the quantities appearing on the right hand side of (4.2). First, noting that $\mathrm{E}[\tau_i(\xi_i)] = \mathrm{E}[\xi_i^{\otimes 2}]$, we have

$$\bar{T} = \sum_{i=1}^n (\tau_i(\xi_i) - \mathrm{E}[\tau_i(\xi_i)]).$$

Thus, by Lemma D.10 with $K = b^2/n$, $\alpha = 1/2$ and $r = 2$, we obtain

$$\mathrm{E}[\|\bar{T}\|_\infty^2] \lesssim \frac{b^4}{n^2} \left(\sqrt{n \log d} + (\log d)^2 \right)^2 \lesssim \frac{b^4 \log d}{n}, \tag{4.14}$$

where the second inequality follows from $\log^3 d \le n$. Next, by Lemma D.8 and Lemma D.10 with $K = b^2/n$, $\alpha = 1/4$ and $r = 1$,

$$\mathrm{E}\left\| \sum_{i=1}^n \{\tau_i(\xi_i)^{\otimes 2} - \mathrm{E}[\tau_i(\xi_i)^{\otimes 2}]\} \right\|_\infty \lesssim \frac{b^4}{n^2} \left(\sqrt{n \log d} + (\log d)^4 \right).$$

Therefore,

$$\begin{aligned} \mathrm{E}\left\| \sum_{i=1}^n \tau_i(\xi_i)^{\otimes 2} \right\|_\infty &\le \left\| \sum_{i=1}^n \mathrm{E}[\tau_i(\xi_i)^{\otimes 2}] \right\|_\infty + \mathrm{E}\left\| \sum_{i=1}^n \{\tau_i(\xi_i)^{\otimes 2} - \mathrm{E}[\tau_i(\xi_i)^{\otimes 2}]\} \right\|_\infty \\ &\lesssim \frac{b^4}{n} + \frac{b^4 \sqrt{\log d}}{n^{3/2}} + \frac{b^4 (\log d)^4}{n^2} \lesssim \frac{b^4}{n} + \frac{b^4 (\log d)^4}{n^2}, \end{aligned} \tag{4.15}$$

where the last inequality follows from $\sqrt{\log d} \le n^{1/6} \le \sqrt{n}$. Similarly, we can show that

$$\mathrm{E}\left\| \sum_{i=1}^n \xi_i^{\otimes 4} \right\|_\infty \lesssim \frac{b^4}{n} + \frac{b^4 (\log d)^4}{n^2}, \tag{4.16}$$

$$\mathrm{E}\left\| \sum_{i=1}^n \xi_i^{\otimes 3} \otimes \tau_i(\xi_i) \right\|_\infty \lesssim \frac{b^5}{n^{3/2}} + \frac{b^5 (\log d)^5}{n^{5/2}}. \tag{4.17}$$

In addition, by the Schwarz inequality,

$$\mathrm{E}\left\| \bar{T} \otimes \sum_{i=1}^n \xi_i^{\otimes 3} \right\|_\infty \le \sqrt{\mathrm{E}\left\| \bar{T} \right\|_\infty^2} \sqrt{\mathrm{E}\left\| \sum_{i=1}^n \xi_i^{\otimes 3} \right\|_\infty^2}.$$

Similarly to the proof of (4.15), we can prove

$$\mathrm{E}\left\| \sum_{i=1}^n \xi_i^{\otimes 3} \right\|_\infty^2 \lesssim \left(\frac{b^3}{\sqrt{n}} + \frac{b^3 (\log d)^3}{n^{3/2}} \right)^2 \lesssim \frac{b^6}{n}, \tag{4.18}$$

where the second inequality follows by the assumption $\log^3 d \le n$. Combining this with (4.14) gives

$$\mathrm{E}\left\|\bar{T}\otimes\sum_{i=1}^n \xi_i^{\otimes 3}\right\|_\infty \lesssim \frac{b^5\sqrt{\log d}}{n}. \tag{4.19}$$

Now, by (4.14)–(4.16),

$$\begin{aligned}
&\frac{\log^2 d}{\sigma_*^4}\left(\mathrm{E}\,\|\bar{T}\|_\infty^2 + \mathrm{E}\left\|\sum_{i=1}^n \tau_i(\xi_i)^{\otimes 2}\right\|_\infty + \mathrm{E}\left\|\sum_{i=1}^n \xi_i^{\otimes 4}\right\|_\infty\right)\\
&\lesssim \frac{\log^2 d}{\sigma_*^4}\left(\frac{b^4\log d}{n} + \frac{b^4(\log d)^4}{n^2}\right) \lesssim \frac{b^4\log^3 d}{\sigma_*^4 n},
\end{aligned}$$

where the second inequality follows by the assumption $\log^3 d \le n$. Also, by (4.17) and (4.19),

$$\begin{aligned}
&\frac{\log^{5/2} d}{\sigma_*^5}\left(\mathrm{E}\left\|\bar{T}\otimes\sum_{i=1}^n \xi_i^{\otimes 3}\right\|_\infty + \mathrm{E}\left\|\sum_{i=1}^n \xi_i^{\otimes 3}\otimes\tau_i(\xi_i)\right\|_\infty\right)\\
&\lesssim \frac{\log^{5/2} d}{\sigma_*^5}\left(\frac{b^5\sqrt{\log d}}{n} + \frac{b^5(\log d)^5}{n^{5/2}}\right) \lesssim \frac{b^5\log^3 d}{\sigma_*^5 n},
\end{aligned}$$

where we used the assumption $\log^3 d \le n$ in the second inequality. All together, we obtain by Theorem 4.1

$$\begin{aligned}
&\sup_{A\in\mathcal{R}}\left|P(S_n\in A) - \int_A p_n(z)dz\right|\\
&\lesssim |\log t|\frac{b^4\log^3 d}{\sigma_*^4 n} + \frac{b^5\log^3 d}{\sigma_*^5 n} + \overline{\sigma}\sqrt{t}\left(\frac{\log d}{\underline{\sigma}} + \frac{\log^{5/2} d}{\sigma_*^4}\left\|\sum_{i=1}^n \mathrm{E}[\xi_i^{\otimes 3}]\right\|_\infty\right)
\end{aligned}$$

for any $t\in(0,1/2]$. With $t = 1/n^2$, we obtain

$$\overline{\sigma}\sqrt{t}\left(\frac{\log d}{\underline{\sigma}} + \frac{\log^{5/2} d}{\sigma_*^4}\left\|\sum_{i=1}^n \mathrm{E}[\xi_i^{\otimes 3}]\right\|_\infty\right) \lesssim \frac{b\log d}{\sigma_* n} + \frac{b^4\log^{5/2} d}{\sigma_*^4 n^{3/2}} \lesssim \frac{b^5\log^3 d}{\sigma_*^5 n}.$$

Consequently, we obtain the desired result. □

4.3 Proof of Theorem 2.2

First consider Case (i). Set $\widetilde{X}_i := X_i - \bar{X}$ for $i = 1,\dots,n$. We apply Theorem 4.1 with $\xi_i = w_i\widetilde{X}_i/\sqrt{n}$ and $t = 1/n^3$ conditional on the data. Note that conditional on the data, ξ_i has a Stein kernel τ_i satisfying $\tau_i(\xi_i) = \mathrm{E}^*[\tau_i^*(w_i)\widetilde{X}_i^{\otimes 2}/n \mid \xi_i]$ by Example 2.3. Therefore, we have

$$\begin{aligned}
&\sup_{A\in\mathcal{R}}\left|P^*(S_n^*\in A) - \int_A \tilde{p}_{n,\gamma}(z)dz\right|\\
&\lesssim (\log n)\frac{\log^2 d}{\sigma_*^4}\left(\mathrm{E}^*\left[\|\bar{T}^*\|_\infty^2\right] + \mathrm{E}^*\left[\left\|\frac{1}{n^2}\sum_{i=1}^n \tau_i^*(w_i)^2\widetilde{X}_i^{\otimes 4}\right\|_\infty\right] + \mathrm{E}^*\left[\left\|\frac{1}{n^2}\sum_{i=1}^n w_i^4\widetilde{X}_i^{\otimes 4}\right\|_\infty\right]\right)\\
&\quad + \frac{\log^{5/2} d}{\sigma_*^5}\left(\mathrm{E}^*\left[\left\|\bar{T}^*\otimes\frac{1}{n^{3/2}}\sum_{i=1}^n w_i^3\widetilde{X}_i^{\otimes 3}\right\|_\infty\right] + \mathrm{E}^*\left[\left\|\frac{1}{n^{5/2}}\sum_{i=1}^n w_i^3\tau_i^*(w_i)\widetilde{X}_i^{\otimes 5}\right\|_\infty\right]\right)
\end{aligned}$$

$$
+ \frac{\overline{\sigma}}{n^{3/2}} \left(\frac{\log d}{\underline{\sigma}} + \frac{\log^2 d}{\sigma_*^3} \|\widehat{\Sigma}_n - \Sigma\|_\infty + \frac{\log^{5/2} d}{\sigma_*^4} \left\| \frac{1}{n^{3/2}} \sum_{i=1}^n \mathrm{E}[w_i^3] \widetilde{X}_i^{\otimes 3} \right\|_\infty \right), \tag{4.20}
$$

where

$$
\tilde{p}_{n,\gamma}(z) = \phi_\Sigma(z) + \frac{1}{2} \langle \widehat{\Sigma}_n - \Sigma, \nabla^2 \phi_\Sigma(z) \rangle - \frac{\gamma}{6n^{3/2}} \sum_{i=1}^n \langle (X_i - \bar{X})^{\otimes 3}, \nabla^3 \phi_\Sigma(z) \rangle
$$

and

$$
\bar{T}^* := \frac{1}{n} \sum_{i=1}^n \tau_i^*(w_i) \widetilde{X}_i^{\otimes 2} - \Sigma.
$$

Next, by Lemmas D.10 and D.12, there exists a universal constant c such that the event

$$
\begin{aligned}
\mathcal{E}_n := \bigcap_{r=1}^2 & \left\{ \left\| \frac{1}{n} \sum_{i=1}^n (X_i^{\otimes r} - \mathrm{E}[X_i^{\otimes r}]) \right\|_\infty \le cb^r \left(\sqrt{\frac{\log(dn)}{n}} + \frac{\log^r(dn)}{n} \right) \right\} \\
& \cap \bigcap_{r=3}^5 \left\{ \max_{1 \le j \le d} \left| \frac{1}{n} \sum_{i=1}^n (|X_{ij}|^r - \mathrm{E}[|X_{ij}|^r]) \right| \le cb^r \left(\sqrt{\frac{\log(dn)}{n}} + \frac{\log^r(dn)}{n} \right) \right\} \\
& \cap \left\{ \max_{1 \le j,k \le d} \frac{1}{n} \sum_{i=1}^n |X_{ij}||X_{ik}| 1_{\{|X_{ij}| \vee |X_{ik}| > 2b \log n\}} \le cb^2 \left(\sqrt{\frac{\log(dn)}{n}} + \frac{\log^2(dn)}{n} \right) \right\} \\
& \cap \left\{ \max_{1 \le j,k \le d} \frac{1}{n} \sum_{i=1}^n X_{ij}^2 X_{ik}^2 1_{\{|X_{ij}| \vee |X_{ik}| \le 2b \log n\}} \le cb^4 \left(1 + \frac{\log(dn) \log^4 n}{n} \right) \right\}
\end{aligned} \tag{4.21}
$$

occurs with probability at least $1 - 1/n$. Recall that $\log^3 d \le n$ by assumption. Hence, on $\mathcal{E}_n$, we have

$$
\|\bar{X}\|_\infty \lesssim b \sqrt{\frac{\log(dn)}{n}} \tag{4.22}
$$

and

$$
\left\| \frac{1}{n} \sum_{i=1}^n X_i^{\otimes 2} - \Sigma \right\|_\infty \lesssim b^2 \sqrt{\frac{\log(dn)}{n}}. \tag{4.23}
$$

Thus, on $\mathcal{E}_n$,

$$
\left\| \widehat{\Sigma}_n - \Sigma \right\|_\infty \le \left\| \frac{1}{n} \sum_{i=1}^n X_i^{\otimes 2} - \Sigma \right\|_\infty + \|\bar{X}\|_\infty^2 \lesssim b^2 \sqrt{\frac{\log(dn)}{n}}. \tag{4.24}
$$

Meanwhile, for every $r \in \{2, 3, 4, 5\}$, we have on $\mathcal{E}_n$

$$
\left\| \frac{1}{n} \sum_{i=1}^n X_i^{\otimes r} \right\| \le \max_{1 \le j \le d} \frac{1}{n} \sum_{i=1}^n |X_{ij}|^r \lesssim b^r \left(1 + \frac{\log^r(dn)}{n} \right). \tag{4.25}
$$

Combining this bound with (4.22), we have on $\mathcal{E}_n$

$$
\max_{1 \le j \le d} \frac{1}{n} \sum_{i=1}^n |\widetilde{X}_{ij}|^r \lesssim b^r \left(1 + \frac{\log^r(dn)}{n} \right). \tag{4.26}
$$

Further, by (4.22), (4.25), the construction of $\mathcal{E}_n$ and $\log^3 d \le n$, we have on $\mathcal{E}_n$

$$
\max_{1 \le j,k \le d} \frac{1}{n} \sum_{i=1}^n |\widetilde{X}_{ij} \widetilde{X}_{ik}| 1_{\{|X_{ij}| \vee |X_{ik}| > 2b \log n\}} \lesssim b^2 \left(\sqrt{\frac{\log d}{n}} + \frac{\log^2(dn)}{n} \right), \tag{4.27}
$$

and

$$\max_{1\le j,k\le d}\frac{1}{n}\sum_{i=1}^{n}\widetilde{X}_{ij}^{2}\widetilde{X}_{ik}^{2}1_{\{|X_{ij}|\vee|X_{ik}|\le 2b\log n\}}\lesssim b^{4}. \tag{4.28}$$

Now we bound the right hand side of (4.20) on the event $\mathcal{E}_n$. First, we have

$$\mathrm{E}^*[\|\bar{T}^*\|_\infty^2]\le 2\,\mathrm{E}^*\left[\max_{1\le j,k\le d}R_{jk}^2\right]+2\|\widehat{\Sigma}_n-\Sigma\|_\infty^2, \tag{4.29}$$

where $R_{jk}:=n^{-1}\sum_{i=1}^{n}\{\tau_i^*(w_i)-1\}\widetilde{X}_{ij}\widetilde{X}_{ik}$. We decompose R_{jk} as

$$\begin{aligned}R_{jk}&=\frac{1}{n}\sum_{i=1}^{n}\{\tau_i^*(w_i)-1\}\widetilde{X}_{ij}\widetilde{X}_{ik}\left(1_{\{|X_{ij}|\vee|X_{ik}|>2b\log n\}}+1_{\{|X_{ij}|\vee|X_{ik}|\le 2b\log n\}}\right)\\&=: R_{1,jk}+R_{2,jk}.\end{aligned}$$

Since

$$\max_{1\le j,k\le d}|R_{1,jk}|\le(b_w^2+1)\max_{1\le j,k\le d}\frac{1}{n}\sum_{i=1}^{n}|\widetilde{X}_{ij}\widetilde{X}_{ik}|1_{\{|X_{ij}|\vee|X_{ik}|>2b\log n\}},$$

we have on $\mathcal{E}_n$

$$\max_{1\le j,k\le d}|R_{1,jk}|\lesssim b_w^2b^2\left(\sqrt{\frac{\log d}{n}}+\frac{\log^2(dn)}{n}\right) \tag{4.30}$$

by (4.27). Meanwhile, by Nemirovski's inequality (e.g. Lemma 14.24 in [13]),

$$\mathrm{E}^*\left[\max_{1\le j,k\le d}R_{2,jk}^2\right]\lesssim\frac{b_w^4\log d}{n}\max_{1\le j,k\le d}\frac{1}{n}\sum_{i=1}^{n}\widetilde{X}_{ij}^{2}\widetilde{X}_{ik}^{2}1_{\{|X_{ij}|\vee|X_{ik}|\le 2b\log n\}}.$$

Hence we have on $\mathcal{E}_n$

$$\mathrm{E}^*\left[\max_{1\le j,k\le d}R_{2,jk}^2\right]\lesssim\frac{b_w^4b^4\log d}{n} \tag{4.31}$$

by (4.28). Combining (4.29)–(4.31) with (4.24), we obtain

$$\mathrm{E}^*\left[\|\bar{T}^*\|_\infty^2\right]\lesssim b_w^4b^4\left(\frac{\log(dn)}{n}+\frac{\log^4(dn)}{n^2}\right). \tag{4.32}$$

Consequently, we have

$$\frac{\log^2 d}{\sigma_*^4}\mathrm{E}^*\left[\|\bar{T}^*\|_\infty^2\right]\lesssim\frac{b_w^4b^4}{\sigma_*^4}\left(\frac{\log^3(dn)}{n}+\frac{\log^6(dn)}{n^2}\right)\lesssim\frac{b_w^4b^4}{\sigma_*^4}\frac{\log^3(dn)}{n}, \tag{4.33}$$

where we used the assumption $\log^3 d\le n$ for the last inequality. Next, we have

$$\left\|\frac{1}{n^2}\sum_{i=1}^{n}\tau_i^*(w_i)^2\widetilde{X}_i^{\otimes 4}\right\|_\infty\le b_w^4\max_{j,k,l,m}\frac{1}{n^2}\sum_{i=1}^{n}|\widetilde{X}_{ij}\widetilde{X}_{ik}\widetilde{X}_{il}\widetilde{X}_{im}|\lesssim\frac{b_w^4b^4}{n}\left(1+\frac{\log^4(dn)}{n}\right), \tag{4.34}$$

where the second inequality follows by the AM-GM inequality and (4.26). Similarly, we can prove

$$\left\|\frac{1}{n^2}\sum_{i=1}^{n}w_i^4\widetilde{X}_i^{\otimes 4}\right\|_\infty\lesssim\frac{b_w^4b^4}{n}\left(1+\frac{\log^4(dn)}{n}\right), \tag{4.35}$$

$$\left\| \frac{1}{n^{3/2}} \sum_{i=1}^n w_i^3 \widetilde{X}_i^{\otimes 3} \right\|_\infty \lesssim \frac{b_w^3 b^3}{\sqrt{n}} \left(1 + \frac{\log^3(dn)}{n}\right) \lesssim \frac{b_w^3 b^3}{\sqrt{n}}, \tag{4.36}$$

$$\left\| \frac{1}{n^{5/2}} \sum_{i=1}^n w_i^3 \tau_i^*(w_i) \widetilde{X}_i^{\otimes 5} \right\|_\infty \lesssim \frac{b_w^5 b^5}{n^{3/2}} \left(1 + \frac{\log^5(dn)}{n}\right). \tag{4.37}$$

By (4.34)–(4.35),

$$\begin{aligned} &\frac{\log^2 d}{\sigma_*^4} \left(\mathrm{E}^* \left[\left\| \frac{1}{n^2} \sum_{i=1}^n \tau_i^*(w_i)^2 \widetilde{X}_i^{\otimes 4} \right\|_\infty \right] + \mathrm{E}^* \left[\left\| \frac{1}{n^2} \sum_{i=1}^n w_i^4 \widetilde{X}_i^{\otimes 4} \right\|_\infty \right] \right) \\ &\lesssim \frac{b_w^4 b^4}{\sigma_*^4} \left(\frac{\log^2 d}{n} + \frac{\log^6(dn)}{n^2} \right) \lesssim \frac{b_w^4 b^4}{\sigma_*^4} \frac{\log^3(dn)}{n}. \end{aligned} \tag{4.38}$$

Also, by (4.37),

$$\begin{aligned} \frac{\log^{5/2} d}{\sigma_*^5} \mathrm{E}^* \left[\left\| \frac{1}{n^{5/2}} \sum_{i=1}^n w_i^3 \tau_i^*(w_i) \widetilde{X}_i^{\otimes 5} \right\|_\infty \right] &\lesssim \frac{b_w^5 b^5}{\sigma_*^5} \left(\frac{\log^{5/2} d}{n^{3/2}} + \frac{\log^{15/2}(dn)}{n^{5/2}} \right) \\ &\lesssim \frac{b_w^5 b^5}{\sigma_*^5} \frac{\log^3(dn)}{n}. \end{aligned} \tag{4.39}$$

Further, by the Schwarz inequality, (4.32) and (4.36),

$$\begin{aligned} \frac{\log^{5/2} d}{\sigma_*^5} \mathrm{E}^* \left[\left\| \bar{T}^* \otimes \frac{1}{n^{3/2}} \sum_{i=1}^n w_i^3 \widetilde{X}_i^{\otimes 3} \right\|_\infty \right] &\lesssim \frac{\log^{5/2} d}{\sigma_*^5} \sqrt{b_w^4 b^4 \left(\frac{\log(dn)}{n} + \frac{\log^4(dn)}{n^2} \right)} \cdot \frac{b_w^3 b^3}{\sqrt{n}} \\ &\le \frac{b_w^5 b^5}{\sigma_*^5} \left(\frac{\log^3(dn)}{n} + \frac{\log^{9/2}(dn)}{n^{3/2}} \right) \lesssim \frac{b_w^5 b^5}{\sigma_*^5} \frac{\log^3(dn)}{n}. \end{aligned} \tag{4.40}$$

Combining (4.20), (4.24), (4.33), (4.36), (4.38)–(4.40) and $b_w \ge 1, b/\sigma_* \ge 1$, we have, on $\mathcal{E}_n$,

$$\begin{aligned} &\sup_{A \in \mathcal{R}} \left| P^*(S_n^* \in A) - \int_A \tilde{p}_{n,\gamma}(z) dz \right| \\ &\lesssim \frac{b_w^5 b^5}{\sigma_*^5} \frac{\log^3(dn)}{n} \log n + \frac{\overline{\sigma}}{n^{3/2}} \left(\frac{\log d}{\underline{\sigma}} + b^2 \frac{\log^{5/2} d}{\sigma_*^3 \sqrt{n}} + b^3 b_w^3 \frac{\log^{5/2} d}{\sigma_*^4 \sqrt{n}} \right) \lesssim \frac{b_w^5 b^5}{\sigma_*^5} \frac{\log^3(dn)}{n} \log n. \end{aligned}$$

It remains to prove

$$\sup_{A \in \mathcal{R}} \left| \int_A \tilde{p}_{n,\gamma}(z) dz - \int_A \hat{p}_{n,\gamma}(z) dz \right| \lesssim \frac{b_w^5 b^5}{\sigma_*^5} \frac{\log^3(dn)}{n} \log n \quad \text{on } \mathcal{E}_n. \tag{4.41}$$

Observe that

$$\sup_{A \in \mathcal{R}} |\langle \widehat{\Sigma}_n - \overline{X^2}, \int_A \nabla^2 \phi_\Sigma(z) dz \rangle| \lesssim \|\bar{X}\|_\infty^2 \frac{\log d}{\sigma_*^2} \tag{4.42}$$

and

$$\begin{aligned} &\sup_{A \in \mathcal{R}} \left| \frac{1}{n^{3/2}} \sum_{i=1}^n \langle (X_i - \bar{X})^{\otimes 3} - X_i^{\otimes 3}, \int_A \nabla^3 \phi_\Sigma(z) dz \rangle \right| \\ &\lesssim \frac{1}{\sqrt{n}} \left(\left\| \frac{1}{n} \sum_{i=1}^n X_i^{\otimes 2} \right\|_\infty \|\bar{X}\|_\infty + \|\bar{X}\|_\infty^3 \right) \frac{\log^{3/2} d}{\sigma_*^3}. \end{aligned}$$

Also, note that $|\gamma| \leq b_w \operatorname{E}[w_1^2] = b_w$. Hence (4.41) follows from (4.22) and (4.25).

