SECOND- AND HIGHER-ORDER GAUSSIAN ANTICONCENTRATION INEQUALITIES AND ERROR BOUNDS IN SLEPIAN'S COMPARISON THEOREM

CUN-HUI ZHANG

ABSTRACT

This paper presents some second- and higher-order Gaussian anticoncentration inequalities in high dimension and error bounds in Slepian's comparison theorem for the distribution functions of the maxima of two Gaussian vectors. The anticoncentration theorems are presented as upper bounds for the sum of the absolute values of the partial derivatives of a certain order for the joint distribution function of a Gaussian vector or weighted sums of such absolute values. Compared with the existing results where the covariance matrix of the entire Gaussian vector is required to be invertible, the bounds for the mth derivatives developed in this paper require only the invertibility of the covariance matrices of all subsets of m random variables. The second-order anticoncentration inequality is used to develop comparison theorems for the joint distribution functions of Gaussian vectors or, equivalently, the univariate distribution functions of their maxima via Slepian's interpolation. The third- and higher-order anticoncentration inequalities are motivated by recent advances in the central limit theorem and consistency of bootstrap for the maximum component of a sum of independent random vectors in high dimension and related applications in statistical inference and machine learning.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 60E15; Secondary 60G15, 62E20, 62E17

KEYWORDS

Anti-concentration, comparison of distributions, Gaussian process, density of maximum

INTERNATIONAL CONGRESS OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 7, pp. 5594–5615
DOI 10.4171/ICM2022/186

Published by EMS Press

1. INTRODUCTION

Let $X = (X_1, \ldots, X_d)^\top$ and $Y = (Y_1, \ldots, Y_d)^\top$ be two Gaussian vectors. Slepian's [31] inequality asserts that when X_i and Y_i have the same mean and variance, and $\mathrm{Var}(X_i - X_j) \le \mathrm{Var}(Y_i - Y_j)$ for all $1 \le i < j \le d$, the maximum of Y_i is stochastically larger than the maximum of X_i,

$$\mathbb{P}\Big\{\max_{1\le i\le d} X_i > t\Big\} \le \mathbb{P}\Big\{\max_{1\le i\le d} Y_i > t\Big\}, \quad \forall t \in \mathbb{R}. \tag{1.1}$$

Variations and extensions of Slepian's inequality have been developed to relax the conditions on the mean and variance of the individual components and pairwise contrasts, and to compare more general functions of the Gaussian vectors. Among such results, the Sudakov–Fernique inequality [15, 32, 33] asserts that

$$\mathbb{E}\Big[\max_{i\le d} X_i\Big] \le \mathbb{E}\Big[\max_{i\le d} Y_i\Big] \tag{1.2}$$

when $\mathbb{E}[X] = \mathbb{E}[Y]$ and $\mathrm{Var}(X_i - X_j) \le \mathrm{Var}(Y_i - Y_j)$ for all $1 \le i < j \le d$. Gordon's [16] inequalities extend (1.1) and (1.2) to the minimax function of Gaussian matrices. Chatterjee [5] provided an error bound in the Sudakov–Fernique inequality

$$\Big|\mathbb{E}\Big[\max_{i\le d} Y_i\Big] - \mathbb{E}\Big[\max_{i\le d} X_i\Big]\Big| \le \sqrt{\Delta \log d} \tag{1.3}$$

under the condition $\mathbb{E}[X] = \mathbb{E}[Y]$, where $\Delta = \max_{1\le i<j\le d} |\mathrm{Var}(Y_i - Y_j) - \mathrm{Var}(X_i - X_j)|$.

Comparison theorems such as the above and related anticoncentration inequalities are used in statistical inference, machine learning, reliability, signal processing, extreme value theory, random matrix theory, empirical processes, and more. See, for example, [1, 18, 21–23, 27, 29, 30, 34] and references therein. Anticoncentration inequalities in Slepian's comparison theorem provide upper bounds for the modulus of continuity of the distribution function of the maximum or the corresponding density function. This paper is motivated by the recent developments in the central limit theorem and bootstrap theory for the maximum component of a sum of independent random vectors in high dimension, specifically a crucial role of the Gaussian anticoncentration theory in these developments [6, 8, 10, 12, 13, 19, 25].

We present in this paper second- and higher-order anticoncentration inequalities for the Gaussian maxima and some of their implications in the comparison of Gaussian distribution functions. These anticoncentration inequalities provide upper bounds for the sum of the absolute values of the derivatives of a given order for the Gaussian joint distribution function and thus upper bounds for the derivatives of the distribution function of the Gaussian maxima. While the second-order anticoncentration inequalities are used in the development of our error bounds in Slepian's comparison theorem, the third- and higher-order anticoncentration inequalities can be used in studies of the central limit theorem and bootstrap in high dimension depending on the order of expansion in the related Slepian's [31] or Lindeberg's [24] interpolations in such applications.