Next consider Case (ii). In this case, $S_n^* \sim N(0, \widehat{\Sigma}_n)$ conditional on the data. Hence, applying Theorem 4.2 and using the bounds (4.22), (4.24) and (4.42), we obtain the desired bound with a simplified argument of the proof for Case (i). $\square$

4.4 Proof of Proposition 2.1

The proof uses the Malliavin–Stein method. We refer to [55, Chapter 2] for undefined notation and concepts used below.

Denote by r_j the j-th row vector of $R^{1/2}$. Then Z has the same law as $(r_1 \cdot G, \dots, r_d \cdot G)^\top$ with $G \sim N(0, I_d)$. Hence we may assume X is of the form $X = \psi(G)$ with $\psi : \mathbb{R}^d \to \mathbb{R}^d$ defined as $\psi_j(x) = F_j^{-1}(\Phi(r_j \cdot x))$ for $j = 1, \dots, d$ and $x \in \mathbb{R}^d$. Note that F_j^{-1} is absolutely continuous and satisfies $|(F_j^{-1})'(t)| \leq (\kappa \min\{t, 1-t\})^{-1}$ a.e. by (2.5); this follows from arguments in Section 5.3 of [11] (see also Propositions A.17 and A.19 in [11]). In particular, F_j^{-1} is locally Lipschitz, and hence ψ_j is locally Lipschitz and its gradient is given by $\nabla \psi_j(x) = \phi(r_j \cdot x)(F_j^{-1})'(\Phi(r_j \cdot x)) r_j$ a.e. Since $|r_j|^2 = R_{jj} = 1$, we obtain

$$|\nabla \psi_j(G)| = \phi(Z_j)(F_j^{-1})'(\Phi(Z_j)) \leq \frac{\phi(Z_j)}{\kappa \min\{\Phi(Z_j), 1 - \Phi(Z_j)\}} \leq \frac{1 + |Z_j|}{\kappa}, \tag{4.43}$$

where the last inequality follows by Birnbaum's inequality. Now, let $H = \mathbb{R}^d$ be the Hilbert space equipped with the canonical inner product. Consider an isonormal Gaussian process over H given by $W(h) = h \cdot G$, $h \in H$. We consider Malliavin calculus with respect to W. The above argument implies that $X_j = \psi_j(G) \in \mathbb{D}^{1,2}$ and $DX_j = \nabla \psi_j(G)$ for every $j = 1, \dots, d$. Therefore, by Proposition 3.7 in [56], the map $\tau : \mathbb{R}^d \to (\mathbb{R}^d)^{\otimes 2}$ defined by $\tau_{jk}(x) = \operatorname{E}[-DL^{-1} X_j \cdot DX_k \mid X = x]$ for $x \in \mathbb{R}^d$ and $j, k = 1, \dots, d$ gives a Stein kernel for X. For any $p \geq 1$ and $j, k = 1, \dots, d$, we have

$$\operatorname{E}[|\tau_{jk}(X)|^p] \leq \sqrt{\operatorname{E}[|DL^{-1}X_j|^{2p}] \operatorname{E}[|DX_k|^{2p}]} \leq \sqrt{\operatorname{E}[|DX_j|^{2p}] \operatorname{E}[|DX_k|^{2p}]},$$

where the first inequality is by the Jensen and Schwarz inequalities and the second by Lemma 5.3.7 in [55]. If p is an even integer, we also have $\operatorname{E}[X_j^p] \leq (p-1)^{p/2} \operatorname{E}[|DX_j|^p]$ by Lemma 5.3.7 in [55]. Since $\|\nabla \psi_j(G)\|_p \lesssim \sqrt{p}\kappa^{-1}$ by (4.43), we obtain the desired result. $\square$

4.5 Cheeger constant of the gamma distribution

Proposition 4.1. *Any gamma distribution has a positive Cheeger constant.*

Proof. Let μ be the gamma distribution with shape ν and rate α. If $\nu \geq 1$, then μ is log-concave, so the claim follows by Proposition 4.1 in [9]. When $\nu < 1$, the density f of μ satisfies $\inf_{0<t<M} f(t) > 0$ for every $M > 0$. Hence, in view of Theorem 1.3 in [10], it suffices to prove $\liminf_{p\uparrow 1} f(F^{-1}(p))/(1-p) > 0$, where F is the distribution function of μ. Since

$$\frac{d}{dp} f(F^{-1}(p)) = \frac{\nu - 1}{F^{-1}(p)} - \alpha \to -\alpha \quad (p \uparrow 1),$$

we have by L'Hôpital's rule $\liminf_{p\uparrow 1} f(F^{-1}(p))/(1-p) = \alpha > 0$. This completes the proof. $\square$

4.6 *Proof of Corollary 2.1*

Replacing X_i by X_i/σ, we may assume $\sigma = 1$ without loss of generality.

Observe that $P(T_n \geq c^G_{1-\alpha}) - \alpha = P(Z < c^G_{1-\alpha}) - P(T_n < c^G_{1-\alpha})$. Thus, in view of (2.4), it suffices to prove

$$\frac{1}{(\log d)^{3/2}} \int_{(-\infty, c^G_{1-\alpha}]^d} \langle \mathrm{E}[\overline{X^3}], \nabla^3 \phi_\Sigma(z) \rangle dz \to -2\sqrt{2}\gamma_X \log(1-\alpha). \tag{4.44}$$

We decompose the left hand side as

$$\begin{aligned}
&\int_{(-\infty, c^G_{1-\alpha}]^d} \langle \mathrm{E}[\overline{X^3}], \nabla^3 \phi_\Sigma(z) \rangle dz \\
&= \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^d \mathrm{E}[X_{ij}^3]\phi''(c^G_{1-\alpha})\Phi(c^G_{1-\alpha})^{d-1} + \frac{3}{n}\sum_{i=1}^n \sum_{j,k:j\neq k} \mathrm{E}[X_{ij}^2 X_{ik}]\phi'(c^G_{1-\alpha})\phi(c^G_{1-\alpha})\Phi(c^G_{1-\alpha})^{d-2} \\
&\quad + \frac{1}{n}\sum_{i=1}^n \sum_{\substack{j,k,l=1 \\ j,k,l \text{ are distinct}}}^d \mathrm{E}[X_{ij}X_{ik}X_{il}]\phi(c^G_{1-\alpha})^3\Phi(c^G_{1-\alpha})^{d-3} \\
&=: \mathbb{I} + \mathbb{II} + \mathbb{III}.
\end{aligned}$$

Since $\Sigma = I_d$ by assumption, we have $c^G_{1-\alpha} = \Phi^{-1}((1-\alpha)^{1/d})$. Hence we have $c^G_{1-\alpha} = \sqrt{2\log d} + o(1)$ (cf. the proof of [39, Proposition 2.1]). We also have $d\phi(c^G_{1-\alpha})/\sqrt{2\log d} \to -\log(1-\alpha)$ by [12, Lemma 10.3]. Consequently,

$$\frac{\mathbb{I}}{(2\log d)^{3/2}} \to -\gamma_X \log(1-\alpha).$$

Meanwhile, noting $n^{-1}\sum_{i=1}^n \mathrm{E}[X_i^{\otimes 2}] = \Sigma = I_d$, we obtain

$$\left| \frac{1}{n}\sum_{i=1}^n \sum_{j,k:j\neq k} \mathrm{E}[X_{ij}^2 X_{ik}] \right| \leq \sum_{j=1}^d \sqrt{\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_{ij}^4]} \sqrt{\frac{1}{n}\sum_{i=1}^n \mathrm{E}\left[\left(\sum_{k:k\neq j} X_{ik}\right)^2\right]} \lesssim b^2 d^{3/2}$$

and

$$\begin{aligned}
&\left| \frac{1}{n}\sum_{i=1}^n \sum_{\substack{j,k,l=1 \\ j,k,l \text{ are distinct}}}^d \mathrm{E}[X_{ij}X_{ik}X_{il}] \right| \\
&\leq \sum_{j,k:j\neq k} \sqrt{\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_{ij}^2 X_{ik}^2]} \sqrt{\frac{1}{n}\sum_{i=1}^n \mathrm{E}\left[\left(\sum_{l:l\neq j,k} X_{il}\right)^2\right]} \lesssim b^2 d^{5/2}.
\end{aligned}$$

Consequently, $\mathbb{II}/(\log d)^{3/2} \to 0$ and $\mathbb{III}/(\log d)^{3/2} \to 0$. Hence we obtain the desired result. □

5 Proofs for Sections 2.2 and 2.3

Given a random vector W, we denote by F_W the distribution function of $W^\vee$.

5.1 Cornish–Fisher expansion

Asymptotic expansion of coverage probability is conventionally derived with the help of Cornish–Fisher expansion (cf. Section 3.5.2 in [38]), so we first develop such expansions for T_n and T_n^* in our setting.

Theorem 5.1 (Cornish–Fisher expansion for T_n)**.** *Under the assumptions of Theorem 2.1, let $\lambda > 0$ be a constant such that $b/\sigma_* \leq \lambda$. Then, for any $\varepsilon \in (0,1/2)$, there exist positive constants c and C depending only on λ and ε such that if*

$$\frac{\varsigma_d^3}{\sigma_*^3}\frac{\log^3 d}{n}\log n \leq c, \tag{5.1}$$

then

$$\sup_{\varepsilon<p<1-\varepsilon}\left|c_p - \left(c_p^G - \frac{Q_n(c_p^G)}{f_\Sigma(c_p^G)}\right)\right| \leq \frac{C}{\sqrt{\log d}}\frac{\varsigma_d^3}{\sigma_*^2}\frac{\log^3 d}{n}\log n, \tag{5.2}$$

where c_p is the p-quantile of T_n.

Theorem 5.2 (Cornish–Fisher expansion for T_n^*)**.** *Under the assumptions of Theorem 2.3, for any $\varepsilon \in (0,1/2)$, there exist positive constants c and C depending only on λ, ε and b_w such that if* (2.9) *holds, then*

$$\sup_{\varepsilon<p<1-\varepsilon}\left|\hat{c}_p - \left(c_p^G - \frac{\hat{Q}_{n,\gamma}(c_p^G)}{f_\Sigma(c_p^G)}\right)\right| \leq \frac{C}{\sqrt{\log d}}\frac{\varsigma_d^3}{\sigma_*^2}\frac{\log^3(dn)}{n}\log n \tag{5.3}$$

with probability at least $1-1/n$, where

$$\hat{Q}_{n,\gamma}(t) := \int_{A(t)}\{\hat{p}_{n,\gamma}(z) - \phi_\Sigma(z)\}dz.$$

The proofs are based on the following abstract result.

Proposition 5.1 (Abstract Cornish–Fisher type expansion for maximum statistics)**.** *Let $\varepsilon \in (0,1/2)$. Also, let W be a random vector in $\mathbb{R}^d$. Suppose that there exist arrays $U \in (\mathbb{R}^d)^{\otimes 2}, V \in (\mathbb{R}^d)^{\otimes 3}$ and a constant $\Delta > 0$ such that*

$$\sup_{t\in\mathbb{R}}\left|P(W^\vee \leq t) - \int_{A(t)} p_{U,V}(z)dz\right| \leq \Delta, \tag{5.4}$$

where $p_{U,V}(z) = \phi_\Sigma(z) + \langle U, \nabla^2\phi_\Sigma(z)\rangle + \langle V, \nabla^3\phi_\Sigma(z)\rangle$. Set

$$\delta := \|U\|_\infty\frac{\log d}{\sigma_*^2} + \|V\|_\infty\frac{\log^{3/2} d}{\sigma_*^3}, \qquad \overline{\Delta} := \Delta + \frac{\varsigma_d}{\sigma_*}(\Delta+\delta)\delta.$$

Then, there exist positive constants c and C depending only on ε such that if $\delta + \overline{\Delta} \leq c$, then

$$\sup_{\varepsilon<p<1-\varepsilon}\left|F_W^{-1}(p) - \left(F_Z^{-1}(p) - \frac{Q_{U,V}(F_Z^{-1}(p))}{f_\Sigma(F_Z^{-1}(p))}\right)\right| \leq \frac{C}{\sqrt{\log d}}\left(\varsigma_d\overline{\Delta} + \frac{\varsigma_d^3}{\sigma_*^2}(\delta+\overline{\Delta})^2\right),$$

where

$$Q_{U,V}(t) = \int_{A(t)}\{p_{U,V}(z) - \phi_\Sigma(z)\}dz.$$

The proof relies on the following novel isoperimetric-type inequality for $Z^\vee$ and its consequence to the second derivative of F_Z^{-1}:

Lemma 5.1. *Let Z be a centered Gaussian vector in $\mathbb{R}^d$. If $Z^\vee$ has a continuous density f, then*

$$f(F_Z^{-1}(p)) \geq \frac{1}{4\sqrt{\operatorname{Var}[Z^\vee]}} \min\left\{\frac{p}{\sqrt{2}}, (1-p)^{3/2}\right\} \tag{5.5}$$

for all $p \in (0,1)$. Moreover, if $\operatorname{Cov}[Z] = \Sigma$, there exists a universal constant $C > 0$ such that

$$|(F_Z^{-1})''(p)| \leq C \left(\frac{\operatorname{Var}[Z^\vee]}{\min\{p^2, (1-p)^3\}}\right)^{3/2} \frac{\log d}{\sigma_*^2} \tag{5.6}$$

for all $p \in (0,1)$.

Proof. See Appendix C. $\square$

Remark 5.1. Under the first assumption of Lemma 5.1, we can also derive the following Gaussian-type isoperimetric inequality for $Z^\vee$: For all $p \in (0,1)$,

$$f(F_Z^{-1}(p)) \geq \frac{1}{\sigma}\phi(\Phi^{-1}(p)), \tag{5.7}$$

where $\sigma := \max_{1\leq j\leq d}\sqrt{\operatorname{Var}[Z_j]}$. In fact, by [5, Proposition 5], (5.7) follows once we prove

$$\phi(\Phi^{-1}(\mathrm{E}[g(Z^\vee)])) \leq \mathrm{E}\left[\sqrt{\phi(\Phi^{-1}(g(Z^\vee)))^2 + \sigma^2 g'(Z^\vee)^2}\right]$$

for any locally Lipschitz function $g : \mathbb{R} \to [0,1]$. The latter follows by applying Bobkov's functional Gaussian isoperimetric inequality to the function $x \mapsto g(\max_{1\leq j\leq d}(\operatorname{Cov}[Z]^{1/2}x)_j)$ (cf. Eq.(2) of [5]). While (5.7) has a better dependence on p than (5.5), it is often the case that $\operatorname{Var}[Z^\vee] = O(1/\log d)$ as already mentioned at the beginning of Section 2.2, so (5.5) is preferable to (5.7) in terms of the dimension dependence.

Proof of Proposition 5.1. Observe that

$$Q_{U,V}(t) = \langle U, \int_{A(t)} \nabla^2\phi_\Sigma(z)dz\rangle + \langle V, \int_{A(t)} \nabla^3\phi_\Sigma(z)dz\rangle.$$

Hence, by Lemmas D.4 and 4.2, there exists a universal constant $C_1 \geq 1$ such that

$$|Q_{U,V}(t)| \leq C_1\delta \tag{5.8}$$

and

$$|Q_{U,V}(t) - Q_{U,V}(s)| \leq C_1\delta\frac{\sqrt{\log d}}{\sigma_*}|t-s| \tag{5.9}$$

for all $t, s \in \mathbb{R}$. Also, for any $p \in (\varepsilon, 1-\varepsilon)$, we have by (5.4)

$$p \leq F_W(F_W^{-1}(p)) \leq F_Z(F_W^{-1}(p)) + Q_{U,V}(F_W^{-1}(p)) + \Delta \tag{5.10}$$

and

$$p \geq F_W(F_W^{-1}(p)-) \geq F_Z(F_W^{-1}(p)) + Q_{U,V}(F_W^{-1}(p)) - \Delta. \tag{5.11}$$

Combining these bounds with (5.8) gives $p - \Delta - C_1\delta \leq F_Z(F_W^{-1}(p)) \leq p + \Delta + C_1\delta$. Therefore, provided that $\Delta + C_1\delta < \varepsilon/2$, we have by the mean value theorem and (5.5)

$$|F_Z^{-1}(p \pm (\Delta + C_1\delta)) - F_Z^{-1}(p)| \leq C_2\sqrt{\operatorname{Var}[Z^\vee]}(\Delta + C_1\delta)$$

for some constant $C_2 \geq 1$ depending only on ε. Thus we obtain

$$|F_W^{-1}(p) - F_Z^{-1}(p)| \leq C_2\sqrt{\mathrm{Var}[Z^\vee]}(\Delta + C_1\delta).$$

This and (5.9) give

$$|Q_{U,V}(F_W^{-1}(p)) - Q_{U,V}(F_Z^{-1}(p))| \leq C_1C_2\delta\frac{\varsigma_d}{\sigma_*}(\Delta + C_1\delta) =: \Delta'.$$

Combining this with (5.10) and (5.11), we obtain

$$p - Q_{U,V}(F_Z^{-1}(p)) - \Delta - \Delta' \leq F_Z(F_W^{-1}(p)) \leq p - Q_{U,V}(F_Z^{-1}(p)) + \Delta + \Delta'. \tag{5.12}$$

Thus, provided that $C_1\delta + \Delta + \Delta' < \varepsilon/2$, we have by Taylor's theorem and (5.6)

$$\begin{aligned}&\left|F_Z^{-1}(p - Q_{U,V}(F_Z^{-1}(p)) \pm (\Delta + \Delta')) - \left(F_Z^{-1}(p) - \frac{Q_{U,V}(F_Z^{-1}(p)) \mp (\Delta + \Delta')}{f_\Sigma(F_Z^{-1}(p))}\right)\right| \\ &\leq C_3\frac{\varsigma_d^3}{\sigma_*^2\sqrt{\log d}}|Q_{U,V}(F_Z^{-1}(p)) \mp (\Delta + \Delta')|^2\end{aligned}$$

for some constant $C_3 \geq 1$ depending only on ε. Combining this with (5.5), (5.8) and (5.12) gives

$$\begin{aligned}&\left|F_W^{-1}(p) - \left(F_Z^{-1}(p) - \frac{Q_{U,V}(F_Z^{-1}(p))}{f_\Sigma(F_Z^{-1}(p))}\right)\right| \\ &\leq \frac{4\sqrt{\mathrm{Var}[Z^\vee]}}{\varepsilon^{3/2}}(\Delta + \Delta') + C_3\frac{\varsigma_d^3}{\sigma_*^2\sqrt{\log d}}(C_1\delta + \Delta + \Delta')^2.\end{aligned}$$

Since $\Delta + \Delta' \leq C_1^2C_2\overline{\Delta}$, this completes the proof. $\square$

Now we turn to the proof of Theorems 5.1 and 5.2.

Proof of Theorem 5.1. First, observe that Lemma D.3 yields

$$\varsigma_d/\sigma_* \gtrsim \underline{\sigma}/\overline{\sigma} \geq \lambda^{-1}. \tag{5.13}$$

Hence, due to (5.1), we may assume

$$\left(1 + \frac{\varsigma_d^3}{\sigma_*^3}\right)\frac{\log^3 d}{n}\log n \leq 1. \tag{5.14}$$

Then, we have (2.4) by Theorem 2.1. Also, observe that $\mathrm{E}[\overline{X^3}] \lesssim b^3/\sqrt{n}$. Hence, in this setting, δ and $\overline{\Delta}$ in Proposition 5.1 are bounded as

$$\delta \leq C_\lambda\sqrt{\frac{\log^3 d}{n}}, \qquad \overline{\Delta} \leq C_\lambda\left(1 + \frac{\varsigma_d}{\sigma_*}\right)\frac{\log^3 d}{n}\log n,$$

where we used (5.14) for the second inequality. Combining these bounds with (5.13) and (5.14) gives

$$\delta + \overline{\Delta} \leq C_\lambda\sqrt{\frac{\log^3 d}{n}\log n}, \qquad \overline{\Delta} \leq C_\lambda\frac{\varsigma_d}{\sigma_*}\frac{\log^3 d}{n}\log n.$$

Consequently, the desired result follows from Proposition 5.1. $\square$

Proof of Theorem 5.2. By the same reasoning as in the proof of Theorem 5.1, we may assume

$$\left(1+\frac{\varsigma_d^3}{\sigma_*^3}\right)\frac{\log^3(dn)}{n}\log n \le 1. \tag{5.15}$$

Let $\mathcal{E}_n$ be the event defined by (4.21). Recall that $P(\mathcal{E}_n) \ge 1-1/n$. Also, by the proof of Theorem 2.2, we have (2.7) on $\mathcal{E}_n$. Further, recall that we have (4.23) and (4.25) on $\mathcal{E}_n$. Hence, on $\mathcal{E}_n$

$$\frac{1}{2}\|\overline{X^2}-\Sigma\|_\infty\frac{\log d}{\sigma_*^2}+\frac{|\gamma|}{6}\|\overline{X^3}\|_\infty\frac{\log^{3/2} d}{\sigma_*^3} \lesssim b^2\frac{\log^{3/2}(dn)}{\sigma_*^2\sqrt{n}}+\frac{|\gamma|b^3\log^{3/2} d}{\sigma_*^3\sqrt{n}} \le C_{\lambda,b_w}\sqrt{\frac{\log^3(dn)}{n}}.$$

Consequently, a similar argument to that in the proof of Theorem 5.1 gives the desired result. ☐

5.2 Proof of Theorem 2.3

Lemma 5.2. *For any $r\in\mathbb{N}$ and $t\in\mathbb{R}$,*

$$\int_{A(t)^c}\nabla^r\phi_\Sigma(z)dz = -\int_{A(t)}\nabla^r\phi_\Sigma(z)dz.$$

Proof. Let $Z\sim N(0,\Sigma)$. Then, for any $x\in\mathbb{R}^d$,

$$\int_{A(t)^c}\phi_\Sigma(z+x)dz = P(Z-x\in A(t)^c) = 1-P(Z-x\in A(t)) = 1-\int_{A(t)}\phi_\Sigma(z+x)dz.$$

Differentiating both sides r times with respect to x and setting $x=0$, we obtain the desired result. ☐

Lemma 5.3 (Anti-concentration inequality for T_n)**.** *Under the assumptions of Theorem 2.1, there exists a universal constant $C>0$ such that*

$$P\left(t\le T_n\le t+\varepsilon\right)\le C\left(\frac{b^5}{\sigma_*^5}\frac{\log^3 d}{n}\log n+\varepsilon\left(\frac{\sqrt{\log d}}{\underline{\sigma}}+\frac{b^3}{\sigma_*^4}\frac{\log^2 d}{\sqrt{n}}\right)\right)$$

for all $t\in\mathbb{R}$ and $\varepsilon>0$.

Proof. The claim immediately follows by combining Theorem 2.1 with Lemmas D.2 and 4.2. ☐

Lemma 5.4. *Let η be a centered random vector in $\mathbb{R}^r$ having a Stein kernel τ such that $\tau_{jk}=\tau_{kj}$ for all j,k. Assume $\mathrm{E}[|\eta|^2]<\infty$. Also, let L and M be a $d\times r$ matrix and an $r\times r$ symmetric matrix, respectively. Then, for any $a\in\mathbb{R}^d$, $\xi:=L\eta+\langle M,\eta^{\otimes 2}-\mathrm{E}[\eta^{\otimes 2}]\rangle a$ has an approximate Stein kernel $(\bar\tau,\beta)$ such that*

$$\bar\tau(\xi)=\mathrm{E}[L\tau(\eta)L^\top+4(L\tau(\eta)M\eta)\otimes a\mid\xi],$$
$$\beta(\xi)=\mathrm{E}[\langle M,2\tau(\eta)-\eta^{\otimes 2}-\mathrm{E}[\eta^{\otimes 2}]\rangle a\mid\xi].$$

Proof. Define a function $g:\mathbb{R}^r\to\mathbb{R}^d$ as $g(y)=Ly+\langle M,y^{\otimes 2}-\mathrm{E}[\eta^{\otimes 2}]\rangle a$, $y\in\mathbb{R}^r$. Take $h\in C_b^2(\mathbb{R}^d)$ arbitrarily and set $H:=h\circ g$. A straightforward computation shows that $H\in C_b^2(\mathbb{R}^d)$ and

$$\nabla H(y)=L^\top\nabla h(g(y))+2(a\cdot\nabla h(g(y)))My,$$
$$\nabla^2 H(y)=L^\top\nabla^2 h(g(y))L+2\{L^\top\nabla^2 h(g(y))a(My)^\top+My(L^\top\nabla^2 h(g(y))a)^\top\}$$

$$+ 2(a \cdot \nabla h(g(y)))M$$

for every $y \in \mathbb{R}^r$. In particular, $\eta \cdot \nabla H(\eta) = (\xi + \langle M, \eta^{\otimes 2} + \mathrm{E}[\eta^{\otimes 2}]\rangle a) \cdot \nabla h(\xi)$. Hence

$$\begin{aligned}
\mathrm{E}[(\xi - \mathrm{E}[\xi]) \cdot \nabla h(\xi)] &= \mathrm{E}[\eta \cdot \nabla H(\eta)] - \mathrm{E}[\langle M, \eta^{\otimes 2} + \mathrm{E}[\eta^{\otimes 2}]\rangle a \cdot \nabla h(\xi)] \\
&= \mathrm{E}[\langle \tau(\eta), \nabla^2 H(\eta)\rangle] - \mathrm{E}[\langle M, \eta^{\otimes 2} + \mathrm{E}[\eta^{\otimes 2}]\rangle a \cdot \nabla h(\xi)] \\
&= \mathrm{E}[\langle \bar{\tau}(\xi), \nabla^2 h(\xi)\rangle] + \mathrm{E}[\beta(\xi) \cdot \nabla h(\xi)].
\end{aligned}$$

This completes the proof. □

Proof of Theorem 2.3. By the same reasoning as in the proof of Theorem 2.1, we may assume $\tau^X_{i,jk} = \tau^X_{i,kj}$ for all $i \in \{1, \dots, n\}$ and $j, k \in \{1, \dots, d\}$ without loss of generality.

By Theorems 5.1 and 5.2, there exist positive constants c and C depending only on λ, ε and b_w such that if (2.9) holds, then we have (5.2) and (5.3) with probability at least $1 - 1/n$. In the sequel we assume (2.9) is satisfied with this c and fix $\alpha \in (\varepsilon, 1-\varepsilon)$ arbitrarily. By (5.5) and Lemmas D.4 and D.11

$$\frac{1}{f_\Sigma(c^G_{1-\alpha})}\left|\hat{Q}_{n,\gamma}(c^G_{1-\alpha}) - \gamma Q_n(c^G_{1-\alpha}) - \frac{1}{2}\langle \overline{X^2} - \Sigma, \Psi_\alpha\rangle\right| \lesssim \frac{|\gamma|\varsigma_d}{\varepsilon^{3/2}\sqrt{\log d}}\frac{b^3}{\sigma_*^3}\frac{\log^{3/2} d}{n}\sqrt{\log n}$$

with probability at least $1 - 1/n$. Combining this with (5.2) and (5.3), we have

$$\left|\hat{c}_{1-\alpha} - \tilde{c}_{1-\alpha} + \frac{\langle \overline{X^2} - \Sigma, \Psi_\alpha\rangle}{2 f_\Sigma(c^G_{1-\alpha})}\right| \le \frac{C_{\lambda,\varepsilon,b_w}}{\sqrt{\log d}}\frac{\varsigma_d^3}{\sigma_*^2}\frac{\log^3(dn)}{n}\log n \tag{5.16}$$

with probability at least $1 - 2/n$, where $\tilde{c}_{1-\alpha} := c_{1-\alpha} + (1-\gamma)Q_n(c^G_{1-\alpha})/f_\Sigma(c^G_{1-\alpha})$. This and Lemma 5.3 give

$$\begin{aligned}
&\left|P(T_n \ge \hat{c}_{1-\alpha}) - P\left(T_n \ge \tilde{c}_{1-\alpha} - \frac{\langle \overline{X^2} - \Sigma, \Psi_\alpha\rangle}{2 f_\Sigma(c^G_{1-\alpha})}\right)\right| \\
&\le C_{\lambda,\varepsilon,b_w}\left(\frac{\log^3 d}{n}\log n + \frac{\varsigma_d^3}{\sigma_*^3}\frac{\log^3(dn)}{n}(\log n)\left(1 + \sqrt{\frac{\log^3 d}{n}}\right)\right) + \frac{1}{n} \\
&\le C_{\lambda,\varepsilon,b_w}\frac{\varsigma_d^3}{\sigma_*^3}\frac{\log^3(dn)}{n}\log n,
\end{aligned} \tag{5.17}$$

where the second inequality follows by (2.9). Now, observe that

$$\widetilde{T}_n := T_n + \frac{\langle \overline{X^2} - \Sigma, \Psi_\alpha\rangle}{2 f_\Sigma(c^G_{1-\alpha})} = \max_{1 \le j \le d}\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_{ij} + U_i), \tag{5.18}$$

where

$$U_i := \frac{\langle X_i^{\otimes 2} - \mathrm{E}[X_i^{\otimes 2}], \widetilde{\Psi}_\alpha\rangle}{\sqrt{n}}, \qquad \widetilde{\Psi}_\alpha := \frac{\Psi_\alpha}{2 f_\Sigma(c^G_{1-\alpha})}. \tag{5.19}$$

By Lemma 5.4, for every i, $\xi_i := (X_i + U_i\mathbf{1}_d)/\sqrt{n}$ has an approximate Stein kernel (τ_i, β_i) such that

$$\tau_i(\xi_i) = n^{-1}\,\mathrm{E}[\tau^X_i(X_i) + 4n^{-1/2}\tau^X_i(X_i)(\widetilde{\Psi}_\alpha X_i) \otimes \mathbf{1}_d \mid \xi_i],$$

$$\beta_i(\xi_i) = n^{-1}\,\mathrm{E}[\langle \widetilde{\Psi}_\alpha, 2\tau_i^X(X_i) - X_i^{\otimes 2} - \mathrm{E}[X_i^{\otimes 2}]\rangle \mathbf{1}_d \mid \xi_i].$$

Hence we can derive an Edgeworth expansion for $\widetilde{T}_n$ by applying Theorem 4.1 to ξ_i. We are going to bound the quantities appearing on the right hand side of (4.2). First, by (5.5) and Lemma D.4

$$\|\widetilde{\Psi}_\alpha\|_1 \le C_\varepsilon \frac{\varsigma_d\sqrt{\log d}}{\sigma_*^2}. \tag{5.20}$$

Hence, by Lemma D.7 and (2.9),

$$\|U_i\|_{\psi_{1/2}} \le C_\varepsilon \frac{b^2}{\sqrt{n}}\frac{\varsigma_d\sqrt{\log d}}{\sigma_*^2} \le C_{\lambda,\varepsilon,b_w}\frac{b}{\log d}, \tag{5.21}$$

and

$$\max_{j,k}\left\| n^{-1/2}\left(\tau_i^X(X_i)(\widetilde{\Psi}_\alpha X_i)\otimes \mathbf{1}_d\right)_{jk}\right\|_{\psi_{1/3}} \le C_\varepsilon \frac{b^3\varsigma_d\sqrt{\log d}}{\sigma_*^2\sqrt{n}} \le C_{\lambda,\varepsilon,b_w}\frac{b^2}{\log d}. \tag{5.22}$$

These estimates allow us to prove (4.15)–(4.17) with b replaced by $C_{\lambda,\varepsilon,b_w}b$ in a similar manner to the proof of Theorem 2.1. Further, observe that

$$\max_j |\,\mathrm{E}[X_{ij}U_i]| \le C_{\lambda,\varepsilon}\frac{b^2}{\sqrt{n}}\frac{\varsigma_d\sqrt{\log d}}{\sigma_*}, \qquad |\,\mathrm{E}[U_i^2]| \le C_{\lambda,\varepsilon,b_w}\frac{b^2}{\sqrt{n}}\frac{\varsigma_d}{\sigma_*\sqrt{\log d}}.$$

Combining these estimates with (5.22), we can also prove (4.14) with b replaced by $C_{\lambda,\varepsilon}b\sqrt{\varsigma_d/\sigma_*}$ similarly to the proof of Theorem 2.1. Next, by Lemma D.7 and (5.20),

$$\max_j \|\beta_{ij}(\xi_i)\|_{\psi_{1/2}} \le C_\varepsilon \frac{b^2}{n}\frac{\varsigma_d\sqrt{\log d}}{\sigma_*^2}.$$

Therefore, similarly to the proof of (4.15), we obtain

$$\frac{\log^2 d}{\sigma_*^4}\left(\mathrm{E}\left\|\sum_{i=1}^n \xi_i^{\otimes 3}\otimes\beta_i(\xi_i)\right\|_\infty + \mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)\otimes\xi_i\otimes\tau_i(\xi_i)\right\|_\infty\right) \le C_{\lambda,\varepsilon,b_w}\frac{\varsigma_d}{\sigma_*}\frac{\log^3 d}{n}$$

and

$$\frac{\log^{3/2} d}{\sigma_*^3}\,\mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)^{\otimes 2}\otimes\xi_i\right\|_\infty \le C_{\lambda,\varepsilon,b_w}\frac{\varsigma_d^2}{\sigma_*^2}\frac{\log^3 d}{n}$$

and

$$\frac{\log^2 d}{\sigma_*^4}\,\mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)\otimes\xi_i\right\|_\infty^2 \le C_{\lambda,\varepsilon,b_w}\frac{\varsigma_d^2}{\sigma_*^2}\frac{\log^3 d}{n}. \tag{5.23}$$

In addition, since $\beta_i(\xi_i)$ has identical components,

$$\frac{\log d}{\sigma_*^2}\left(\mathrm{E}\,\|B\|_\infty^2 + \mathrm{E}\left\|\sum_{i=1}^n \beta_i(\xi_i)^{\otimes 2}\right\|_\infty\right) = \frac{2\log d}{\sigma_*^2}\sum_{i=1}^n \mathrm{E}[\beta_{i1}(\xi_i)^2] \le C_{\lambda,\varepsilon}\frac{\log^3 d}{n}. \tag{5.24}$$

Combining (5.23) and (5.24) with (4.14), (4.18) and the AM-GM inequality gives

$$\frac{\log^2 d}{\sigma_*^4}\left(\mathrm{E}\left\|B\otimes\sum_{i=1}^n \xi_i^{\otimes 3}\right\|_\infty + \mathrm{E}\left\|\bar{T}\otimes\sum_{i=1}^n \beta_i(\xi_i)\otimes\xi_i\right\|_\infty\right) \le C_{\lambda,\varepsilon,b_w}\frac{\varsigma_d^2}{\sigma_*^2}\frac{\log^3 d}{n}$$

and

$$\frac{\log^{3/2} d}{\sigma_*^3} \mathrm{E}\left\| B \otimes \sum_{i=1}^n \beta_i(\xi_i) \otimes \xi_i \right\|_\infty \le C_{\lambda,\varepsilon,b_w} \frac{\varsigma_d^2}{\sigma_*^2} \frac{\log^3 d}{n}.$$

All together, we can proceed as in the proof of Theorem 2.1 and then obtain

$$\sup_{t\in\mathbb{R}} \left| P(\widetilde{T}_n \le t) - \int_{A(t)} (p_n(z) + q_n(z))\, dz \right| \le C_{\lambda,\varepsilon,b_w} \frac{\varsigma_d^2}{\sigma_*^2} \frac{\log^3 d}{n} \log n,$$

where

$$q_n(z) = \frac{1}{2n} \sum_{i=1}^n \langle 2\, \mathrm{E}[X_i U_i] \mathbf{1}_d^\top + \mathrm{E}[U_i^2] \mathbf{1}_d^{\otimes 2}, \nabla^2 \phi_\Sigma(z) \rangle - \frac{1}{6n^{3/2}} \sum_{i=1}^n \langle \mathrm{E}[(X_i + U_i \mathbf{1}_d)^{\otimes 3} - X_i^{\otimes 3}], \nabla^3 \phi_\Sigma(z) \rangle.$$

Therefore, in view of Theorem 2.1, it remains to prove

$$\left| \int_{A(\tilde{c}_{1-\alpha})^c} p_n(z) dz - \int_{A(c_{1-\alpha})^c} p_n(z) dz + (1-\gamma) Q_n(c_{1-\alpha}^G) \right| \le C_{\lambda,\varepsilon,b_w} \frac{\varsigma_d^2}{\sigma_*^2} \frac{\log^3(dn)}{n}, \tag{5.25}$$

$$\left| \int_{A(\tilde{c}_{1-\alpha})^c} q_n(z) dz + \mathrm{E}[R_n(\alpha)] \right| \le C_{\lambda,\varepsilon,b_w} \frac{\varsigma_d^3}{\sigma_*^3} \frac{\log^3(dn)}{n} \log n. \tag{5.26}$$

Let us prove (5.25). By Lemma 4.2,

$$\left| \int_{A(\tilde{c}_{1-\alpha})^c} p_n(z) dz - \int_{A(c_{1-\alpha})^c} p_n(z) dz + \{ F_Z(\tilde{c}_{1-\alpha}) - F_Z(c_{1-\alpha}) \} \right| \lesssim \frac{|1-\gamma| b^3 \log^2 d}{\sigma_*^4 \sqrt{n}} \left| \frac{Q_n(c_{1-\alpha}^G)}{f_\Sigma(c_{1-\alpha}^G)} \right|.$$

Also, by Taylor's theorem and Lemma D.5,

$$\left| F_Z(\tilde{c}_{1-\alpha}) - F_Z(c_{1-\alpha}) - (1-\gamma) Q_n(c_{1-\alpha}^G) \right| \lesssim \frac{|1-\gamma| \log d}{\sigma_*^2} \left| \frac{Q_n(c_{1-\alpha}^G)}{f_\Sigma(c_{1-\alpha}^G)} \right|^2.$$

Further, Lemma D.4 and (5.5) yield

$$\left| \frac{Q_n(c_{1-\alpha}^G)}{f_\Sigma(c_{1-\alpha}^G)} \right| \lesssim \frac{\varsigma_d}{\sqrt{\varepsilon^3 \log d}} \cdot \frac{b^3}{\sqrt{n}} \cdot \frac{\log^{3/2} d}{\sigma_*^3} \le C_{\lambda,\varepsilon} \frac{\varsigma_d \log d}{\sqrt{n}}. \tag{5.27}$$

Combining these three estimates gives (5.25). Next, to prove (5.26), consider the following decomposition:

$$\begin{aligned}
&\int_{A(\tilde{c}_{1-\alpha})^c} q_n(z) dz \\
&= \frac{1}{n} \sum_{i=1}^n \langle \mathrm{E}[X_i U_i] \mathbf{1}_d^\top, \int_{A(\tilde{c}_{1-\alpha})^c} \nabla^2 \phi_\Sigma(z) dz \rangle + \frac{1}{2n} \sum_{i=1}^n \langle \mathrm{E}[U_i^2] \mathbf{1}_d^{\otimes 2}, \int_{A(\tilde{c}_{1-\alpha})^c} \nabla^2 \phi_\Sigma(z) dz \rangle \\
&\quad - \frac{1}{6n^{3/2}} \sum_{i=1}^n \langle \mathrm{E}[(X_i + U_i \mathbf{1}_d)^{\otimes 3} - X_i^{\otimes 3}], \int_{A(\tilde{c}_{1-\alpha})^c} \nabla^3 \phi_\Sigma(z) dz \rangle \\
&=: I + II + III.
\end{aligned}$$

We can rewrite I as

$$
\begin{aligned}
I &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j,k=1}^{d} \mathrm{E}[X_{ij}U_i]\int_{A(\tilde{c}_{1-\alpha})^c} \partial_{jk}\phi_\Sigma(z)dz \\
&= \frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j,k,l,m=1}^{d} \mathrm{E}[X_{ij}X_{il}X_{im}]\widetilde{\Psi}_{\alpha,lm}\int_{A(\tilde{c}_{1-\alpha})^c} \partial_{jk}\phi_\Sigma(z)dz \\
&= \frac{1}{\sqrt{n}}\langle \mathrm{E}[\overline{X^3}]\otimes \mathbf{1}_d, \widetilde{\Psi}_\alpha \otimes \int_{A(\tilde{c}_{1-\alpha})^c} \nabla^2\phi_\Sigma(z)dz\rangle = -\frac{1}{\sqrt{n}}\langle \mathrm{E}[\overline{X^3}]\otimes \mathbf{1}_d, \widetilde{\Psi}_\alpha \otimes \int_{A(\tilde{c}_{1-\alpha})} \nabla^2\phi_\Sigma(z)dz\rangle,
\end{aligned}
$$

where we used Lemma 5.2 for the last equality. We are going to prove $\tilde{c}_{1-\alpha}$ in the last expression can be replaced by $c^G_{1-\alpha}$. By Lemma 4.2 and (5.20),

$$
\left\|\widetilde{\Psi}_\alpha \otimes \left(\int_{A(\tilde{c}_{1-\alpha})} \nabla^2\phi_\Sigma(z)dz - \int_{A(c^G_{1-\alpha})} \nabla^2\phi_\Sigma(z)dz\right)\right\|_1 \lesssim \frac{\varsigma_d \log^2 d}{\sigma_*^5}|\tilde{c}_{1-\alpha} - c^G_{1-\alpha}|.
$$

Also, by (5.2) and (5.27),

$$
|\tilde{c}_{1-\alpha} - c^G_{1-\alpha}| \le C_{\lambda,\varepsilon,b_w}\left(\frac{\varsigma_d \log d}{\sqrt{n}} + \frac{\varsigma_d^3}{\sigma_*^2}\frac{\log^{5/2}(dn)}{n}\log n\right).
$$

Consequently, we deduce

$$
|I + \mathrm{E}[R_n(\alpha)]| \le C_{\lambda,\varepsilon,b_w}\frac{b^3\varsigma_d \log^2 d}{\sigma_*^5\sqrt{n}}|\tilde{c}_{1-\alpha} - c^G_{1-\alpha}| \le C_{\lambda,\varepsilon,b_w}\frac{\varsigma_d^3}{\sigma_*^3}\frac{\log^3(dn)}{n}\log n,
$$

where we also used (2.9) and (5.13) for the last inequality. Meanwhile, by Lemma D.4 and (5.21),

$$
|II| \le C_\varepsilon \frac{b^4}{n}\cdot\frac{\varsigma_d^2 \log d}{\sigma_*^4}\frac{\log d}{\sigma_*^2} \le C_{\lambda,\varepsilon}\frac{\varsigma_d^2}{\sigma_*^2}\frac{\log^2 d}{n}
$$

and

$$
\begin{aligned}
|III| &\lesssim \frac{1}{\sqrt{n}}\max_{1\le i\le n}\left(\|\mathrm{E}[X_i^{\otimes 2}U_i]\|_\infty + \|\mathrm{E}[X_iU_i^2]\|_\infty + |\mathrm{E}[U_i^3]|\right)\frac{\log^{3/2} d}{\sigma_*^3} \\
&\le C_{\lambda,\varepsilon,b_w}\frac{b^4}{n}\frac{\varsigma_d\sqrt{\log d}}{\sigma_*^2}\frac{\log^{3/2} d}{\sigma_*^3} \le C_{\lambda,\varepsilon,b_w}\frac{\varsigma_d}{\sigma_*}\frac{\log^2 d}{n}.
\end{aligned}
$$

All together, we complete the proof. □

5.3 *Proof of Corollary 2.2*

First, replacing X_i by X_i/σ, we may assume $\sigma = 1$ without loss of generality. Note that we have $\sigma_*^{-1} \lesssim \lambda$ under this assumption. Next, set $\rho := \max_{1\le j<k\le d}|\Sigma_{jk}| < 1$. Note that we have $1-\rho^2 \ge \sigma_*^2$ because $1-\rho^2$ coincides with the minimum principal minor of Σ of size 2.