We present below some error bounds in Slepian's comparison theorem as consequences of our results in Sections 2 and 3.

Theorem 1. *Let $X = (X_1, \dots, X_d)^\top$ and $Y = (Y_1, \dots, Y_d)^\top$ be two Gaussian vectors with $\mathbb{E}[X] = \mathbb{E}[Y] = \mu$. Let $\sigma_i = \sqrt{\mathrm{Var}(X_i) \wedge \mathrm{Var}(Y_i)}$, $\Delta_{i,j} = \{\mathrm{Cov}(Y_i, Y_j) - \mathrm{Cov}(X_i, X_j)\}/(\sigma_i \sigma_j)$, $\Delta_+^{\mathrm{cross}} = \max_{1 \le i < j \le d} (\Delta_{i,j})_+$, and $\Delta^{\mathrm{diag}} = \max_{1 \le i \le d} |\Delta_{i,i}|$. Then, for $d \ge 2$,*

$$\mathbb{P}\Big\{\max_{1\le k\le d} Y_k \le t\Big\} - \mathbb{P}\Big\{\max_{1\le k\le d} X_k \le t\Big\} \le \sqrt{\Delta^*}(4\log d), \tag{1.4}$$

where $\Delta^* = (\Delta_+^{\mathrm{cross}} + \Delta^{\mathrm{diag}})/2 + \max_{1\le i\le j\le d} |\Delta_{i,j}|/(2\log d)$. Moreover, for $d \ge 2$,

$$\begin{aligned} &\mathbb{P}\Big\{\max_{1\le k\le d} Y_k \le t\Big\} - \mathbb{P}\Big\{\max_{1\le k\le d} X_k \le t\Big\} \\ &\quad \le \bigg(2\frac{\Delta_+^{\mathrm{cross}} \vee \Delta^{\mathrm{diag}}}{1-\rho^*} + \frac{\Delta^{\mathrm{diag}}}{2}\bigg) \min\big\{2\log d, \big(\nu_*(t)+1\big)^2\big\}, \end{aligned} \tag{1.5}$$

where $\rho^* = \max_{i<j\le d} |\mathrm{Corr}(X_i, X_j)| \vee |\mathrm{Corr}(Y_i, Y_j)|$ and $\nu_*(t) = 1 \vee \max_{i\le d} |t - \mu_i|/\sigma_i$.

Theorem 1 is proved in Section 3. In Theorem 1, (1.4) is a sharper and more explicit version of Corollary 5.1 of [9]. Under the conditions for (1.1), $\Delta_+^{\mathrm{cross}}(s) = \Delta^{\mathrm{diag}}(s) = 0$ in (1.5), so Theorem 1 contains Slepian's inequality as a special case. Inequality (1.5) improves upon (1.4) when $\sqrt{\Delta^*}/(1-\rho^*)$ is small. Although quantities of different order of smoothness are concerned, the error bounds in Theorem 1 are of a similar form to that of (1.3).

The rest of the paper is organized as follows. We present second-order anticoncentration inequalities in Section 2, comparison theorems for the Gaussian joint distribution functions in Section 3, and higher-order anticoncentration inequalities in Section 4.

We use the following notation to shorten mathematical expressions in the rest of the paper. For positive integers $m < d$, $[d] = \{1, \dots, d\}$, $i_{1:m} = (i_1, \dots, i_m)$, $[d]^m = \{i_{1:m} : i_j \in [d]\ \forall\, j \in [m]\}$, $[d]^m_{\ne} = \{i_{1:m} \in [d]^m : i_j \ne i_k\ \forall\, j \ne k\}$, $[d]^m_{<} = \{i_{1:m} \in [d]^m : i_1 < \cdots < i_m\}$, $[d]_{i_{1:m}} = \{k \in [d] : k \ne i_j\ \forall j \in [m]\}$, and $[d]^2_{i_{1:m},\ne} = \{(j,k) \in [d]^2_{\ne} : j \in [d]_{i_{1:m}}, k \in [d]_{i_{1:m}}\}$. As usual, we denote by $\varphi(t)$ and $\Phi(t)$, respectively, the $N(0,1)$ density and distribution functions, $\|\cdot\|_2$ the Euclidean norm, $\|f\|_{L_\infty} = \sup_{x\in\mathbb{R}^d} |f(x)|$, $a \wedge b = \min(a,b)$, $a \vee b = \max(a,b)$, and $x_+ = \max(x, 0)$.