We begin by proving the following inequalities for every $\alpha \in (\varepsilon, 1-\varepsilon)$:

$$
|c^G_{1-\alpha}| \le C_\varepsilon\sqrt{\log d}, \qquad \exp\left(-(c^G_{1-\alpha})^2/2\right) \le C_{\lambda,\varepsilon,K}d^{-1}\sqrt{\log d}. \tag{5.28}
$$

The first one is an immediate consequence of Lemma A.6 in [6]. Meanwhile, by Eq.(4.2.9) in [43] (see also Eq.(4.2.1) in [43]), we have for any $u > 0$

$$|P(Z^\vee \le u) - \Phi(u)^d| \le C_\lambda \sum_{1\le j<k\le d} |\Sigma_{jk}| \exp\left(-\frac{u^2}{1+|\Sigma_{jk}|}\right).$$

Set $\mathcal{I} := \{(j,k) \in \{1,\dots,d\}^2 : |\Sigma_{jk}| > u^{-2}\}$. Observe that $dK^2 \ge \sum_{j,k=1}^d \Sigma_{jk}^2 \ge u^{-2}\sum_{(j,k)\in\mathcal{I}} |\Sigma_{jk}|$. Hence we obtain

$$\begin{aligned}|P(Z^\vee \le u) - \Phi(u)^d| &\le C_\lambda \sum_{(j,k)\in\mathcal{I}} |\Sigma_{jk}| e^{-u^2/(1+\rho)} + C_\lambda \sum_{(j,k)\notin\mathcal{I}} |\Sigma_{jk}| e^{-u^2/(1+u^{-2})} \\ &\le C_\lambda dK^2 u^2 e^{-u^2/(1+\rho)} + C_\lambda d^{3/2} K e^{-u^2}.\end{aligned}$$

Further, observe that $1-(\varepsilon/2)^{1/d} < -d^{-1}\log(\varepsilon/2) \le 1/2$. Thus, with $u = \Phi^{-1}((\varepsilon/2)^{1/d}) = -\Phi^{-1}(1-(\varepsilon/2)^{1/d})$, we have by Lemma 10.3 in [12]

$$e^{-u^2/2} = \sqrt{2\pi}\phi(u) \le -\sqrt{2\pi} d^{-1}\log(\varepsilon/2)\sqrt{2\log(-d/\log(\varepsilon/2))} \le C_\varepsilon d^{-1}\sqrt{\log d}. \tag{5.29}$$

In addition, by the well-known inequality $\Phi(-s) = 1-\Phi(s) \le e^{-s^2/2}$ for all $s \ge 0$, we deduce $u \le \sqrt{-2\log(1-(\varepsilon/2)^{1/d})} \le C_\varepsilon\sqrt{\log d}$. Hence we obtain

$$P(Z^\vee \le u) \le \varepsilon/2 + C_{\lambda,\varepsilon}\left(d^{-(1-\rho)/(1+\rho)} K^2 (\log d)^{1+1/(1+\rho)} + Kd^{-1/2}\log d\right).$$

Therefore, if the second term on the right hand side is less than $\varepsilon/2$, then $P(Z^\vee \le u) \le \varepsilon < 1-\alpha$, so $c^G_{1-\alpha} > u$. Hence the second bound in (5.28) follows by (5.29). Otherwise, we have $d \le d_0$ for some constant d_0 depending only on λ, ε and K, so the second bound in (5.28) trivially holds with sufficiently large $C_{\lambda,\varepsilon,K}$.

Now we turn to the main body of the proof. Since the left hand side of (2.10) is bounded by 1, we may assume (2.9) holds with the constant c in Theorem 2.3. Also, we have $f_\Sigma(c^G_{1-\alpha}) \ge C_\varepsilon \varsigma_d/\sqrt{\log d}$ by (5.5). Thus, the proof completes once we show that

$$|\operatorname{E}[\langle \overline{X^3} \otimes \mathbf{1}_d, \Psi_\alpha^{\otimes 2}\rangle]| \le C_{\lambda,\varepsilon,K} d^{-1/2}\log^2 d \tag{5.30}$$

for any $\alpha \in (\varepsilon, 1-\varepsilon)$. Observe that

$$\langle \overline{X^3} \otimes \mathbf{1}_d, \Psi_\alpha^{\otimes 2}\rangle = \frac{1}{n}\sum_{i=1}^n \left\langle X_i^{\otimes 3} \otimes \mathbf{1}_d, \Psi_\alpha^{\otimes 2}\right\rangle = \frac{1}{n}\sum_{i=1}^n \langle X_i^{\otimes 2}, \Psi_\alpha\rangle \langle X_i \otimes \mathbf{1}_d, \Psi_\alpha\rangle.$$

Thus, by the Schwarz inequality

$$|\operatorname{E}[\langle \overline{X^3} \otimes \mathbf{1}_d, \Psi_\alpha^{\otimes 2}\rangle]| \le \sqrt{\frac{1}{n}\sum_{i=1}^n \operatorname{E}[\langle X_i^{\otimes 2}, \Psi_\alpha\rangle^2] \frac{1}{n}\sum_{i=1}^n \operatorname{E}[\langle X_i \otimes \mathbf{1}_d, \Psi_\alpha\rangle^2]}.$$

We have

$$\frac{1}{n}\sum_{i=1}^n \operatorname{E}[\langle X_i^{\otimes 2}, \Psi_\alpha\rangle^2] = \frac{1}{n}\sum_{i=1}^n \left\langle \operatorname{E}[X_i^{\otimes 4}], \Psi_\alpha^{\otimes 2}\right\rangle \le b^4 \|\Psi_\alpha\|_1^2 \lesssim \lambda^4 \log^2 d,$$

where the last inequality follows from Lemma D.4. Also,

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\langle X_i\otimes\mathbf{1}_d,\Psi_\alpha\rangle^2]=\frac{1}{n}\sum_{i=1}^{n}\sum_{j,k,l,m=1}^{d}\mathrm{E}[X_{ij}X_{ik}]\Psi_{\alpha,jl}\Psi_{\alpha,km}$$
$$=\sum_{j,k,l,m=1}^{d}\Sigma_{jk}\Psi_{\alpha,jl}\Psi_{\alpha,km}\leq K\sum_{j=1}^{d}\left(\sum_{l=1}^{d}\Psi_{\alpha,jl}\right)^2.$$

Therefore, (5.30) follows once we show

$$\sum_{j=1}^{d}\left(\sum_{l=1}^{d}\Psi_{\alpha,jl}\right)^2\leq C_{\lambda,\varepsilon,K}d^{-1}\log^2 d. \tag{5.31}$$

Below we write $t=c_{1-\alpha}^G$ for short. Fix $j\in\{1,\dots,d\}$ arbitrarily. A straightforward computation shows

$$\Psi_{\alpha,jj}=-\mathrm{E}[(\Sigma^{-1}Z)_j 1_{\{\max_{k:k\neq j}Z_k\leq t\}}\mid Z_j=t]\phi(t),$$

where $Z\sim N(0,\Sigma)$. Hence

$$|\Psi_{\alpha,jj}|\leq\mathrm{E}[|(\Sigma^{-1}Z)_j|\mid Z_j=t]\phi(t)\leq(\Sigma^{-1})_{jj}|t|\phi(t)+\mathrm{E}[|(\Sigma^{-1})_{j,-j}\cdot Z_{-j}|\mid Z_j=t]\phi(t), \tag{5.32}$$

where $(\Sigma^{-1})_{j,-j}:=((\Sigma^{-1})_{jk})_{1\leq k\leq d:k\neq j}$ and $Z_{-j}:=(Z_k)_{1\leq k\leq d:k\neq j}$. It is well-known that the conditional distribution of Z_{-j} given $Z_j=t$ is $N(t\Sigma_{j,-j},\Sigma_{-j,-j}-\Sigma_{j,-j}^{\otimes 2})$ with $\Sigma_{j,-j}:=(\Sigma_{jk})_{1\leq k\leq d:k\neq j}$ and $\Sigma_{-j,-j}:=(\Sigma_{kl})_{1\leq k,l\leq d:k,l\neq j}$ (see e.g. Theorem 1.2.5 in [36]). Therefore,

$$\mathrm{E}[(\Sigma^{-1})_{j,-j}\cdot Z_{-j}\mid Z_j=t]=(\Sigma^{-1})_{j,-j}\cdot t\Sigma_{j,-j}=t\sum_{k:k\neq j}(\Sigma^{-1})_{jk}\Sigma_{jk}=t\left(1-(\Sigma^{-1})_{jj}\right)$$

and

$$\begin{aligned}\mathrm{Var}[(\Sigma^{-1})_{j,-j}\cdot Z_{-j}\mid Z_j=t]&=(\Sigma^{-1})_{j,-j}^\top(\Sigma_{-j,-j}-\Sigma_{j,-j}^{\otimes 2})(\Sigma^{-1})_{j,-j}\\&=\sum_{k,l:k,l\neq j}(\Sigma^{-1})_{jk}(\Sigma^{-1})_{jl}(\Sigma_{kl}-\Sigma_{jk}\Sigma_{jl})\\&=\sum_{k:k\neq j}(\Sigma^{-1})_{jk}\left(-(\Sigma^{-1})_{jj}\Sigma_{kj}-\Sigma_{jk}+\Sigma_{jk}(\Sigma^{-1})_{jj}\right)\\&=(\Sigma^{-1})_{jj}-1.\end{aligned}$$

Consequently,

$$|\,\mathrm{E}[|(\Sigma^{-1})_{j,-j}\cdot Z_{-j}|\mid Z_j=t]|\leq\left|t\left(1-(\Sigma^{-1})_{jj}\right)\right|+\sqrt{(\Sigma^{-1})_{jj}-1}\leq C_\lambda(1+|t|). \tag{5.33}$$

Combining (5.32) and (5.33) with (5.28) gives

$$|\Psi_{\alpha,jj}|\leq C_\lambda(1+|t|)\phi(t)\leq C_{\lambda,\varepsilon,K}d^{-1}\log d. \tag{5.34}$$

Next, fix $l\in\{1,\dots,d\}\setminus\{j\}$ arbitrarily. Then we have

$$|\Psi_{\alpha,jl}|\leq P\left(\max_{k:k\neq j,l}Z_k\leq t\mid Z_j=t,Z_l=t\right)\phi_{\Sigma[j,l]}(t,t)\leq\phi_{\Sigma[j,l]}(t,t),$$

where $\Sigma[j,l] = (\Sigma_{pq})_{p,q\in\{j,l\}}$. Hence

$$|\Psi_{\alpha,jl}| \le \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{t^2}{1+\Sigma_{jl}}\right).$$

Let $\mathcal{I}_j := \{k \in \{1,\dots,d\} : |\Sigma_{jk}| > t^{-2}\}$. Then we have

$$K^2 \ge \sum_{k=1}^d \Sigma_{jk}^2 \ge \sum_{k\in\mathcal{I}_j} \Sigma_{jk}^2 > t^{-4}\#\mathcal{I}_j,$$

where $\#\mathcal{I}_j$ is the number of elements in $\mathcal{I}_j$. Hence we obtain

$$\begin{aligned}\sum_{l:l\neq j} |\Psi_{\alpha,jl}| &\le C_\lambda \sum_{l\in\mathcal{I}_j} e^{-t^2/(1+\rho)} + C_\lambda \sum_{l\notin\mathcal{I}_j} e^{-t^2/(1+t^{-2})} \le C_\lambda(K^2t^4e^{-t^2/(1+\rho)} + de^{-t^2})\\ &\le C_{\lambda,\varepsilon,K}\left(d^{-2/(1+\rho)}(\log d)^{2+1/(1+\rho)} + d^{-1}\log d\right) \le C_{\lambda,\varepsilon,K} d^{-1}\log d,\end{aligned}$$

where the third inequality follows from (5.28). Combining this with (5.34) gives (5.31). $\square$

5.4 *Proof of Corollary 2.3*

In view of Theorem 2.3, the asserted claim immediately follows from (4.44), (5.30) and (5.5). $\square$

5.5 *Proof of Corollary 2.4*

Let us prove (2.11). Since $\sigma_* \ge \sqrt{1-\rho}$ and $\varsigma_d = O(b\sqrt{\log d})$, we have $\frac{\varsigma_d^3}{\sigma_*^3}\frac{\log^3(dn)}{n}\log n = o(n^{-1/2})$ by assumption. Also, observe that

$$\begin{aligned}&(1-\gamma)Q_n(c_{1-\alpha}^G) + \mathrm{E}[R_n(\alpha)]\\ &= -(1-\gamma)\frac{\rho^{3/2}\,\mathrm{E}[U^3]}{6\sqrt{n}}\langle \mathbf{1}_d^{\otimes 3}, \int_{A(c_{1-\alpha}^G)} \nabla^3\phi_\Sigma(z)dz\rangle + \frac{\rho^{3/2}\,\mathrm{E}[U^3]}{2\sqrt{n}}\frac{\left\langle \mathbf{1}_d^{\otimes 2}, \Psi_\alpha\right\rangle^2}{f_\Sigma(c_{1-\alpha}^G)} + (1-\rho)^{3/2}\Upsilon_{n,\gamma}(\alpha).\end{aligned}$$

Therefore, in view of Theorem 2.3, it suffices to prove

$$\begin{aligned}\langle \mathbf{1}_d^{\otimes 3}, \int_{A(c_{1-\alpha}^G)} \nabla^3\phi_\Sigma(z)dz\rangle &= \rho^{-\frac{3}{2}}(z_\alpha^2-1)\phi(z_\alpha) + o(1),\\ \frac{\left\langle \mathbf{1}_d^{\otimes 2}, \Psi_\alpha\right\rangle^2}{f_\Sigma(c_{1-\alpha}^G)} &= \rho^{-\frac{3}{2}} z_\alpha^2\phi(z_\alpha) + o(1).\end{aligned}$$

Observe that Z has the same law as $\sqrt{\rho}\zeta\mathbf{1}_d + \sqrt{1-\rho}G$ with $\zeta \sim N(0,1)$ and $G \sim N(0, I_d)$ independent. Hence, $F_Z(t) = \mathrm{E}[\Phi((t-\sqrt{1-\rho}G^\vee)/\sqrt{\rho})]$ for every $t\in\mathbb{R}$. Set $a_d := \sqrt{1-\rho}\,\mathrm{E}[G^\vee]$. Since $\mathrm{Var}[G^\vee] = O(1/\log d) = o(1)$, we have $F_Z(t) = \Phi((t-a_d)/\sqrt{\rho}) + o(1)$ uniformly in $t\in\mathbb{R}$ by the Lipschitz continuity of Φ. Hence $1-\alpha = F_Z(c_{1-\alpha}^G) = \Phi((c_{1-\alpha}^G - a_d)/\sqrt{\rho}) + o(1)$ and thus $(c_{1-\alpha}^G - a_d)/\sqrt{\rho} = \Phi^{-1}(1-\alpha) + o(1) = -z_\alpha + o(1)$. Then, for any integer $r\ge 0$, the Lipschitz continuity of $\phi^{(r)}$ gives

$$\begin{aligned}f_\Sigma^{(r)}(c_{1-\alpha}^G) &= \rho^{-\frac{r+1}{2}}\,\mathrm{E}[\phi^{(r)}((c_{1-\alpha}^G - \sqrt{1-\rho}G^\vee)/\sqrt{\rho})]\\ &= (-1)^r\rho^{-\frac{r+1}{2}}\phi^{(r)}(z_\alpha) + o(1).\end{aligned} \tag{5.35}$$

Combining this with (D.1) gives

$$\langle \mathbf{1}_d^{\otimes 3}, \int_{A(c_{1-\alpha}^G)} \nabla^3 \phi_\Sigma(z)dz\rangle = \rho^{-\frac{3}{2}}\phi''(z_\alpha) + o(1) = \rho^{-\frac{3}{2}}(z_\alpha^2 - 1)\phi(z_\alpha) + o(1)$$

and

$$\frac{\langle \mathbf{1}_d^{\otimes 2}, \Psi_\alpha\rangle^2}{f_\Sigma(c_{1-\alpha}^G)} = \frac{\rho^{-2}\phi'(z_\alpha)^2}{\rho^{-\frac{1}{2}}\phi(z_\alpha)} + o(1) = \rho^{-\frac{3}{2}} z_\alpha^2 \phi(z_\alpha) + o(1).$$

This completes the proof of (2.11).

Next we prove $\Upsilon_{n,\gamma}(\alpha) = o(n^{-1/2})$ when $\gamma = 1$. In view of (5.35) with $r = 0$, it suffices to prove

$$\langle \mathrm{E}[V^{\otimes 3}] \otimes \mathbf{1}_d, \Psi_\alpha^{\otimes 2}\rangle = o(1).$$

Observe that

$$\begin{aligned} |\langle \mathrm{E}[V^{\otimes 3}] \otimes \mathbf{1}_d, \Psi_\alpha^{\otimes 2}\rangle| &= |\mathrm{E}[\langle V^{\otimes 2}, \Psi_\alpha\rangle \langle V \otimes \mathbf{1}_d, \Psi_\alpha\rangle]| \\ &\le \sqrt{\mathrm{E}[\langle V^{\otimes 2}, \Psi_\alpha\rangle^2]\,\mathrm{E}[\langle V \otimes \mathbf{1}_d, \Psi_\alpha\rangle^2]}. \end{aligned}$$

We have $\mathrm{E}[\langle V^{\otimes 2}, \Psi_\alpha\rangle^2] = \langle \mathrm{E}[V^{\otimes 4}], \Psi_\alpha^{\otimes 2}\rangle = O(\log^2 d)$ by Lemma D.4. Meanwhile, observe that ϕ_Σ is symmetric as a d-variate function due to the structure of Σ. Hence, $\Psi_{\alpha,jj} = \Psi_{\alpha,11}$ and $\Psi_{\alpha,jk} = \Psi_{\alpha,12}$ if $j \neq k$. Therefore,

$$\begin{aligned} \mathrm{E}[\langle V \otimes \mathbf{1}_d, \Psi_\alpha\rangle^2] &= \mathrm{E}\left[\left(\sum_{j=1}^d V_j\right)^2\right] (\Psi_{\alpha,11} + (d-1)\Psi_{\alpha,12})^2 \\ &= d\,(\Psi_{\alpha,11} + (d-1)\Psi_{\alpha,12})^2 = \frac{1}{d}\left(\sum_{j,k=1}^d \Psi_{jk}\right)^2. \end{aligned}$$

Since $\sum_{j,k=1}^d \Psi_{jk} = O(1)$ by (D.1) and (5.35), we conclude $\mathrm{E}[\langle V \otimes \mathbf{1}_d, \Psi_\alpha\rangle^2] = O(d^{-1})$. Consequently, we obtain $\langle \mathrm{E}[V^{\otimes 3}] \otimes \mathbf{1}_d, \Psi_\alpha^{\otimes 2}\rangle = O(d^{-1/2}\log d) = o(1)$. □

5.6 Proof of Theorem 2.4

For $\alpha \in (0,1)$, we denote by $\hat{c}_{1-\alpha}^*$ the $(1-\alpha)$-quantile of T_n^{**} under P^{**}.

Lemma 5.5. *Under the assumptions of Theorem 2.4, there exist positive constants c and C depending only on $\lambda, \varepsilon, b_w$ and b_v such that if* (2.9) *holds, then*

$$\sup_{\varepsilon<\alpha<1-\varepsilon} P\left(\left|P^*(T_n^* \ge \hat{c}_{1-\alpha}^*) - (\alpha - R_n(\alpha))\right| > C\frac{\varsigma_d^3}{\sigma_*^3}\frac{\log^3(dn)}{n}\log n\right) \le \frac{5}{n}.$$

Proof. The proof is basically a straightforward modification of that of Theorem 2.3. We only give a sketch of the proof with emphasis on relatively major changes.

Fix $\alpha \in (\varepsilon, 1-\varepsilon)$ arbitrarily. First, it is not difficult to see that an analogous result to Theorem 5.2 holds for T_n^{**}. Thus, by a similar argument to the proof of (5.16), we can find a constant c depending only on $\lambda, \varepsilon, b_w$ and b_v and an event $\mathcal{E}_n^*(\alpha)$ satisfying the following conditions:

(i) If (2.9) holds, then we have on $\mathcal{E}_n^*(\alpha)$

$$\left|\hat{c}_{1-\alpha}^* - \hat{c}_{1-\alpha} + \frac{1}{2f_\Sigma(c_{1-\alpha}^G)}\langle\frac{1}{n}\sum_{i=1}^n(w_i^2-1)X_i^{\otimes 2}, \Psi_\alpha\rangle\right| \leq \frac{C_{\lambda,\varepsilon,b_w,b_v}}{\sqrt{\log d}}\frac{\varsigma_d^3}{\sigma_*^2}\frac{\log^3(dn)}{n}\log n.$$

(ii) (2.7) holds on $\mathcal{E}_n^*(\alpha)$.

(iii) $P(\mathcal{E}_n^*(\alpha)) \geq 1 - 1/n^2$.