2. ANTICONCENTRATION INEQUALITIES FOR GAUSSIAN MAXIMA

Let $X = (X_1, \dots, X_d)^\top$ be a multivariate Gaussian vector with a joint distribution function

$$G(x) = G(x_1, \dots, x_d) = \mathbb{P}\big\{X_i \le x_i\ \forall i \in [d]\big\}. \tag{2.1}$$

Let $X_{\max} = \max_{i\in[d]} X_i$ and denote the distribution function of the maximum by

$$G_{\max}(t) = \mathbb{P}\{X_{\max} \le t\} = G(t, \dots, t). \tag{2.2}$$

While concentration inequalities provide upper bounds for the deviation of $X_{\max}$ from its center, e.g., the median, anticoncentration inequalities bound

$$\mathbb{P}\{a < X_{\max} \le a + \varepsilon\} = G_{\max}(a+\varepsilon) - G_{\max}(a)$$

or the density of $X_{\max}$ from the above.

Among existing results on the anticoncentration of $X_{\max}$, Nazarov's [26] inequality,

$$G_{\max}(a+\varepsilon) - G_{\max}(a) \le \varepsilon \frac{2+\sqrt{2\log d}}{\min_{i\in[d]}\sqrt{\mathrm{Var}(X_i)}}, \qquad \forall \varepsilon > 0, \tag{2.3}$$

has found important applications in statistics and machine learning, including bootstrap and central limit theorem [6,8,12,13,19,35]. In terms of the joint distribution G, Nazarov's inequality can be written as an ℓ_1-bound for the gradient of G,

$$\frac{d}{dt}G_{\max}(t) = \sum_{i=1}^{d} \frac{\partial}{\partial x_i} G(x)|_{x_i=t,\forall i\in[d]} \le \frac{2+\sqrt{2\log d}}{\min_{i\in[d]}\sqrt{\mathrm{Var}(X_i)}}. \tag{2.4}$$

In our development of comparison theorems for Gaussian maxima, the second derivative

$$\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}\mathbb{P}\Big\{\max_{k\in[d]}(X_k - x_k) \le t\Big\}$$

is involved in Slepian's interpolation. As t can be absorbed into x_k, what we need is a proper upper bound for the second derivative of the distribution function G,

$$G_{i,j}(x) = \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}\mathbb{P}\big\{X_i \le x_i \ \forall i \in [d]\big\}.$$

In fact, a weighted ℓ_1-norm of $G_{i,j}(x)$ is used in our analysis. Such bounds for the Hessian of $G(x)$ can be viewed as second-order anticoncentration inequalities.

For a standard Gaussian vector $Z = (Z_1, \dots, Z_d)^\top$ with $\mathbb{E}[Z] = 0$ and $\mathbb{E}[ZZ^\top] = I_d$, [3] proved the following anticoncentration inequality of general order:

$$\sup_x \sum_{(i_1,\dots,i_m)\in[d]^m} \left| \frac{\partial^m \mathbb{P}\{Z_k \le x_k \ \forall k \in [d]\}}{\partial x_{i_1}\cdots\partial x_{i_m}} \right| \le C_m(\log d)^{m/2}$$

for some constant C_m depending on m only. Further development of such results and their applications can be found in [2,4,10,11,14,17,20,25,28,36,37] among others. In particular, for applications to Gaussian and bootstrap approximation of the maxima of sums of independent random vectors, the Gaussian vector X was assumed to have a nonsingular covariance matrix Σ and the transformation $Z = \Sigma^{-1/2}X$ was taken to study the anticoncentration of $X_{\max}$ [2,10,14]. The resulting anticoncentration inequality can be written as

$$\sup_x \sum_{(i_1,\dots,i_m)\in[d]^m} \left| \frac{\partial^m G(x)}{\partial x_{i_1}\cdots\partial x_{i_m}} \right| \le \frac{C_m(\log d)^{m/2}}{\lambda_{\min}^{m/2}(\Sigma)}, \tag{2.5}$$

where $\lambda_{\min}$ stands for the smallest eigenvalue. However, the dependence of (2.5) on the smallest eigenvalue is restrictive. We provide below second-order and in Section 4 higher-order anticoncentration inequalities which replace $\lambda_{\min}(\Sigma)$ in (2.5) by the minimum of the eigenvalues of diagonal blocks of dimension m. Such results can be viewed as extensions of Nazarov's inequality (2.4) to higher order.