In the sequel we assume (2.9) is satisfied with the above c. Then, by a similar argument to the proof of (5.17), we obtain

$$\left|P^*(T_n^* \geq \hat{c}_{1-\alpha}^*) - P^*\left(T_n^* \geq \hat{c}_{1-\alpha} - J_n^*(\alpha)\right)\right| \leq C_{\lambda,\varepsilon,b_w,b_v}\frac{\varsigma_d^3}{\sigma_*^3}\frac{\log^3(dn)}{n}\log n + P^*(\mathcal{E}_n^*(\alpha)^c),$$

where $J_n^*(\alpha) := \langle n^{-1}\sum_{i=1}^n(w_i^2-1)X_i^{\otimes 2}, \widetilde{\Psi}_\alpha\rangle$ and $\widetilde{\Psi}_\alpha$ is defined as in (5.19). Since $P(P^*(\mathcal{E}_n^*(\alpha)^c) \geq 1/n) \leq n\,\mathrm{E}[P^*(\mathcal{E}_n^*(\alpha)^c)] \leq 1/n$ by Markov's inequality and (iii), we conclude

$$\left|P^*(T_n^* \geq \hat{c}_{1-\alpha}^*) - P^*\left(T_n^* \geq \hat{c}_{1-\alpha} - J_n^*(\alpha)\right)\right| \leq C_{\lambda,\varepsilon,b_w,b_v}\frac{\varsigma_d^3}{\sigma_*^3}\frac{\log^3(dn)}{n}\log n$$

with probability at least $1-1/n$. As in the proof of Theorem 2.3, we derive an Edgeworth expansion for $T_n^* + J_n^*(\alpha)$ by applying Theorem 4.1 to $\xi_i := (w_i\widetilde{X}_i + (w_i^2-1)U_i\mathbf{1}_d)/\sqrt{n}$ conditional on the data, where $\widetilde{X}_i := X_i - \bar{X}$ and $U_i := \langle X_i^{\otimes 2}, \widetilde{\Psi}_\alpha\rangle/\sqrt{n}$. By Lemma 5.4, conditional on the data, ξ_i has an approximate Stein kernel (τ_i, β_i) such that

$$\begin{aligned}\tau_i(\xi_i) &= n^{-1}\,\mathrm{E}^*[\tau^*(w_i)\widetilde{X}_i^{\otimes 2} + V_i \mid \xi_i],\\ \beta_i(\xi_i) &= n^{-1}\,\mathrm{E}^*[\langle X_i^{\otimes 2}, \widetilde{\Psi}_\alpha\rangle(2\tau^*(w_i) - w_i^2 - 1)\mathbf{1}_d \mid \xi_i],\end{aligned}$$

where $V_i := 4n^{-1/2}(\tau^*(w_i)\langle X_i^{\otimes 2}, \widetilde{\Psi}_\alpha\rangle w_i)\widetilde{X}_i \otimes \mathbf{1}_d$. As in the proof of Theorem 2.3, it is not difficult to check that we have the estimates corresponding to (4.34)–(4.37) in the present setting with probability at least $1-1/(2n)$ by a similar argument to the proof of Theorem 2.2. Meanwhile, by the Schwarz inequality,

$$\left\|\frac{1}{n}\sum_{i=1}^n V_i\right\|_\infty \leq \frac{4b_w^3}{\sqrt{n}}\sqrt{\frac{1}{n}\sum_{i=1}^n\langle X_i^{\otimes 2}, \widetilde{\Psi}_\alpha\rangle^2}\sqrt{\max_{1\leq j\leq d}\frac{1}{n}\sum_{i=1}^n\widetilde{X}_{ij}^2}.$$

Observe that $\langle X_i^{\otimes 2}, \widetilde{\Psi}_\alpha\rangle^2 = \langle X_i^{\otimes 4}, \widetilde{\Psi}_\alpha^{\otimes 2}\rangle$. Hence, by (5.20) and Lemma D.11,

$$\frac{1}{n}\sum_{i=1}^n\langle X_i^{\otimes 2}, \widetilde{\Psi}_\alpha\rangle^2 \leq C_\varepsilon\frac{b^4\varsigma_d^2\log d}{\sigma_*^4} \leq C_{\lambda,\varepsilon}\frac{b^2\varsigma_d^2\log d}{\sigma_*^2} \tag{5.36}$$

with probability at least $1-1/(2n)$. Combining these estimates with (4.26) and the argument to prove (4.32), we obtain the estimate corresponding to (4.32) with $b_w b$ replaced by $C_{\lambda,\varepsilon,b_w,b_v}b\sqrt{\varsigma_d/\sigma_*}$ with probability at least $1-1/n$. Moreover, by (4.22), (4.25), (5.20) and (5.36), we have

$$\frac{\log^2 d}{\sigma_*^4}\left(\mathrm{E}^*\left\|\sum_{i=1}^n\xi_i^{\otimes 3}\otimes\beta_i(\xi_i)\right\|_\infty + \mathrm{E}^*\left\|\sum_{i=1}^n\beta_i(\xi_i)\otimes\xi_i\otimes\tau_i(\xi_i)\right\|_\infty\right) \leq C_{\lambda,\varepsilon,b_w,b_v}\frac{\varsigma_d}{\sigma_*}\frac{\log^3(dn)}{n},$$

$$\frac{\log^{3/2} d}{\sigma_*^3} \mathrm{E}^* \left\| \sum_{i=1}^n \beta_i(\xi_i)^{\otimes 2} \otimes \xi_i \right\|_\infty \leq C_{\lambda,\varepsilon,b_w,b_v} \frac{\varsigma_d^2}{\sigma_*^2} \frac{\log^3(dn)}{n},$$

$$\frac{\log^2 d}{\sigma_*^4} \mathrm{E}^* \left\| \sum_{i=1}^n \beta_i(\xi_i) \otimes \xi_i \right\|_\infty^2 \leq C_{\lambda,\varepsilon,b_w,b_v} \frac{\varsigma_d^2}{\sigma_*^2} \frac{\log^3(dn)}{n},$$

$$\frac{\log d}{\sigma_*^2} \left(\mathrm{E}^* \left\| \sum_{i=1}^n \beta_i(\xi_i) \right\|_\infty^2 + \mathrm{E}^* \left\| \sum_{i=1}^n \beta_i(\xi_i)^{\otimes 2} \right\|_\infty \right) \leq C_{\lambda,\varepsilon,b_w} \frac{\log^3 d}{n}$$

with probability at least $1-1/n$. All together, we can proceed as in the proof of Theorem 2.2 and then obtain

$$\sup_{t\in\mathbb{R}} \left| P^*(T_n^* + J_n^*(\alpha) \leq t) - \int_{A(t)} (\hat{p}_{n,1}(z) + \hat{q}_n(z))\, dz \right| \leq C_{\lambda,\varepsilon,b_w,b_v} \frac{\varsigma_d^2}{\sigma_*^2} \frac{\log^3(dn)}{n} \log n$$

with probability at least $1 - 3/n$, where

$$\hat{q}_n(z) = \frac{1}{2n} \sum_{i=1}^n \langle 2\widetilde{X}_i U_i \mathbf{1}_d^\top + U_i^2 \mathbf{1}_d^{\otimes 2}, \nabla^2 \phi_\Sigma(z) \rangle - \frac{1}{6n^{3/2}} \sum_{i=1}^n \langle \mathrm{E}^*[(w_i \widetilde{X}_i + U_i \mathbf{1}_d)^{\otimes 3}] - \widetilde{X}_i^{\otimes 3}, \nabla^3 \phi_\Sigma(z) \rangle.$$

The remaining proof is a minor modification of the proof of (5.26), so we omit the details. $\square$

Lemma 5.6. *Under the assumptions of Theorem 2.4, there exists a constant $C > 0$ depending only on λ and ε such that*

$$|\mathrm{E}[R_n(\alpha+\delta)] - \mathrm{E}[R_n(\alpha)]| \leq C\delta \frac{\varsigma_d^3}{\sigma_*^3} \sqrt{\frac{\log^3 d}{n}} \tag{5.37}$$

for any $\alpha \in (\varepsilon, 1-2\varepsilon)$ and $\delta \in (0, \varepsilon]$.

Proof. By Lemma D.4 and (5.5),

$$|\mathrm{E}[R_n(\alpha+\delta)] - \mathrm{E}[R_n(\alpha)]| \leq C_{\lambda,\varepsilon} \frac{b^3}{\sqrt{n}} \left(\frac{\log^2 d}{\sigma_*^4} \left| \frac{1}{f_\Sigma(c_{1-\alpha-\delta}^G)} - \frac{1}{f_\Sigma(c_{1-\alpha}^G)} \right| + \frac{\log d}{\sigma_*^2} \frac{\varsigma_d}{\sqrt{\log d}} |\Psi_{\alpha+\delta} - \Psi_\alpha| \right).$$

Noting that $(F_Z^{-1})'(p) = 1/f_\Sigma(c_p^G)$ for all $p \in (0,1)$, we obtain by the mean value theorem and (5.6)

$$\left| \frac{1}{f_\Sigma(c_{1-\alpha-\delta}^G)} - \frac{1}{f_\Sigma(c_{1-\alpha}^G)} \right| \leq C_\varepsilon \delta \frac{\varsigma_d^3}{\sigma_*^2 \sqrt{\log d}}.$$

Also, by Lemma 4.2, the mean value theorem and (5.5),

$$|\Psi_{\alpha+\delta} - \Psi_\alpha| \lesssim \frac{\log^{3/2} d}{\sigma_*^3} |c_{1-\alpha-\delta}^G - c_{1-\alpha}^G| \leq C_\varepsilon \delta \frac{\varsigma_d \log d}{\sigma_*^3}.$$

Combining these bounds gives (5.37). $\square$

Proof of Theorem 2.4. Denote by c_1 and C_1 the constants c and C in Theorem 2.3, respectively. Also, denote by c_2 and C_2 the constants c and C in Lemma 5.5, respectively. Since the left hand side of (2.12) is bounded by 1, we may assume (2.9) holds with $c = c_1 \wedge c_2$ without loss of generality. Then, for each $\alpha \in (\varepsilon, 1-\varepsilon)$, the event

$$\mathcal{E}_n(\alpha) := \left\{ \left| P^*(T_n^* \geq \hat{c}^*_{1-\alpha}) - (\alpha - R_n(\alpha)) \right| \leq C_2 \frac{\varsigma_d^3}{\sigma_*^3} \frac{\log^3(dn)}{n} \log n \right\}$$

occurs with probability at least $1 - 5/n$. Meanwhile, by (2.9), (5.5) and Lemmas D.4 and D.10, there exists a constant $C_3 > 0$ depending only on λ, ε such that the event

$$\mathcal{E}_n := \left\{ \sup_{\varepsilon < \alpha < 1-\varepsilon} |R_n(\alpha) - \mathrm{E}[R_n(\alpha)]| \leq C_3 \frac{\varsigma_d}{\sigma_*} \frac{\log^3(dn)}{n} \right\}$$

occurs with probability at least $1 - 1/n$ and

$$\sup_{\varepsilon < \alpha < 1-\varepsilon} |\mathrm{E}[R_n(\alpha)]| \leq C_3 \frac{\varsigma_d}{\sigma_*} \sqrt{\frac{\log^3 d}{n}}. \tag{5.38}$$

Further, let C_4 be the constant C in Lemma 5.6. Set

$$\Delta_n := C_2 \frac{\varsigma_d^3}{\sigma_*^3} \frac{\log^3(dn)}{n} \log n + C_3 \frac{\varsigma_d}{\sigma_*} \frac{\log^3(dn)}{n} + C_3 C_4 \frac{\varsigma_d^4}{\sigma_*^4} \frac{\log^3 d}{n}.$$

Since the left hand side of (2.12) is bounded by 1, we may assume without loss of generality

$$C_3 \frac{\varsigma_d}{\sigma_*} \sqrt{\frac{\log^3 d}{n}} + 3\Delta_n \leq \varepsilon. \tag{5.39}$$

Now fix $\alpha \in (2\varepsilon, 1-2\varepsilon)$ arbitrarily. Set $\alpha_{n,1} := \alpha - \Delta_n$ and $\alpha'_{n,1} := \alpha_{n,1} + \mathrm{E}[R_n(\alpha_{n,1})]$. By (5.38) and (5.39), $\alpha_{n,1}, \alpha'_{n,1} \in (\varepsilon, 1-\varepsilon)$. Hence, on $\mathcal{E}_n(\alpha'_{n,1}) \cap \mathcal{E}_n$,

$$\begin{aligned} P^*(\hat{F}_n^*(T_n^*) > 1 - \alpha'_{n,1}) &\leq P^*(T_n^* \geq \hat{c}^*_{1-\alpha'_{n,1}}) \leq \alpha'_{n,1} - R_n(\alpha'_{n,1}) + C_2 \frac{\varsigma_d^3}{\sigma_*^3} \frac{\log^3(dn)}{n} \log n \\ &\leq \alpha_{n,1} + \mathrm{E}[R_n(\alpha_{n,1})] - \mathrm{E}[R_n(\alpha'_{n,1})] + C_3 \frac{\varsigma_d}{\sigma_*} \frac{\log^3(dn)}{n} + C_2 \frac{\varsigma_d^3}{\sigma_*^3} \frac{\log^3(dn)}{n} \log n \leq \alpha, \end{aligned}$$

where the last inequality follows from (5.37) and (5.38). This yields $\hat{\beta}_\alpha \leq 1 - \alpha'_{n,1}$ on $\mathcal{E}_n(\alpha'_{n,1}) \cap \mathcal{E}_n$. Hence

$$\begin{aligned} P(T_n \geq \hat{c}_{\hat{\beta}_\alpha}) &\geq P(T_n \geq \hat{c}_{1-\alpha'_{n,1}}) - \frac{6}{n} \geq \alpha'_{n,1} - \mathrm{E}[R_n(\alpha'_{n,1})] - C_{\lambda,\varepsilon,b_w} \frac{\varsigma_d^3}{\sigma_*^3} \frac{\log^3(dn)}{n} \log n \\ &\geq \alpha + \mathrm{E}[R_n(\alpha_{n,1})] - \mathrm{E}[R_n(\alpha'_{n,1})] - C_{\lambda,\varepsilon,b_w,b_v} \frac{\varsigma_d^4}{\sigma_*^4} \frac{\log^3(dn)}{n} \log n \\ &\geq \alpha - C_{\lambda,\varepsilon,b_w,b_v} \frac{\varsigma_d^4}{\sigma_*^4} \frac{\log^3(dn)}{n} \log n, \end{aligned} \tag{5.40}$$

where the second inequality is by Theorem 2.3, the third by $\Delta_n \leq C_{\lambda,\varepsilon,b_w,b_v} \frac{\varsigma_d^4}{\sigma_*^4} \frac{\log^3 d}{n} \log n$ and the fourth by (5.37) and (5.38). Similarly, with $\alpha_{n,2} := \alpha + 2\Delta_n$ and $\alpha'_{n,2} := \alpha_{n,2} + \mathrm{E}[R_n(\alpha_{n,2})]$, we have on $\mathcal{E}_n(\alpha'_{n,2}) \cap \mathcal{E}_n$

$$P^*(\hat{F}_n^*(T_n^*) > 1 - \alpha'_{n,2} - \Delta_n) \geq P^*(\hat{F}_n^*(T_n^*) \geq 1 - \alpha'_{n,2}) = P^*(T_n^* \geq \hat{c}^*_{1-\alpha'_{n,2}})$$

$$\geq \alpha'_{n,2} - R_n(\alpha'_{n,2}) - C_2 \frac{\varsigma_d^3}{\sigma_*^3} \frac{\log^3(dn)}{n} \log n > \alpha.$$

Hence $\hat{\beta}_\alpha > 1 - \alpha'_{n,2} - \Delta_n$ on $\mathcal{E}_n(\alpha'_{n,2}) \cap \mathcal{E}_n$. Therefore, a similar argument to (5.40) yields

$$P(T_n \geq \hat{c}_{\hat{\beta}_\alpha}) \leq \alpha + C_{\lambda,\varepsilon,b_w,b_v} \frac{\varsigma_d^4}{\sigma_*^4} \frac{\log^3(dn)}{n} \log n.$$

Combining this and (5.40) gives the desired result. □

Appendix

A Nearly optimal high-dimensional CLT under the sub-exponential condition

Theorem A.1. *Set $\sigma_j := \sqrt{\mathrm{Var}[S_{n,j}]}$ for $j = 1, \dots, d$. Suppose that there exists a constant $B \geq 1$ such that $\max_{i,j} \|X_{ij}/\sigma_j\|_{\psi_1} \leq B$ and $\max_j n^{-1} \sum_{i=1}^n \mathrm{E}[(X_{ij}/\sigma_j)^4] \leq B^2$. Then there exists a universal constant $C > 0$ such that*

$$\sup_{A \in \mathcal{R}} |P(S_n \in A) - P(Z \in A)| \leq \frac{C}{\rho_*^2} \sqrt{\frac{B^2 \log^3(dn)}{n}} \log n, \tag{A.1}$$

where $Z \sim N(0, \mathrm{Cov}[S_n])$ and ρ_*^2 is the minimum eigenvalue of the correlation matrix of S_n.

Proof. As announced, the proof is a combination of [24, Theorem 2.1] and a simple truncation argument used in [39, Section 5.2] and [32, Section 4.3]. Denote by δ the left hand side of (A.1). Considering X_{ij}/σ_j instead of X_{ij}, we may assume $\sigma_j = 1$ for all j without loss of generality. Further, since $\delta \leq 1$, we may also assume

$$\frac{1}{\rho_*^2} \sqrt{\frac{B^2 \log^3(dn)}{n}} \log n \leq 1. \tag{A.2}$$

Next, let $\kappa_n := 2B \log n$. For $i = 1, \dots, n$ and $j = 1, \dots, d$, define $\hat{X}_{ij} := X_{ij} 1_{\{|X_{ij}| \leq \kappa_n\}} - \mathrm{E}[X_{ij} 1_{\{|X_{ij}| \leq \kappa_n\}}]$ and set $\hat{X}_i = (\hat{X}_{i1}, \dots, \hat{X}_{id})^\top$ and $\hat{S}_n := n^{-1/2} \sum_{i=1}^n \hat{X}_i$. Note that $\max_{1 \leq i \leq n} \|\hat{X}_i\|_\infty \leq 2\kappa_n$. Then, by a similar argument to the proof of Eq.(4.19) in [32], we obtain

$$\delta \lesssim \frac{1}{n} + \frac{B \log(dn) \sqrt{\log d}}{\sqrt{n}} \log n + \hat{\delta},$$

where $\hat{\delta} := \sup_{A \in \mathcal{R}} |P(\hat{S}_n \in A) - P(Z \in A)|$. Since $\rho_*^2 \leq \sigma_1^2 = 1$, it remains to prove

$$\hat{\delta} \lesssim \frac{1}{\rho_*^2} \sqrt{\frac{B^2 \log^3(dn)}{n}} \log n. \tag{A.3}$$

We prove this bound by applying Theorem 2.1 in [24] with $\psi = 2\kappa_n$. This gives

$$\hat{\delta} \lesssim (\log n) \left(\Delta_0 + \sqrt{\frac{B^2 \log^3 d}{n \rho_*^4}} + \frac{\kappa_n^2 \log^2 d}{n \rho_*^2} \right) + \frac{\kappa_n \log^{3/2} d}{\rho_* \sqrt{n}}, \tag{A.4}$$

where $\Delta_0 := \frac{\log d}{\rho_*^2} \| \mathrm{Cov}(\hat{S}_n) - \mathrm{Cov}(S_n) \|_\infty$. By (A.2) and $\rho_* \leq 1$,

$$\frac{\kappa_n^2 \log^2 d}{n \rho_*^2} = \frac{4}{\rho_*^2} \sqrt{\frac{B^2 \log^3 d}{n}} \sqrt{\frac{B^2 (\log d)(\log^2 n)}{n}} \log n \leq \frac{4}{\rho_*^2} \sqrt{\frac{B^2 \log^3(dn)}{n}}$$

and

$$\frac{\kappa_n \log^{3/2} d}{\rho_* \sqrt{n}} \leq \frac{2}{\rho_*^2} \sqrt{\frac{B^2 \log^3(dn)}{n}} \log n.$$

Further, by Eq.(26) in [39] and (A.2),

$$\Delta_0 \lesssim \frac{\log d}{\rho_*^2} e^{-\kappa_n/(2B)} B^2 \log n = \frac{1}{\rho_*^2} \frac{B^2 \log d}{n} \log n \leq \frac{1}{\rho_*^2} \sqrt{\frac{B^2 \log^3(dn)}{n}}.$$

Consequently, we obtain (A.3) from (A.4). □

B Proofs of the auxiliary results in Section 4.1

B.1 Proof of Lemma 4.2

We divide the proof into four steps.

Step 1. First we reduce the proof to the case $\Sigma = I_d$. Let $Z \sim N(0, \Sigma)$ and $Z' \sim N(0, \Sigma - \sigma_*^2 I_d)$. Then, for any $A \in \mathcal{R}$, $u, v \in \mathbb{R}_+^d$ and $x \in \mathbb{R}^d$, we have

$$\int_{A^{u,v} \setminus A} \phi_\Sigma(x + z) dz = \mathrm{E}[1_{A^{u,v} \setminus A}(Z - x)] = \mathrm{E}\left[\int_{\mathbb{R}^d} 1_{A^{u,v} \setminus A}(\sigma_* z + Z') \phi_d(z + x/\sigma_*) dz\right].$$

Differentiating both sides r times with respect to x and then setting $x = 0$, we obtain

$$\int_{A^{u,v} \setminus A} \nabla^r \phi_\Sigma(z) dz = \frac{1}{\sigma_*^r} \mathrm{E}\left[\int_{\mathbb{R}^d} 1_{A^{u,v} \setminus A}(\sigma_* z + Z') \phi_d(z) dz\right].$$

Observe that $1_{A^{u,v} \setminus A}(\sigma_* z + Z') = 1_{(\sigma_*^{-1}(A - Z'))^{u/\sigma_*, v/\sigma_*} \setminus (\sigma_*^{-1}(A - Z'))}(z)$ and $\sigma_*^{-1}(A - Z') \in \mathcal{R}$. Hence

$$\left\| \int_{A^{u,v} \setminus A} \nabla^r \phi_\Sigma(z) dz \right\|_1 \leq \frac{1}{\sigma_*^r} \sup_{A \in \mathcal{R}} \left\| \int_{A^{u/\sigma_*, v/\sigma_*} \setminus A} \nabla^r \phi_d(z) dz \right\|_1.$$

Therefore, the claim for general Σ follows from that for $\Sigma = I_d$.

Step 2. In this and the next steps, we show that the quantity inside $\sup_{A \in \mathcal{R}}$ on the left hand side of (4.3) can be replaced by a weighted surface integral of $\nabla^r \phi_\Sigma$ over the boundary of A. Note that an analogous result for the case $r = 0$ is standard in the literature; see e.g. Proposition 1.1 in [57]. For $A \in \mathcal{R}$, $u, v \in \mathbb{R}_+^d$ and $\varepsilon > 0$, set

$$I_A(u, v) = \left\| \int_{A^{u,v} \setminus A} \nabla^r \phi_d(z) dz \right\|_1, \qquad K(\varepsilon) = \sup_{A \in \mathcal{R}; u, v \in [0, \varepsilon]^d} \frac{I_A(u, v)}{\varepsilon}.$$

In this step, we prove

$$\sup_{\varepsilon > 0} K(\varepsilon) = \limsup_{\varepsilon \downarrow 0} K(\varepsilon).$$

Take $\varepsilon > 0$ arbitrarily. For any $A \in \mathcal{R}$ and $u, v \in \mathbb{R}_+^d$, observe that $A^{u/2, v/2} \in \mathcal{R}$ and $A^{u,v} \setminus A$ is the disjoint union of $(A^{u/2, v/2})^{u/2, v/2} \setminus (A^{u/2, v/2})$ and $A^{u/2, v/2} \setminus A$. This implies that

$$\sup_{A \in \mathcal{R}; u, v \in [0, \varepsilon]^d} I_A(u, v) \leq 2 \sup_{A \in \mathcal{R}; u, v \in [0, \varepsilon/2]^d} I_A(u, v),$$

and thus $K(\varepsilon) \le K(\varepsilon/2)$. Repeating this procedure gives $K(\varepsilon) \le K(\varepsilon/2^n)$ for $n = 1, 2, \ldots$. Hence

$$K(\varepsilon) \le \limsup_{n\to\infty} K(\varepsilon/2^n) \le \limsup_{\eta\downarrow 0} K(\eta).$$

Since ε is arbitrary, we obtain the desired result.

Step 3. For any Borel set $A \subset \mathbb{R}^{d-1}$, $j \in \{1, \ldots, d\}$ and $s \in \mathbb{R}$, define

$$J_{A,j}(s) = \int_A \nabla^r \phi_d(z|_{z_j=s}) dz_1 \cdots \widehat{dz_j} \cdots dz_d,$$

where $z|_{z_j=s} = (z_1, \ldots, z_{j-1}, s, z_{j+1}, \ldots, z_d)^\top$ and $\widehat{dz_j}$ means that dz_j is omitted. Then, for $u, v \in \mathbb{R}^d$ and $A = \prod_{j=1}^d [a_j, b_j] \in \mathcal{R}$, set

$$L_A(u, v) = \sum_{j=1}^d \{u_j J_{A^j,j}(a_j) + v_j J_{A^j,j}(b_j)\},$$

where $A^j = \prod_{k:k\neq j} [a_k, b_k]$. In this step, we prove

$$\limsup_{\varepsilon\downarrow 0} K(\varepsilon) = \limsup_{\varepsilon\downarrow 0} \sup_{A\in\mathcal{R}; u,v\in[0,\varepsilon]^d} \frac{\|L_A(u, v)\|_1}{\varepsilon}.$$

Fix $A = \prod_{j=1}^d [a_j, b_j] \in \mathcal{R}$, $\varepsilon > 0$ and $u, v \in [0, \varepsilon]^d$. For $j_1, \ldots, j_r \in \{1, \ldots, d\}$, we set

$$A^{u,v}_{j_1,\ldots,j_r} = \{x \in A^{u,v} : x_j \notin [a_j, b_j] \text{ for } j \in \{j_1, \ldots, j_r\} \text{ and } x_j \in [a_j, b_j] \text{ for } j \notin \{j_1, \ldots, j_r\}\}.$$

Then, $A^{u,v} \setminus A = \bigcup_{r=1}^d \bigcup_{1\le j_1<\cdots<j_r\le d} A^{u,v}_{j_1,\ldots,j_r}$ and this is a disjoint union. Hence we have

$$I_A(u, v) = \left\| \sum_{r=1}^d \sum_{1\le j_1<\cdots<j_r\le d} \int_{A^{u,v}_{j_1,\ldots,j_r}} \nabla^r \phi_d(z) dz \right\|_1.$$

One can easily check that

$$\sup_{A\in\mathcal{R}; u,v\in[0,\varepsilon]^d} \sum_{r=2}^d \sum_{1\le j_1<\cdots<j_r\le d} \int_{A^{u,v}_{j_1,\ldots,j_r}} \|\nabla^r \phi_d(z)\|_1 \, dz = O(\varepsilon^2) \quad \text{as } \varepsilon \downarrow 0.$$

Hence we obtain

$$\limsup_{\varepsilon\downarrow 0} K(\varepsilon) = \limsup_{\varepsilon\downarrow 0} \sup_{A\in\mathcal{R}; u,v\in[0,\varepsilon]^d} \left\| \frac{1}{\varepsilon} \sum_{j=1}^d \int_{A^{u,v}_j} \nabla^r \phi_d(z) dz \right\|_1.$$

For any $j \in \{1, \ldots, d\}$, observe that $A^{u,v}_j$ is the disjoint union of

$$\underline{A}^u_j := [a_1, b_1] \times \cdots [a_{j-1}, b_{j-1}] \times [a_j - u_j, a_j) \times [a_{j+1}, b_{j+1}] \times \cdots \times [a_d, b_d]$$

and

$$\overline{A}^v_j := [a_1, b_1] \times \cdots [a_{j-1}, b_{j-1}] \times (b_j, b_j + v_j] \times [a_{j+1}, b_{j+1}] \times \cdots \times [a_d, b_d].$$

Therefore, we have

$$\begin{aligned}
&\left\|\sum_{j=1}^d \int_{A_j^{u,v}} \nabla^r \phi_d(z)dz - L_A(u,v)\right\|_1 \\
&= \left\|\sum_{j=1}^d \left(\int_{\underline{A}_j^u} \{\nabla^r \phi_d(z) - \nabla^r \phi_d(z|_{z_j=a_j})\}dz + \int_{\overline{A}_j^v} \{\nabla^r \phi_d(z) - \nabla^r \phi_d(z|_{z_j=b_j})\}dz\right)\right\|_1 \\
&\leq 2\varepsilon \sum_{j=1}^d \int_{\mathbb{R}^{d-1}} \sup_{s,t\in\mathbb{R}:|s-t|\leq\varepsilon} \left\|\nabla^r \phi_d(z|_{z_j=s}) - \nabla^r \phi_d(z|_{z_j=t})\right\|_1 dz_1 \cdots \widehat{dz_j} \cdots dz_d.
\end{aligned}$$

Thus,

$$\limsup_{\varepsilon\downarrow 0} \sup_{A\in\mathcal{R};u,v\in[0,\varepsilon]^d} \frac{1}{\varepsilon}\left\|\sum_{j=1}^d \int_{A_j^{u,v}} \nabla^r \phi_d(z)dz - L_A(u,v)\right\|_1 = 0.$$

This gives the desired result.

Step 4. It remains to prove

$$\limsup_{\varepsilon\downarrow 0} \sup_{A\in\mathcal{R};u,v\in R_+(\varepsilon)} \frac{\|L_A(u,v)\|_1}{\varepsilon} \leq C_r(\log d)^{(r+1)/2}. \tag{B.1}$$

We first note that if A is an orthant, i.e. $a_j = -\infty$ for all j, then (B.1) immediately follows from the fundamental theorem of calculus and Lemma D.4. In fact, we have in this case

$$\frac{\|L_A(u,v)\|_1}{\varepsilon} \leq \left\|\int_A \nabla^{r+1}\phi_d(z)dz\right\|_1$$

for any $\varepsilon > 0$ and $u, v \in [0,\varepsilon]^d$. In the following we show that the proof is essentially reduced to this case by a similar argument to the proof of [29, Lemma 2.2]. For every $q \in \{1,\dots,r\}$, set

$$\mathcal{N}_q(r) = \{(\nu_1,\dots,\nu_q) \in \mathbb{Z}^q : \nu_1,\dots,\nu_q \geq 0, \nu_1 + \cdots + \nu_q = r\}.$$

Also, for any $m \in \mathbb{N}$, let

$$\mathcal{J}_m(d) = \{(j_1,\dots,j_m) \in \{1,\dots,d\}^m : j_1,\dots,j_m \text{ are distinct}\}.$$

Then, for any $A \in \mathcal{R}$, $j \in \{1,\dots,d\}$ and $s \in \mathbb{R}$, we have

$$\begin{aligned}
\left\|J_{A^j,j}(s)\right\|_1 &= \sum_{j_1,\dots,j_r=1}^d \left|\int_{A^j} \partial_{j_1,\dots,j_r}\phi_d(z|_{z_j=s})dz_1 \cdots \widehat{dz_j} \cdots dz_d\right| \\
&\leq C_r \sum_{q=1}^r \sum_{(\nu_1,\dots,\nu_q)\in\mathcal{N}_q(r)} \sum_{(j_1,\dots,j_q)\in\mathcal{J}_q(d)} \left|\int_{A^j} \partial_{j_1}^{\nu_1}\cdots\partial_{j_q}^{\nu_q}\phi_d(z|_{z_j=s})dz_1 \cdots \widehat{dz_j} \cdots dz_d\right| \\
&=: C_r \sum_{q=1}^r \sum_{\nu=(\nu_1,\dots,\nu_q)\in\mathcal{N}_q(r)} \sum_{\boldsymbol{j}=(j_1,\dots,j_q)\in\mathcal{J}_q(d)} \Lambda_j(\nu,\boldsymbol{j}).
\end{aligned}$$

For each $r = 1, \dots, q$, the cardinality of the set $\mathcal{N}_q(r)$ is bounded by a constant depending only on r. Therefore, to prove (B.1), it suffices to show that

$$\sum_{j=1}^{d} \sum_{\boldsymbol{j}=(j_1,\dots,j_r)\in\mathcal{J}_q(d)} \Lambda_j(\nu,\boldsymbol{j}) \le C_r(\log d)^{(r+1)/2} \tag{B.2}$$

for any (fixed) $q \in \{1,\dots,r\}$, $\nu = (\nu_1,\dots,\nu_q) \in \mathcal{N}_q(r)$, $A = \prod_{j=1}^d [a_j, b_j] \in \mathcal{R}$ and $s \in \{a_j, b_j\}$, $j = 1,\dots,d$.

To prove (B.2), we introduce additional notation. For a non-negative integer m, H_m denotes the m-th Hermite polynomial, i.e. $H_m(t) = (-1)^m\phi(t)^{-1}\phi^{(m)}(t)$. When $m \ge 1$, we set $h_m(t) = H_{m-1}(t)\phi(t)$. Also, we denote by t_m the maximum root of H_m. For example, $t_1 = 0, t_2 = 1, t_3 = \sqrt{3}$. Finally, set $M_m := \max_{0\le t\le t_m} |H_{m-1}(t)| < \infty$ and define

$$\tilde{h}_m(t) = M_m\phi(t)1_{[0,t_m]}(t) + h_m(t)1_{(t_m,\infty)}(t).$$

The function $\tilde{h}_m$ satisfies the following properties by Lemma A.1 in [29]:

$$\tilde{h}_m \text{ is decreasing on } [0,\infty). \tag{B.3}$$

$$|h_m(t)| \le \tilde{h}_m(|t|) \text{ for all } t \in \mathbb{R}. \tag{B.4}$$

Now, we fix $\boldsymbol{j} = (j_1,\dots,j_q) \in \mathcal{J}_q(d)$ and $j \in \{1,\dots,d\}$ for a while. Set

$$\nu = \begin{cases} \nu_p & \text{if } j = j_p \text{ for some } p \in \{1,\dots,q\}, \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

$$\begin{aligned} \Lambda_j(\nu,\boldsymbol{j}) &= |h_{\nu+1}(s)| \left(\prod_{p:j_p\ne j} \left| h_{\nu_p}(b_{j_p}) - h_{\nu_p}(a_{j_p}) \right| \right) \prod_{k:k\ne j_1,\dots,j_q,j} \{\Phi(b_k) - \Phi(a_k)\} \\ &\le \tilde{h}_{\nu+1}(|s|) \left(\prod_{p:j_p\ne j} \left(\tilde{h}_{\nu_p}(|b_{j_p}|) + \tilde{h}_{\nu_p}(|a_{j_p}|) \right) \right) \prod_{k:k\ne j_1,\dots,j_q,j} \{\Phi(b_k) + \Phi(-a_k) - 1\}, \end{aligned}$$

where the last inequality follows from (B.4) and the identity $1 - \Phi(t) = \Phi(-t)$. Set $c_k = |a_k| \wedge |b_k|$ for $k = 1,\dots,d$. Then we have $\Phi(b_k) + \Phi(-a_k) - 1 \le \min\{\Phi(b_k), \Phi(-a_k)\} \le \Phi(c_k)$. Combining this with (B.3) gives

$$\Lambda_j(\nu,\boldsymbol{j}) \le 2^q \tilde{h}_{\nu+1}(c_j) \left(\prod_{p:j_p\ne j} \tilde{h}_{\nu_p}(c_{j_p}) \right) \prod_{k:k\ne j_1,\dots,j_q,j} \Phi(c_k).$$

Now, observe that $\tilde{h}_m(t) \le C_m(1 + t^{m-1})\phi(t)$ for any $m \in \mathbb{N}$ and $t \ge 0$ by construction. Hence, if $\max_{p=1,\dots,q} c_{j_p} \le \sqrt{4(r+1)\log d}$,

$$\Lambda_j(\nu,\boldsymbol{j}) \le \begin{cases} C_r(\log d)^{(r-q+1)/2} \left(\prod_{p=1}^q \phi(c_{j_p})\right) \prod_{k:k\ne j_1,\dots,j_q} \Phi(c_k) & \text{if } j \in \{j_1,\dots,j_q\}, \\ C_r(\log d)^{(r-q)/2}\phi(c_j) \left(\prod_{p=1}^q \phi(c_{j_p})\right) \prod_{k:k\ne j_1,\dots,j_q,j} \Phi(c_k) & \text{otherwise,} \end{cases}$$

where we used the identity $\sum_{p=1}^q \nu_p = r$. Besides, if $\max_{p=1,\dots,q} c_{j_p} > \sqrt{4(r+1)\log d}$,

$$\Lambda_j(\nu, \boldsymbol{j}) \le C_r \prod_{p=1}^q e^{-c_{j_p}^2/4} \le C_r d^{-r-1}.$$

Consequently,

$$\begin{aligned}
&\sum_{j=1}^d \sum_{\boldsymbol{j}=(j_1,\dots,j_q)\in\mathcal{J}_q(d)} \Lambda_j(\nu,\boldsymbol{j}) \\
&\le C_r + C_r \sum_{(j_1,\dots,j_q)\in\mathcal{J}_q(d)} (\log d)^{(r-q+1)/2} \left(\prod_{p=1}^q \phi(c_{j_p})\right) \prod_{k:k\neq j_1,\dots,j_q} \Phi(c_k) \\
&\quad + C_r \sum_{(j_1,\dots,j_q)\in\mathcal{J}_q(d)} \sum_{j:j\neq j_1,\dots,j_q} (\log d)^{(r-q)/2}\phi(c_j) \left(\prod_{p=1}^q \phi(c_{j_p})\right) \prod_{k:k\neq j_1,\dots,j_q,j} \Phi(c_k).
\end{aligned}$$

With $A' = \prod_{j=1}^d (-\infty, c_j]$, we can rewrite the right hand side of the above inequality as

$$\begin{aligned}
C_r + C_r(\log d)^{(r-q+1)/2} &\sum_{(j_1,\dots,j_q)\in\mathcal{J}_q(d)} \int_{A'} \partial_{j_1,\dots,j_q}\phi_d(z)dz \\
&+ C_r(\log d)^{(r-q)/2} \sum_{(j_1,\dots,j_{q+1})\in\mathcal{J}_{q+1}(d)} \int_{A'} \partial_{j_1,\dots,j_{q+1}}\phi_d(z)dz.
\end{aligned}$$

This quantity is bounded by

$$C_r\left(1 + (\log d)^{(r-q+1)/2}\left\|\int_{A'} \nabla^q\phi_d(z)dz\right\|_1 + (\log d)^{(r-q)/2}\left\|\int_{A'} \nabla^{q+1}\phi_d(z)dz\right\|_1\right). \tag{B.5}$$

For any $m \in \mathbb{N}$, observe that

$$\left\|\int_{A'} \nabla^m\phi_d(z)dz\right\|_1 = \lim_{a\to-\infty}\left\|\int_{\prod_{j=1}^d [a,c_j]} \nabla^m\phi_d(z)dz\right\|_1 \le \sup_{A\in\mathcal{R}}\left\|\int_A \nabla^m\phi_d(z)dz\right\|_1.$$

Therefore, by Lemma D.4, the quantity in (B.5) is bounded by $C_r(\log d)^{(r+1)/2}$. This gives (B.2). □

B.2 Proof of Lemma 4.3

The proof of (4.6) is an almost straightforward multi-dimensional extension of that of [33, Lemma 2.1], and the proof of (4.7) is its simplification. Let $Z \sim N(0,\Sigma)$ be independent of $(\xi_i)_{i=1}^n$. Set $W(s) = \sqrt{1-s}W + \sqrt{s}Z$ for every $s \in [0,1]$. The following two lemmas summarize the role of the approximate Stein kernel in the proof.

Lemma B.1. *Under the assumptions of Lemma 4.3, suppose additionally that $h \in C_b^r(\mathbb{R}^d)$. Then, for any $t \in [0,1]$, we have the following:*

(a) If $U \in (\mathbb{R}^d)^{\otimes r}$ is non-random,

$$\begin{aligned}
&\mathrm{E}[\langle U, \nabla^r h(W(t))\rangle] - \mathrm{E}[\langle U, \nabla^r h(Z)\rangle] \\
&= \frac{1}{2}\int_t^1 \left(\mathrm{E}[\langle U\otimes\bar{T}, \nabla^{r+2}h(W(s))\rangle] + \frac{\mathrm{E}[\langle U\otimes B, \nabla^{r+1}h(W(s))\rangle]}{\sqrt{1-s}}\right)ds.
\end{aligned}$$

(b) *For every $i = 1, \dots, n$, let $v_i : \mathbb{R}^d \to (\mathbb{R}^d)^{\otimes r}$ be a measurable function such that $v_i(\xi_i), v_i(\xi_i) \otimes \xi_i, v_i(\xi_i) \otimes \tau_i(\xi_i)$ and $v_i(\xi_i) \otimes \beta_i(\xi_i)$ are integrable. Set $V := \sum_{i=1}^n v_i(\xi_i)$. Then*

$$
\begin{aligned}
&\mathrm{E}[\langle V, \nabla^r h(W(t))\rangle] - \mathrm{E}[\langle V, \nabla^r h(Z)\rangle] \\
&= \frac{1}{2}\int_t^1 \left(\mathrm{E}[\langle V \otimes \bar{T}, \nabla^{r+2} h(W(s))\rangle] - \sum_{i=1}^n \mathrm{E}[\langle v_i(\xi_i) \otimes \tau_i(\xi_i), \nabla^{r+2} h(W(s))\rangle] \right) ds \\
&\quad + \frac{1}{2}\int_t^1 \frac{1}{\sqrt{1-s}} \left(\mathrm{E}[\langle V \otimes B, \nabla^{r+1} h(W(s))\rangle] + \sum_{i=1}^n \mathrm{E}[\langle v_i(\xi_i) \otimes (\xi_i - \beta_i(\xi_i)), \nabla^{r+1} h(W(s))\rangle] \right) ds.
\end{aligned}
$$

Proof. Since the proof of (a) is similar to that of (b) and simpler, we prove (b) only.

By the fundamental theorem of calculus, we obtain

$$
\begin{aligned}
&\mathrm{E}[\langle V, \nabla^r h(W(t))\rangle] - \mathrm{E}[\langle V, \nabla^r h(Z)\rangle] \\
&= -\int_t^1 \frac{\partial}{\partial s} \mathrm{E}[\langle V, \nabla^r h(W(s))]ds \\
&= \frac{1}{2}\int_t^1 \left(\frac{\mathrm{E}[\langle V \otimes W, \nabla^{r+1} h(W(s))\rangle]}{\sqrt{1-s}} - \frac{\mathrm{E}[\langle V \otimes Z, \nabla^{r+1} h(W(s))\rangle]}{\sqrt{s}} \right) ds. \qquad \text{(B.6)}
\end{aligned}
$$

Since Z is independent of (W, V), an application of the multivariate Stein identity conditional on (W, V) gives for any $0 < s < 1$

$$
\frac{\mathrm{E}[\langle V \otimes Z, \nabla^{r+1} h(W(s))\rangle]}{\sqrt{s}} = \mathrm{E}[\langle V \otimes \Sigma, \nabla^{r+2} h(W(s))\rangle]. \qquad \text{(B.7)}
$$

Meanwhile, we rewrite $\mathrm{E}[\langle V \otimes W, \nabla^3 h(W(s))\rangle]$ as

$$
\begin{aligned}
&\mathrm{E}[\langle V \otimes W, \nabla^{r+1} h(W(s))\rangle] \\
&= \sum_{i=1}^n \mathrm{E}[\langle V \otimes \xi_i, \nabla^{r+1} h(W(s))\rangle] \\
&= \sum_{i=1}^n \mathrm{E}[\langle V^{(i)} \otimes \xi_i, \nabla^{r+1} h(W(s))\rangle] + \sum_{i=1}^n \mathrm{E}[\langle v_i(\xi_i) \otimes \xi_i, \nabla^{r+1} h(W(s))\rangle], \qquad \text{(B.8)}
\end{aligned}
$$

where $V^{(i)} = V - v_i(\xi_i)$. For $i = 1, \dots, n$, since (τ_i, β_i) is an approximate Stein kernel for ξ_i and ξ_i is independent of $V^{(i)}$ and $W - \xi_i$, we have

$$
\begin{aligned}
&\mathrm{E}[\langle V^{(i)} \otimes \xi_i, \nabla^{r+1} h(W(s))\rangle] \\
&= \sqrt{1-s}\, \mathrm{E}[\langle V^{(i)} \otimes \tau_i(\xi_i), \nabla^{r+2} h(W(s))\rangle] + \mathrm{E}[\langle V^{(i)} \otimes \beta_i(\xi_i), \nabla^{r+1} h(W(s))\rangle].
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\sum_{i=1}^n \mathrm{E}[\langle V^{(i)} \otimes \xi_i, \nabla^{r+1} h(W(s))\rangle] \\
&= \sqrt{1-s} \left(\sum_{i=1}^n \mathrm{E}[\langle V \otimes \tau_i(\xi_i), \nabla^{r+2} h(W(s))\rangle] - \sum_{i=1}^n \mathrm{E}[\langle v_i(\xi_i) \otimes \tau_i(\xi_i), \nabla^{r+2} h(W(s))\rangle] \right)
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{i=1}^{n} \mathrm{E}[\langle V \otimes \beta_i(\xi_i), \nabla^{r+1} h(W(s))\rangle] - \sum_{i=1}^{n} \mathrm{E}[\langle v_i(\xi_i) \otimes \beta_i(\xi_i), \nabla^{r+1} h(W(s))\rangle] \\
&= \sqrt{1-s}\left(\mathrm{E}[\langle V \otimes T, \nabla^{r+2} h(W(s))\rangle] - \sum_{i=1}^{n} \mathrm{E}[\langle v_i(\xi_i) \otimes \tau_i(\xi_i), \nabla^{r+2} h(W(s))\rangle]\right) \\
&+ \mathrm{E}[\langle V \otimes B, \nabla^{r+1} h(W(s))\rangle] - \sum_{i=1}^{n} \mathrm{E}[\langle v_i(\xi_i) \otimes \beta_i(\xi_i), \nabla^{r+1} h(W(s))\rangle]. \qquad \text{(B.9)}
\end{aligned}
$$

Inserting (B.7)–(B.9) into (B.6) gives the desired result. □

Lemma B.2. *Let ξ be a centered random vector in $\mathbb{R}^d$. Suppose that ξ has a Stein kernel τ such that $\mathrm{E}[\|\xi\|_\infty^3] + \mathrm{E}[\|\tau(\xi) \otimes \xi^{\otimes 2}\|_\infty] < \infty$. Then, $\mathrm{E}[\|\tau(\xi) \otimes \xi\|_\infty] < \infty$ and for any $f \in C_b^4(\mathbb{R}^d)$,*

$$
\mathrm{E}[\langle \tau(\xi) \otimes \xi, \nabla^3 f(\xi)\rangle] = \frac{1}{2}\left(\mathrm{E}[\langle \xi^{\otimes 3}, \nabla^3 f(\xi)\rangle] - \mathrm{E}[\langle \tau(\xi) \otimes \xi^{\otimes 2}, \nabla^4 f(\xi)\rangle]\right). \qquad \text{(B.10)}
$$

Proof. Take $j, k, l \in \{1, \dots, d\}$ arbitrarily. Then we have

$$
\begin{aligned}
\mathrm{E}[|\tau_{jk}(\xi)\xi_l|] &= \mathrm{E}[|\tau_{jk}(\xi)\xi_l|; |\xi_l| \le 1] + \mathrm{E}[|\tau_{jk}(\xi)\xi_l|; |\xi_l| > 1] \\
&\le \mathrm{E}[|\tau_{jk}(\xi)|] + \mathrm{E}[|\tau_{jk}(\xi)|\xi_l^2] \le \mathrm{E}[\|\tau(\xi)\|_\infty] + \mathrm{E}[\|\tau(\xi) \otimes \xi^{\otimes 2}\|_\infty] < \infty.
\end{aligned}
$$

Hence $\mathrm{E}[\|\tau(\xi) \otimes \xi\|_\infty] < \infty$.

Next we prove (B.10). By a standard approximation argument, it suffices to prove (B.10) when f is compactly supported. For every $j = 1, \dots, d$, define a function $g_j : \mathbb{R}^d \to \mathbb{R}$ as

$$
g_j(x) = \langle x^{\otimes 2}, \nabla^2 \partial_j f(x)\rangle = \sum_{u,v=1}^{d} x_u x_v \partial_{juv} f(x), \qquad x \in \mathbb{R}^d.
$$

Observe that g_j is compactly supported because f is compactly supported. For $j, k \in \{1, \dots, d\}$ and $x \in \mathbb{R}^d$, we have

$$
\partial_k g_j(x) = 2\sum_{v=1}^{d} x_v \partial_{jkv} f(x) + \sum_{u,v=1}^{d} x_u x_v \partial_{jkuv} f(x).
$$

Hence we obtain

$$
\begin{aligned}
\mathrm{E}[\langle \tau(\xi) \otimes \xi, \nabla^3 f(\xi)\rangle] &= \sum_{j,k,v=1}^{d} \mathrm{E}[\tau_{jk}(\xi)\xi_v \partial_{jkv} f(\xi)] \\
&= \frac{1}{2}\sum_{j,k=1}^{d} \mathrm{E}[\tau_{jk}(\xi)\partial_k g_j(\xi)] - \frac{1}{2}\sum_{j,k,u,v=1}^{d} \mathrm{E}[\tau_{jk}(\xi)\xi_u\xi_v \partial_{jkuv} f(\xi)].
\end{aligned}
$$

The second term on the last line is equal to $\frac{1}{2}\mathrm{E}[\langle \tau(\xi) \otimes \xi^{\otimes 2}, \nabla^4 f(\xi)\rangle]$. To evaluate the first term, define a function $G : \mathbb{R}^d \to \mathbb{R}$ as $G(x) = \sum_{k=1}^d x_k \int_0^1 g_k(\theta x) d\theta$, $x \in \mathbb{R}^d$. Then, using the relation $\partial_k g_j = \partial_j g_k$, one can easily verify $\partial_j G = g_j$ for all $j = 1, \dots, d$. Moreover, G is bounded. In fact, since $g_1, \dots, g_d$ are compactly supported, there exists a constant $K > 0$ such that $g_j(x) = 0$ for all j whenever $\|x\|_\infty > K$. Then, for any $x \in \mathbb{R}^d$ with $\|x\|_\infty > K$,

$$|G(x)| \leq \sum_{k=1}^d |x_k| \int_0^{K/\|x\|_\infty} |g_k(\theta x)| d\theta \leq dK \max_{1 \leq k \leq d} \sup_{x \in \mathbb{R}^d} |g_k(x)|.$$

It is straightforward to check that this estimate is still true when $\|x\|_\infty \leq K$. Hence G is bounded. As a result, $G \in C_b^2(\mathbb{R}^d)$ and thus

$$\begin{aligned}\sum_{j,k=1}^d \mathrm{E}[\tau_{jk}(\xi)\partial_k g_j(\xi)] &= \sum_{j,k=1}^d \mathrm{E}[\tau_{jk}(\xi)\partial_{kj} G(\xi)] = \sum_{j=1}^d \mathrm{E}[\xi_j \partial_j G(\xi)] = \sum_{j=1}^d \mathrm{E}[\xi_j g_j(\xi)] \\ &= \mathrm{E}[\langle \xi^{\otimes 3}, \nabla^3 f(\xi)\rangle],\end{aligned}$$

where the second equality follows from the definition of Stein kernel. Combining these identities gives the desired result. ☐

Proof of Lemma 4.3. First we prove (4.6). It suffices to prove the claim when $h \in C_b^\infty(\mathbb{R}^d)$. To see this, let $t_1 = t/2$ and $t_2 = t/(2-t)$. One can easily check that $t_1, t_2 \in (0, 1]$ and $h_t = (h_{t_1})_{t_2}$. Since $h_{t_1} \in C_b^\infty(\mathbb{R}^d)$, the general case follows by applying the claim to $h = h_{t_1}$ and $t = t_2$.

Observe that $\mathrm{E}[h_t(W)] = \mathrm{E}[h(W(t))]$. Therefore, Lemma B.1(a) gives

$$\mathrm{E}[h_t(W)] - \mathrm{E}[h(Z)] = \frac{1}{2}\int_t^1 \left(\mathrm{E}[\langle \bar{T}, \nabla^2 h(W(s))\rangle] + \frac{\mathrm{E}[B \cdot \nabla h(W(s))]}{\sqrt{1-s}}\right) ds. \tag{B.11}$$

Next, for $t < s < 1$, Lemma B.1 also gives

$$\begin{aligned}&\mathrm{E}[\langle \bar{T}, \nabla^2 h(W(s))\rangle] - \mathrm{E}[\langle \bar{T}, \nabla^2 h(Z)\rangle] \\ &= \frac{1}{2}\int_s^1 \left(\mathrm{E}[\langle \bar{T}^{\otimes 2}, \nabla^4 h(W(u))\rangle] - \sum_{i=1}^n \mathrm{E}[\langle \tau_i(\xi_i)^{\otimes 2}, \nabla^4 h(W(u))\rangle]\right) du \\ &\quad + \frac{1}{2}\int_s^1 \frac{1}{\sqrt{1-u}}\left(\mathrm{E}[\langle \bar{T} \otimes B, \nabla^3 h(W(u))\rangle] + \sum_{i=1}^n \mathrm{E}[\langle \tau_i(\xi_i) \otimes (\xi_i - \beta_i(\xi_i)), \nabla^3 h(W(u))\rangle]\right) du\end{aligned} \tag{B.12}$$

and

$$\begin{aligned}&\mathrm{E}[B \cdot \nabla h(W(s))] - \mathrm{E}[B \cdot \nabla h(Z)] \\ &= \frac{1}{2}\int_s^1 \left(\mathrm{E}[\langle B \otimes \bar{T}, \nabla^3 h(W(u))\rangle] - \sum_{i=1}^n \mathrm{E}[\langle \beta_i(\xi_i) \otimes \tau_i(\xi_i), \nabla^3 h(W(u))\rangle]\right) du \\ &\quad + \frac{1}{2}\int_s^1 \frac{1}{\sqrt{1-u}}\left(\mathrm{E}[\langle B^{\otimes 2}, \nabla^2 h(W(u))\rangle] + \sum_{i=1}^n \mathrm{E}[\langle \beta_i(\xi_i) \otimes (\xi_i - \beta_i(\xi_i)), \nabla^2 h(W(u))\rangle]\right) du.\end{aligned} \tag{B.13}$$

Using the definition of the approximate Stein kernel, we can easily verify that $\beta_i(\xi_i)$ are centered. Hence

$$\mathrm{E}[B \cdot \nabla h(Z)] = \mathrm{E}[B] \cdot \mathrm{E}[\nabla h(Z)] = 0. \tag{B.14}$$

Meanwhile, for $s < u < 1$, Lemma B.2 gives

$$\begin{aligned}
&\sum_{i=1}^n \mathrm{E}[\langle \tau_i(\xi_i) \otimes \xi_i, \nabla^3 h(W(u))\rangle] \\
&= \frac{1}{2}\sum_{i=1}^n \left(\mathrm{E}[\langle \xi_i^{\otimes 3} - \beta_i(\xi_i) \otimes \xi_i^{\otimes 2}, \nabla^3 h(W(u))\rangle] - \sqrt{1-u}\,\mathrm{E}[\langle \tau_i(\xi_i) \otimes \xi_i^{\otimes 2}, \nabla^4 h(W(u))\rangle]\right).
\end{aligned} \tag{B.15}$$

Then, Lemma B.1 again gives

$$\begin{aligned}
&\mathrm{E}[\langle \xi_i^{\otimes 3}, \nabla^3 h(W(u))\rangle] - \mathrm{E}[\langle \xi_i^{\otimes 3}, \nabla^3 h(Z)\rangle] \\
&= \frac{1}{2}\int_u^1 \sum_{i=1}^n \mathrm{E}[\langle \xi_i^{\otimes 3} \otimes (\bar{T} - \tau_i(\xi_i)), \nabla^5 h(W(v))\rangle] dv \\
&\quad + \frac{1}{2}\int_u^1 \frac{1}{\sqrt{1-v}} \sum_{i=1}^n \mathrm{E}[\langle \xi_i^{\otimes 3} \otimes (\xi_i + B - \beta_i(\xi_i)), \nabla^4 h(W(v))\rangle] dv
\end{aligned} \tag{B.16}$$

and

$$\begin{aligned}
&\sum_{i=1}^n \mathrm{E}[\langle \beta_i(\xi_i) \otimes \xi_i, \nabla^2 h(W(u))\rangle] - \sum_{i=1}^n \mathrm{E}[\langle\langle \beta_i(\xi_i) \otimes \xi_i, \nabla^2 h(Z)\rangle] \\
&= \frac{1}{2}\int_u^1 \sum_{i=1}^n \mathrm{E}[\langle \beta_i(\xi_i) \otimes \xi_i \otimes (\bar{T} - \tau_i(\xi_i)), \nabla^4 h(W(v))\rangle] dv \\
&\quad + \frac{1}{2}\int_u^1 \frac{1}{\sqrt{1-v}} \sum_{i=1}^n \mathrm{E}[\langle \beta_i(\xi_i) \otimes \xi_i \otimes (\xi_i + B - \beta_i(\xi_i)), \nabla^3 h(W(v))\rangle] dv.
\end{aligned} \tag{B.17}$$

Further, note that we have by integration by parts

$$\begin{aligned}
\int_{\mathbb{R}^d} h_t(z) \nabla^r \phi_\Sigma(z) dz &= (-\sqrt{1-t})^r \int_{\mathbb{R}^d} \mathrm{E}[\nabla^r h(\sqrt{1-t}z + \sqrt{t}Z)]\phi_\Sigma(z) dz \\
&= (-\sqrt{1-t})^r\, \mathrm{E}[\nabla^r h(Z)]
\end{aligned}$$

for any $r \in \mathbb{N}$. Also, from the definition of the approximate Stein kernel, we can easily verify $\langle \mathrm{E}[\xi_i^{\otimes 2}], M\rangle = \langle \mathrm{E}[\tau_i(\xi_i) + \beta_i(\xi_i) \otimes \xi_i], M\rangle$ for any $i = 1, \dots, n$ and symmetric $d \times d$ matrix M. Hence

$$\begin{aligned}
&\frac{1-t}{2}\mathrm{E}[\langle \bar{T}, \nabla^2 h(Z)\rangle] + \frac{1}{4}\mathrm{E}\left[\left\langle \sum_{i=1}^n \beta_i(\xi_i) \otimes \xi_i, \nabla^2 h(Z)\right\rangle\right] \int_t^1 \frac{1}{\sqrt{1-s}}\left(\int_s^1 \frac{1}{\sqrt{1-u}} du\right) ds \\
&= \frac{1-t}{2}\langle \Sigma_W - \Sigma, \mathrm{E}[\nabla^2 h(Z)]\rangle = \frac{1}{2}\int_{\mathbb{R}^d} h_t(z)\langle \Sigma_W - \Sigma, \nabla^2 \phi_\Sigma(z)\rangle dz
\end{aligned} \tag{B.18}$$

and

$$\frac{\mathrm{E}[\langle \xi_i^{\otimes 3}, \nabla^3 h(Z)\rangle]}{8} \int_t^1 \left(\int_s^1 \frac{1}{\sqrt{1-u}} du\right) ds = -\frac{1}{6}\int_{\mathbb{R}^d} h_t(z) \langle \mathrm{E}[\xi_i^{\otimes 3}], \nabla^3 \phi_\Sigma(z)\rangle dz. \tag{B.19}$$

Finally, observe that

$$\nabla^r h_s(W) = (1-s)^{r/2}\,\mathrm{E}[\nabla^r h(W(s)) \mid \xi_1, \dots, \xi_n] \tag{B.20}$$

for any $r \in \mathbb{N}$ and $s \in [0, 1]$. Combining (B.11)–(B.20) gives (4.6).

Next we prove (4.7). As above, we may assume $h \in C_b^\infty(\mathbb{R}^d)$. By Lemma B.1(a),

$$\mathrm{E}[h_t(W)] - \mathrm{E}[h(Z)] = \frac{1}{2}\int_t^1 \mathrm{E}[\langle \Sigma_W - \Sigma, \nabla^2 h(W(s))\rangle]ds$$

and

$$\mathrm{E}[\langle \Sigma_W - \Sigma, \nabla^2 h(W(s))\rangle] - \mathrm{E}[\langle \Sigma_W - \Sigma, \nabla^2 h(Z)\rangle] = \frac{1}{2}\int_s^1 \mathrm{E}[\langle (\Sigma_W - \Sigma)^{\otimes 2}, \nabla^4 h(W(u))\rangle]du.$$

Combining these two identities with (B.20) gives (4.7). □

B.3 Proof of Lemma 4.1

As already mentioned, the proof is a straightforward modification of [8, Lemma 11.4]. Let

$$\delta := \sup\left\{\left|\int h_y d(\mu - \nu)\right| : y \in \mathbb{R}^d\right\}.$$

Assume first that

$$\delta = \sup\left\{\int h_y d(\mu - \nu) : y \in \mathbb{R}^d\right\}. \tag{B.21}$$

Then, given any $\eta > 0$, there exists a vector $z \in \mathbb{R}^d$ such that $\int h_z d(\mu - \nu) \geq \delta - \eta$. In this case, we have

$$\begin{aligned}
&\int_{[-\varepsilon,\varepsilon]^d}\left[\int M_{h_z}(y + x; \varepsilon)(\mu - \nu)(dy)\right]K(dx)\\
&\geq \int_{[-\varepsilon,\varepsilon]^d}\left[\int h_z(y)\mu(dy) - \int M_{h_z}(y + x; \varepsilon)\nu(dy)\right]K(dx)\\
&= \int_{[-\varepsilon,\varepsilon]^d}\left[\int h_z(y)(\mu - \nu)(dy) - \int \{M_{h_z}(y + x; \varepsilon) - h_z(y)\}\,\nu(dy)\right]K(dx)\\
&\geq \int_{[-\varepsilon,\varepsilon]^d}\left[\delta - \eta - \int \{M_{h_z}(y + x; \varepsilon) - h_z(y + x)\}\,\nu(dy) - \int \{h_z(y + x) - h_z(y)\}\,\nu(dy)\right]K(dx)\\
&\geq \int_{[-\varepsilon,\varepsilon]^d}[\delta - \eta - \tau^*(h; \varepsilon) - \tilde{\tau}^*(h; \varepsilon)]\,K(dx) = \alpha\,[\delta - \eta - \tau^*(h; \varepsilon) - \tilde{\tau}^*(h; \varepsilon)]
\end{aligned}$$

and

$$\begin{aligned}
&\int_{\mathbb{R}^d\setminus[-\varepsilon,\varepsilon]^d}\left[\int M_{h_z}(y + x; \varepsilon)(\mu - \nu)(dy)\right]K(dx)\\
&\geq \int_{\mathbb{R}^d\setminus[-\varepsilon,\varepsilon]^d}\left[\int h_z(y + x)\mu(dy) - \int M_{h_z}(y + x; \varepsilon)\nu(dy)\right]K(dx)\\
&= \int_{\mathbb{R}^d\setminus[-\varepsilon,\varepsilon]^d}\left[\int h_z(y + x)(\mu - \nu)(dy) - \int \{M_{h_z}(y + x; \varepsilon) - h_z(y + x)\}\,\nu(dy)\right]K(dx)\\
&\geq \int_{\mathbb{R}^d\setminus[-\varepsilon,\varepsilon]^d}[-\delta - \tau^*(h; \varepsilon)]\,K(dx) = (1 - \alpha)\,[-\delta - \tau^*(h; \varepsilon)]\,.
\end{aligned}$$

Consequently, we obtain

$$\gamma^*(h; \varepsilon) \geq \int M_{h_z}(x; \varepsilon)(\mu - \nu) * K(dx) \geq (2\alpha - 1)\delta - \tau^*(h; \varepsilon) - \alpha\tilde{\tau}^*(h; \varepsilon) - \alpha\eta.$$

Letting $\eta \downarrow 0$, we obtain the desired result. If instead of (B.21) we have

$$\delta = \sup\left\{-\int h_y d(\mu-\nu) : y \in \mathbb{R}^d\right\},$$

then, given any $\eta > 0$, we can find $z \in \mathbb{R}^d$ such that $-\int h_z d(\mu-\nu) \geq \delta - \eta$. Now look at $-h_z$ (instead of h_z) and note that $M_{-h_y}(\cdot;\varepsilon) = -m_{h_y}(\cdot;\varepsilon)$ and

$$\int \{h_y - m_{h_y}(x;\varepsilon)\}\nu(dx) = \int \{M_{-h_y}(x;\varepsilon) - (-h_y)\}\nu(dx)$$

for every $y \in \mathbb{R}^d$. Proceeding exactly as above, we obtain

$$\gamma^*(h;\varepsilon) \geq -\int m_{h_z}(x;\varepsilon)(\mu-\nu) * K(dx) \geq (2\alpha-1)\delta - \tau^*(h;\varepsilon) - \alpha\tilde{\tau}^*(h;\varepsilon) - \alpha\eta.$$

Thus we complete the proof. □

C Proof of Lemma 5.1

Let us prove (5.5). We write $F = F_Z$ and $x = F^{-1}(p)$ for short. First we consider the case $p \geq 1/2$. Since the function $\mathbb{R}^d \ni z \mapsto z^\vee \in \mathbb{R}$ is convex, $\Phi^{-1} \circ F$ is concave by Corollary A.2.9 in [59]. Hence, for any $y > x$,

$$\frac{(\Phi^{-1}\circ F)(y) - (\Phi^{-1}\circ F)(x)}{y-x} \leq (\Phi^{-1}\circ F)'(x) = \frac{f(x)}{\phi(\Phi^{-1}(F(x)))}.$$

Let $y = F^{-1}(p) + c$ with c a positive constant specified later. Then,

$$\frac{f(F^{-1}(p))}{\phi(\Phi^{-1}(p))} \geq \frac{\Phi^{-1}(F(F^{-1}(p)+c)) - \Phi^{-1}(p)}{c}.$$

Thus, we need to choose c so that $F(F^{-1}(p)+c)$ has an appropriate lower bound. Noting $p \geq 1/2$, we have

$$\begin{aligned} 1 - F(F^{-1}(p)+c) &= P\left(Z^\vee > F^{-1}(p)+c\right) \leq P\left(Z^\vee > F^{-1}(1/2)+c\right) \\ &\leq \frac{\mathrm{E}[(Z^\vee - F^{-1}(1/2))^2]}{c^2} \leq \frac{2\,\mathrm{Var}[Z^\vee]}{c^2}, \end{aligned}$$

where we used the following inequality for the last bound: For any random variable Y and its median m, $\mathrm{E}[(Y-m)^2] = \mathrm{Var}[Y] + (\mathrm{E}[Y]-m)^2 \leq 2\,\mathrm{Var}[Y]$. Thus, letting $c = \sqrt{4\,\mathrm{Var}[Z^\vee]/(1-p)}$, we obtain $F(F^{-1}(p)+c)) \geq 1-(1-p)/2 = (1+p)/2$. Consequently,

$$f(F^{-1}(p)) \geq \phi(\Phi^{-1}(p))\frac{\Phi^{-1}((1+p)/2) - \Phi^{-1}(p)}{2\sqrt{\mathrm{Var}[Z^\vee]}}\sqrt{1-p}. \tag{C.1}$$

Also, by the fundamental theorem of calculus,

$$\Phi^{-1}((1+p)/2) - \Phi^{-1}(p) = \int_p^{(1+p)/2} \frac{1}{\phi(\Phi^{-1}(u))}du \geq \frac{1-p}{2\phi(\Phi^{-1}(p))}, \tag{C.2}$$

where the last inequality holds because $\phi\circ\Phi^{-1}$ is decreasing on $[1/2, 1)$. Combining (C.1) with (C.2) gives (5.5).

Next consider the case $p < 1/2$. Since the function $\log \circ \Phi$ is increasing and concave, $\log \circ F = (\log \circ \Phi) \circ (\Phi^{-1} \circ F)$ is concave. Hence $(\log \circ F)' = f/F$ is non-increasing. Thus we obtain

$$\frac{f(F^{-1}(p))}{p} \geq \frac{f(F^{-1}(1/2))}{F(F^{-1}(1/2))} \geq \frac{1}{2^{5/2}\sqrt{\mathrm{Var}[Z^\vee]}},$$

where the last inequality follows from (5.5) for $p = 1/2$ which was already proved in the above.

It remains to prove (5.6) when $\mathrm{Cov}[Z] = \Sigma$. An elementary computation shows

$$(F_Z^{-1})'(u) = \frac{1}{f_\Sigma(F_Z^{-1}(u))}, \qquad (F_Z^{-1})''(u) = -\frac{f_\Sigma'(F_Z^{-1}(u))}{f_\Sigma(F_Z^{-1}(u))^3}$$

for all $u \in (0, 1)$. Thus, the desired result follows from (5.5) and Lemma D.5. $\square$

D Technical tools

D.1 Inequalities related to multivariate normal distributions

Let Z be a centered Gaussian vector in $\mathbb{R}^d$. Set $\overline{\sigma} := \max_{1\leq j\leq d}\sqrt{\mathrm{Var}[Z_j]}$ and $\underline{\sigma} := \min_{1\leq j\leq d}\sqrt{\mathrm{Var}[Z_j]}$.

Lemma D.1. $\mathrm{E}[Z^\vee] \leq \overline{\sigma}\sqrt{2\log d}$ *and* $\mathrm{Var}[Z^\vee] \leq \overline{\sigma}^2$.

Proof. The first bound follows from [12, Theorem 2.5]. The second one follows from [12, Theorem 5.8]. $\square$

Lemma D.2 (Nazarov's inequality)**.** *If* $\underline{\sigma} > 0$*, then for any* $x \in \mathbb{R}^d$ *and* $\varepsilon > 0$*,*

$$P\left(0 < \max_{1\leq j\leq d}(Z_j - x_j) \leq \varepsilon\right) \leq \frac{\varepsilon}{\underline{\sigma}}(\sqrt{2\log d} + 2).$$

Proof. See [21, Theorem 1]. $\square$

Lemma D.3. *There exists a universal constant* $c > 0$ *such that* $\sqrt{\mathrm{Var}[Z^\vee]\log d} \geq c\underline{\sigma}^2/\overline{\sigma}$*.*

Proof. By a straightforward modification of the proof of Theorem 1.8 in [27], we can prove

$$\sqrt{\mathrm{Var}[Z^\vee]}(\underline{\sigma} + \mathrm{E}[Z^\vee]) \geq c'\underline{\sigma}^2$$

for some universal constant $c' > 0$. In fact, this follows by applying the arguments in the proof of Theorem 1.8 in [27] to $\mathbf{X} = Z/\overline{\sigma}$ with $t = \sqrt{1 - \underline{\sigma}^2/(4m^2)}$ when $m := \mathrm{E}[Z^\vee] > \underline{\sigma}/2$ (Lemma 2.2 in [27] is applied with $\lambda = \underline{\sigma}^2/(4m\overline{\sigma})$). Then, since $\mathrm{E}[Z^\vee] \leq \overline{\sigma}\sqrt{2\log d}$ by Lemma D.1, we obtain the desired result. $\square$

Lemma D.4 (Anderson–Hall–Titterington's bound)**.** *For any* $r \in \mathbb{N}$*,*

$$\sup_{A\in\mathcal{R}}\left\|\int_A \nabla^r \phi_\Sigma(z)dz\right\|_1 \leq C_r \frac{\log^{r/2} d}{\sigma_*^r},$$

where $C_r > 0$ *is a constant depending only on* r*.*

Proof. Let $Z \sim N(0, \Sigma)$ and $Z' \sim N(0, \Sigma - \sigma_*^2 I_d)$. Then, for any $A \in \mathcal{R}$ and $x \in \mathbb{R}^d$, we have

$$\int_A \phi_\Sigma(x + z)dz = \mathrm{E}[1_A(Z - x)] = \mathrm{E}\left[\int_{\mathbb{R}^d} 1_A(\sigma_* z + Z')\phi_d(z + x/\sigma_*)dz\right].$$

Differentiating both sides r times with respect to x and then setting $x = 0$, we obtain

$$\int_A \nabla^r \phi_\Sigma(z)dz = \frac{1}{\sigma_*^r} \mathrm{E}\left[\int_{\mathbb{R}^d} 1_A(\sigma_* z + Z')\nabla^r \phi_d(z)dz\right].$$

Hence

$$\left\|\int_A \nabla^r \phi_\Sigma(z)dz\right\|_1 \leq \frac{1}{\sigma_*^r} \mathrm{E}\left[\left\|\int_{\mathbb{R}^d} 1_A(\sigma_* z + Z')\nabla^r \phi_d(z)dz\right\|_1\right] \leq C_r \frac{\log^{r/2} d}{\sigma_*^r},$$

where the last inequality follows by Lemma 2.2 in [29] because $\{z \in \mathbb{R}^d : \sigma_* z + Z' \in A\} \in \mathcal{R}$. □

Lemma D.5. *For any integer $r \geq 0$,*

$$f_\Sigma^{(r)}(t) = \int_{(-\infty,t]^d} \langle \mathbf{1}_d^{\otimes(r+1)}, \nabla^{r+1}\phi_\Sigma(z)\rangle dz \quad \textit{for all } t \in \mathbb{R}. \tag{D.1}$$

Moreover, there exists a constant $C_r > 0$ depending only on r such that $\sup_{t\in\mathbb{R}} |f_\Sigma^{(r)}(t)| \leq C_r(\sqrt{\log d}/\sigma_*)^{r+1}$.

Proof. Observe that

$$F_Z(t) = \int_{(-\infty,t]^d} \phi_\Sigma(z)dz = \int_{(-\infty,0]^d} \phi_\Sigma(z + t\mathbf{1}_d)dz$$

for all $t \in \mathbb{R}$. Differentiating this equation $r + 1$ times with respect to t gives

$$f_\Sigma^{(r)}(t) = \int_{(-\infty,0]^d} \langle \mathbf{1}_d^{\otimes(r+1)}, \nabla^{r+1}\phi_\Sigma(z + t\mathbf{1}_d)\rangle dz = \int_{(-\infty,t]^d} \langle \mathbf{1}_d^{\otimes(r+1)}, \nabla^{r+1}\phi_\Sigma(z)\rangle dz.$$

Hence we obtain (D.1). The second claim follows from Lemma D.4. □

D.2 Inequalities related to sub-Weibull norms

Lemma D.6. *Let ξ be a random variable. Suppose that there is a constant $A > 0$ such that $\|\xi\|_p \leq Ap^{1/\alpha}$ for all $p \geq 1$. Then $\|\xi\|_{\psi_\alpha} \leq C_\alpha A$.*

Proof. See Lemma A.5 in [40]. □

Lemma D.7. *For any $\alpha \in (0, 1)$, there exists a constant $C_\alpha > 0$ depending only on α such that*

$$\left\|\sum_{i=1}^n \xi_i\right\|_{\psi_\alpha} \leq C_\alpha \sum_{i=1}^n \|\xi_i\|_{\psi_\alpha}$$

for any random variables $\xi_1, \dots, \xi_n$.

Proof. This follows from Lemma C.2 in [17] and the triangle inequality for the Orlicz norm associated with a convex function. □

Lemma D.8. *Let ξ_1, ξ_2 be two random variables such that $\|\xi_1\|_{\psi_{\alpha_1}} + \|\xi_2\|_{\psi_{\alpha_2}} < \infty$ for some $\alpha_1, \alpha_2 > 0$. Then we have $\|\xi_1\xi_2\|_{\psi_\alpha} \leq \|\xi_1\|_{\psi_{\alpha_1}}\|\xi_2\|_{\psi_{\alpha_2}}$, where $\alpha > 0$ is defined by the equation $1/\alpha = 1/\alpha_1 + 1/\alpha_2$.*

Proof. See [41, Proposition D.2]. □

Lemma D.9. *Let $\xi_1, \dots, \xi_n$ be independent random variables such that* $\max_{1\le i\le n} \|\xi_i\|_{\psi_\alpha} \le K$ *for some $K > 0$ and $\alpha \in (0, 1]$. Then, there is a constant $C_\alpha > 0$ depending only on α such that for any $p \ge 1$,*

$$\left\| \sum_{i=1}^n (\xi_i - \mathrm{E}[\xi_i]) \right\|_p \le C_\alpha K \left(\sqrt{pn} + p^{1/\alpha} \right).$$

Proof. See Lemma 2.1 in [30]. □

Lemma D.10. *Let $Y_1, \dots, Y_n$ be independent random vectors in $\mathbb{R}^d$. Suppose that there exist constants $K > 0$ and $\alpha \in (0, 1]$ such that* $\max_{1\le i\le n} \max_{1\le j\le d} \|Y_{ij}\|_{\psi_\alpha} \le K$*. Then, there exists a constant $C_\alpha > 0$ depending only on α such that*

$$\mathrm{E} \left\| \sum_{i=1}^n (Y_i - \mathrm{E}[Y_i]) \right\|_\infty^r \le C_\alpha^r K^r \left(\sqrt{nr \log d} + (r \log d)^{1/\alpha} \right)^r \tag{D.2}$$

for any $r \ge 1$ and

$$P\left(\left\| \sum_{i=1}^n (Y_i - \mathrm{E}[Y_i]) \right\|_\infty > C_\alpha K \left(\sqrt{an \log(dn)} + a^{1/\alpha} \log^{1/\alpha}(dn) \right) \right) \le \frac{1}{n^a} \tag{D.3}$$

for any $a \ge 1$.

Proof. By Lemma D.9, we have for any $p \ge 2$

$$\max_{1\le j\le d} \left\| \sum_{i=1}^n (Y_{ij} - \mathrm{E}[Y_{ij}]) \right\|_p \le C'_\alpha K \left(\sqrt{pn} + p^{1/\alpha} \right), \tag{D.4}$$

where $C'_\alpha > 0$ depends only on α. Therefore, with $p = r \log d$, we have

$$\begin{aligned} \mathrm{E} \left\| \sum_{i=1}^n (Y_i - \mathrm{E}[Y_i]) \right\|_\infty^r &\le \left(\mathrm{E} \left\| \sum_{i=1}^n (Y_i - \mathrm{E}[Y_i]) \right\|_\infty^p \right)^{r/p} \le d^{r/p} \max_{1\le j\le d} \left\| \sum_{i=1}^n (Y_{ij} - \mathrm{E}[Y_{ij}]) \right\|_p^r \\ &\le d^{r/p} (C'_\alpha)^r K^r \left(\sqrt{nr \log d} + (r \log d)^{1/\alpha} \right)^r. \end{aligned}$$

Since $d^{r/p} = e^{r \log d/p} = e \le e^r$, we obtain (D.2) with $C_\alpha = eC'_\alpha$. Also, by the union bound, Markov's inequality and (D.4), we have for any $t > 0$ and $p \ge 2$

$$P\left(\left\| \sum_{i=1}^n (Y_i - \mathrm{E}[Y_i]) \right\|_\infty > t \right) \le d \left(t^{-1} C'_\alpha K \left(\sqrt{pn} + p^{1/\alpha} \right) \right)^p.$$

Applying this estimate with $p = a \log(dn)$ and $t = eC'_\alpha K \left(\sqrt{pn} + p^{1/\alpha} \right)$, we obtain (D.3) □

Lemma D.11. *Let $r \in \mathbb{N}$. If* (2.2) *is satisfied, there exists a constant $C_r > 0$ depending only on r such that*

$$P\left(\left| \frac{1}{n} \sum_{i=1}^n (\langle X_i^{\otimes r}, V \rangle - \mathrm{E}[\langle X_i^{\otimes r}, V \rangle]) \right| > a^r C_r \|V\|_1 b^r \sqrt{\frac{\log n}{n}} \right) \le \frac{1}{n^a}$$

for any $a \ge 1$ and $V \in (\mathbb{R}^d)^{\otimes r}$.

Proof. By Lemmas D.7 and D.8, for every $i = 1, \dots, n$,

$$\left\| \langle X_i^{\otimes r}, V \rangle \right\|_{\psi_{1/r}} \le C_r \sum_{j_1,\dots,j_r=1}^{d} \|X_{ij_1} \cdots X_{ij_r}\|_{\psi_{1/r}} |V_{j_1,\dots,j_r}| \le C_r \|V\|_1 \max_j \|X_{ij}\|_{\psi_1}^r .$$

Therefore, by Lemma D.9, there exists a constant $C_r' > 0$ depending only on r such that

$$\left\| \frac{1}{n} \sum_{i=1}^{n} (\langle X_i^{\otimes r}, V \rangle - \mathrm{E}[\langle X_i^{\otimes r}, V \rangle]) \right\|_p \le C_r' \|V\|_1 b^r \left(\sqrt{\frac{p}{n}} + \frac{p^r}{n} \right)$$

for any $p \ge 1$. Hence, with $p = a \log n$, we have by Markov's inequality

$$P\left(\left| \frac{1}{n} \sum_{i=1}^{n} (\langle X_i^{\otimes r}, V \rangle - \mathrm{E}[\langle X_i^{\otimes r}, V \rangle]) \right| > e C_r' \|V\|_1 b^r \left(\sqrt{\frac{p}{n}} + \frac{p^r}{n} \right) \right) \le e^{-p} = \frac{1}{n^a}.$$

Noting $n^{-1} \log^r n = \sqrt{n^{-1} \log n} \cdot n^{-1/2} \log^{r-1/2} n \le C_r \sqrt{n^{-1} \log n}$, we complete the proof. □

Lemma D.12. *If* (2.2) *is satisfied, there exists a universal constant $C > 0$ such that for any $a \ge 1$,*

$$P\left(\max_{1 \le j,k \le d} \frac{1}{n} \sum_{i=1}^{n} |X_{ij}||X_{ik}| 1_{\{|X_{ij}| \vee |X_{ik}| > 2b \log n\}} > C b^2 \left(\sqrt{\frac{a \log(dn)}{n}} + \frac{a^2 \log^2(dn)}{n} \right) \right) \le \frac{1}{n^a} \quad \text{(D.5)}$$

and

$$P\left(\max_{1 \le j,k \le d} \frac{1}{n} \sum_{i=1}^{n} X_{ij}^2 X_{ik}^2 1_{\{|X_{ij}| \vee |X_{ik}| \le 2b \log n\}} > C b^4 \left(1 + a \frac{\log(dn) \log^4 n}{n} \right) \right) \le \frac{1}{n^a}. \quad \text{(D.6)}$$

Proof. Let us prove (D.5). Write $Y_{i,jk} := |X_{ij}||X_{ik}| 1_{\{|X_{ij}| \vee |X_{ik}| > 2b \log n\}}$ for short. Since $\|Y_{i,jk}\|_{\psi_{1/2}} \le b^2$, we have by Lemma D.10

$$\max_{1 \le j,k \le d} \frac{1}{n} \left| \sum_{i=1}^{n} (Y_{i,jk} - \mathrm{E}[Y_{i,jk}]) \right| \lesssim b^2 \left(\sqrt{\frac{a \log(dn)}{n}} + \frac{a^2 \log^2(dn)}{n} \right)$$

with probability at least $1 - 1/n^a$. Moreover, observe that

$$|\mathrm{E}[Y_{i,jk}]| \le \max_{1 \le j,k \le d} \sqrt{\mathrm{E}[X_{ij}^2 X_{ik}^2] P(|X_{ij}| > 2b \log n)} \lesssim \frac{b^2}{n}.$$

Hence we conclude

$$\max_{1 \le j,k \le d} \frac{1}{n} \left| \sum_{i=1}^{n} Y_{i,jk} \right| \lesssim b^2 \left(\sqrt{\frac{a \log(dn)}{n}} + \frac{a \log^2(dn)}{n} \right)$$

with probability at least $1 - 1/n^a$. This proves (D.5).

Next we prove (D.6). Set

$$\zeta := \max_{1 \le j,k \le d} \frac{1}{n} \sum_{i=1}^{n} X_{ij}^2 X_{ik}^2 1_{\{|X_{ij}| \vee |X_{ik}| \le 2b \log n\}}.$$

By Lemma E.5 in [20] with $\eta = 3$ and $B = 16b^4 \log^4 n/n$, we have for any $t > 0$

$$P\left(\zeta \geq 4\,\mathrm{E}[\zeta] + 16\frac{b^4 \log^4 n}{n}t\right) \leq e^{-t}.$$

Further, $\mathrm{E}[\zeta] \lesssim b^4 + \frac{b^4 \log^4 n}{n}\log d$ by Lemma 9 in [19]. Hence there exists a universal constant $C_2 > 0$ such that

$$P\left(\zeta \geq C_2 b^4\left(1 + \frac{\log^4 n}{n}(\log d + t)\right)\right) \leq e^{-t}$$

for any $t > 0$. Applying this with $t = a\log n$ gives (D.6). □

Acknowledgments The author thanks Xiao Fang and Ryo Imai for valuable discussions about the subject of this paper. Thanks are also due to Kouki Akatsuka for his helpful comments about several imperfections contained in the original version of the paper. This work was partly supported by JST CREST Grant Number JPMJCR2115 and JSPS KAKENHI Grant Numbers JP22H00834, JP22H00889, JP22H01139, JP24K14848.

References

[1] Alonso-Gutiérrez, D. & Bastero, J. (2015). *Approaching the Kannan-Lovász-Simonovits and variance conjectures*. Springer, Switzerland.

[2] Anderson, N. H., Hall, P. & Titterington, D. (1998). Edgeworth expansions in very-high-dimensional problems. *J. Statist. Plann. Inference* **70**, 1–18.

[3] Artstein, S., Ball, K. M., Barthe, F. & Naor, A. (2004). Solution of Shannon's problem on the monotonicity of entropy. *J. Amer. Math. Soc.* **17**, 975–982.

[4] Barbour, A. (1986). Asymptotic expansions based on smooth functions in the central limit theorem. *Probab. Theory Relat. Fields* **72**, 289–303.

[5] Barthe, F. & Maurey, B. (2000). Some remarks on isoperimetry of Gaussian type. *Ann. Inst. Henri Poincaré Probab. Stat.* **36**, 419–434.

[6] Belloni, A., Chernozhukov, V., Chetverikov, D., Hansen, C. & Kato, K. (2018). High-dimensional econometrics and regularized GMM. Working paper. arXiv: 1806.01888.

[7] Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *J. Amer. Statist. Assoc.* **83**, 687–697.

[8] Bhattacharya, R. N. & Rao, R. R. (2010). *Normal approximation and asymptotic expansions*. SIAM, Philadelphia.

[9] Bobkov, S. G. (1999). Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.* **27**, 1903–1921.

[10] Bobkov, S. G. & Houdré, C. (1997). Isoperimetric constants for product probability measures. *Ann. Probab.* **25**, 184–205.

[11] Bobkov, S. G. & Ledoux, M. (2019). *One-dimensional empirical measures, order statistics, and Kantorovich transport distances*, vol. 261 of *Memoirs of the American Mathematical Society*. American Mathematical Society, Providence.

[12] Boucheron, S., Lugosi, G. & Massart, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, Oxford.

[13] Bühlmann, P. & van de Geer, S. (2011). *Statistics for high-dimensional data*. Springer, Berlin Heidelberg.

[14] Cacoullos, T. & Papathanasiou, V. (1992). Lower variance bounds and a new proof of the central limit theorem. *J. Multivariate Anal.* **43**, 173–184.

[15] Chang, J. & Hall, P. (2015). Double-bootstrap methods that use a single double-bootstrap simulation. *Biometrika* **102**, 203–214.

[16] Chatterjee, S. (2014). *Superconcentration and related topics*. Springer, Switzerland.

[17] Chen, X. & Kato, K. (2019). Randomized incomplete U-statistics in high dimensions. *Ann. Statist.* **47**, 3127–3156.

[18] Chernozhukov, V., Chetverikov, D. & Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41**, 2786–2819.

[19] Chernozhukov, V., Chetverikov, D. & Kato, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162**, 47–70.

[20] Chernozhukov, V., Chetverikov, D. & Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45**, 2309–2353.

[21] Chernozhukov, V., Chetverikov, D. & Kato, K. (2017). Detailed proof of Nazarov's inequality. Unpublished paper. arXiv: 1711.10696.

[22] Chernozhukov, V., Chetverikov, D., Kato, K. & Koike, Y. (2022). Improved central limit theorem and bootstrap approximation in high dimensions. *Ann. Statist.* **50**, 2562–2586.

[23] Chernozhukov, V., Chetverikov, D., Kato, K. & Koike, Y. (2023). High-dimensional data bootstrap. *Annu. Rev. Stat. Appl.* **10**, 427–449.

[24] Chernozhukov, V., Chetverikov, D. & Koike, Y. (2023). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab.* **33**, 2374–2425.

[25] Davidson, R. & MacKinnon, J. G. (2007). Improving the reliability of bootstrap tests with the fast double bootstrap. *Comput. Statist. Data Anal.* **51**, 3259–3281.

[26] Deng, H. & Zhang, C.-H. (2020). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Ann. Statist.* **48**, 3643–3671.

[27] Ding, J., Eldan, R. & Zhai, A. (2015). On multiple peaks and moderate deviations for the supremum of a Gaussian field. *Ann. Probab.* **43**, 3468–3493.

[28] Döbler, C. (2025). On existence and uniqueness of univariate Stein kernels. *Electron. Commun. Probab.* **30**, 1–11.

[29] Fang, X. & Koike, Y. (2021). High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.* **31**, 1660–1686.

[30] Fang, X. & Koike, Y. (2023). From p-Wasserstein bounds to moderate deviations. *Electron. J. Probab.* **28**, 1–52.

[31] Fang, X. & Koike, Y. (2024). Sharp high-dimensional central limit theorems for log-concave distributions. *Ann. Inst. Henri Poincaré Probab. Stat.* **60**, 2129–2156.

[32] Fang, X., Koike, Y., Liu, S.-H. & Zhao, Y.-K. (2023). High-dimensional central limit theorems by Stein's method in the degenerate case. Working paper. arXiv: 2305.17365v1.

[33] Fang, X. & Liu, S.-H. (2025). Edgeworth expansion by Stein's method. *Bernoulli* **31**, 2018–2041.

[34] Fathi, M. (2019). Stein kernels and moment maps. *Ann. Probab.* **47**, 2172–2185.

[35] Fujikoshi, Y. & Ulyanov, V. V. (2020). *Non-asymptotic analysis of approximations for multivariate statistics*. Springer, Singapore.

[36] Fujikoshi, Y., Ulyanov, V. V. & Shimizu, R. (2010). *Multivariate statistics*. Wiley, Hoboken, New Jersey.

[37] Gaunt, R. E. (2016). Rates of convergence in normal approximation under moment conditions via new bounds on solutions of the Stein equation. *J. Theoret. Probab.* **29**, 231–247.

[38] Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer, New York.

[39] Koike, Y. (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* **4**, 643–696.

[40] Koike, Y. (2023). High-dimensional central limit theorems for homogeneous sums. *J. Theoret. Probab.* **36**, 1–45.

[41] Kuchibhotla, A. K. & Chakrabortty, A. (2022). Moving beyond sub-Gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *Inf. Inference* **11**, 1389–1456.

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

[43] Leadbetter, M. R., Lindgren, G. & Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer, New York.

[44] Ledoux, M., Nourdin, I. & Peccati, G. (2015). Stein's method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.* **25**, 256–306.

[45] Lee, S. M. S. & Young, G. A. (1999). The effect of Monte Carlo approximation on coverage error of double-bootstrap confidence intervals. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61**, 353–366.

[46] Ley, C., Reinert, G. & Swan, Y. (2017). Stein's method for comparison of univariate distributions. *Probab. Surv.* **14**, 1–52.

[47] Liu, R. Y. & Singh, K. (1987). On a partial correction by the bootstrap. *Ann. Statist.* **15**, 1713–1718.

[48] Lopes, M. E. (2022). Central limit theorem and bootstrap approximation in high dimensions with near $1/\sqrt{n}$ rates. *Ann. Statist.* **50**, 2492–2513.

[49] Lopes, M. E., Lin, Z. & 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.

[50] Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.* **21**, 255–285.

[51] Mijoule, G., Raič, M., Reinert, G. & Swan, Y. (2023). Stein's density method for multivariate continuous distributions. *Electron. J. Probab.* **28**, 1–40.

[52] Nankervis, J. C. (2005). Computational algorithms for double bootstrap confidence intervals. *Comput. Statist. Data Anal.* **49**, 461–475.

[53] Nelsen, R. B. (2006). *An introduction to copulas*. Springer, New York, 2nd edn.

[54] Nourdin, I. & Peccati, G. (2009). Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145**, 75–118.

[55] Nourdin, I. & Peccati, G. (2012). *Normal approximations with Malliavin calculus: From Stein's method to universality*. Cambridge University Press, Cambridge.

[56] Nourdin, I., Peccati, G. & Swan, Y. (2014). Entropy and the fourth moment phenomenon. *J. Funct. Anal.* **266**, 3170–3207.

[57] Raič, M. (2019). A multivariate Berry–Esseen theorem with explicit constants. *Bernoulli* **25**, 2824–2853.

[58] Stein, C. (1986). *Approximate computation of expectations*. Institute of Mathematical Statistics, Hayward, California.

[59] van der Vaart, A. W. & Wellner, J. A. (2023). *Weak convergence and empirical processes*. Springer, Switzerland, 2nd edn.

[60] Zhilova, M. (2022). New Edgeworth-type expansions with finite sample guarantees. *Ann. Statist.* **50**, 2545–2561.