Before we present the second-order anticoncentration inequality, we give a variation of Nazarov's inequality to explain our approach and write a short proof of it as a road map of the proof in higher order.

Theorem 2. *Let $G(x)$ be the joint distribution function (2.1) of a Gaussian vector $(X_1, \dots, X_d)^\top$ with $X_i \sim N(\mu_i, \sigma_i^2)$. Let $G_i(x) = (\partial/\partial x_i)G(x)$. Let $h(t)$ be a function and $t_0 \geq 0$ such that $h(t)\varphi(t) \leq h(t_0)\varphi(t_0)$ for $t \leq t_0$, $h(t) \wedge h'(t) \geq 0$ for $t \geq t_0$, and $th(t) - h'(t)$ is nonnegative and increasing in $[t_0, \infty)$. Let $a_1^* = t_0 \vee \max_{i\in[d]}(x_i - \mu_i)/\sigma_i$. Then,*

$$\sum_{i=1}^{d} \sigma_i G_i(x) h\big((x_i - \mu_i)/\sigma_i\big) \leq \min\big\{h\big(a_1^*\big)\big(a_1^* + 1 \wedge \big(1/a_1^*\big)\big), h(t_0 \vee \sqrt{2\log d})\sqrt{2\log d}\big\}, \quad d \geq 2. \tag{2.6}$$

In particular, for $\beta \geq 0$, (2.6) holds for $h(t) = |t|^\beta$ with $t_0 = \sqrt{\beta}$.

For $h(t) = 1$ and $d \geq 2$, (2.6) slightly improves Nazarov's inequality (2.4). Inequality (2.6) with $h(t) = |t|^{m-1}$ is useful in bounding the mth order derivative of $G(x)$. The following corollary demonstrates another way of utilizing the choice $h(\cdot)$ in Theorem 2.

Corollary 1. *Let $d \geq 2$. If $\sigma_i \geq \underline{\sigma} > 0$, then*

$$\frac{d}{dt}\mathbb{P}\Big\{\max_{i\in[d]} X_i \leq t\Big\} \leq \frac{\sqrt{2\log d}}{\underline{\sigma}}.$$

If $|t - \mu_i| \geq a$ with a certain $a > 0$, then

$$\frac{d}{dt}\mathbb{P}\Big\{\max_{i\in[d]} X_i \leq t\Big\} \leq (2/a)\log d.$$

Compared with existing literature, Corollary 1 provides an alternative bound to deal with high heteroskedasticity. The second bound in the corollary follows from (2.6) with $h(t) = |t|/a$ as $G_i(x) \leq |(x_i - \mu_i)/\sigma_i|\sigma_i G_i(x)/a$. Typically, for $\mathbb{E}[X_i] = 0$, the magnitude of $\mathbb{E}[\max_{i\in[d]} X_i]$ is of the order $\overline{\sigma}\sqrt{\log d}$ for some $\overline{\sigma}$ representing the average of σ_i or larger and the probability outside a small neighborhood of $\mathbb{E}[\max_{i\in[d]} X_i]$ is very small due to Gaussian concentration, so that the most interesting application of (2.4) is for $t \geq a$ with $a \asymp \overline{\sigma}\sqrt{\log d}$. In such applications, Corollary 1 replaces $\min_{i\in[d]} \sigma_i$ in (2.4) with a quantity of the order $\overline{\sigma}$. This is related to a variation of (2.3) in [7] where the $\sqrt{\log d}$ rate is replaced by $O(\mathbb{E}[\max_{k\leq d} X_k/\sigma_k] + \sqrt{1 \vee \log(\min_i \sigma_i/\varepsilon)})$ for centered X. Another variation of (2.4) can be found in [13] where the upper bound is

$$1/\sigma_{(1)} + \max_{1\leq j\leq d}(1 + \sqrt{2\log j})/\sigma_{(j)}$$

where $\sigma_{(1)} \leq \cdots \leq \sigma_{(d)}$ are the ordered values of $\sigma_1, \dots, \sigma_d$. This variation of Nazarov's inequality is extended to higher orders in Section 4.

Proof of Theorem 2. Let $\phi_i(t) = \varphi((t - \mu_i)/\sigma_i)/\sigma_i$ be the density of X_i, $X_i' = (X_i - x_i)/\sigma_i$ and $\rho_{i,j} = \mathrm{Corr}(X_i, X_j)$. Because X_i is independent of $X_k' - \rho_{i,k}X_i'$,

$$\begin{aligned} G_i(x) &= \mathbb{P}\big\{X_k \leq x_k, \forall k \in [d]_i | X_i = x_i\big\}\phi_i(x_i) \\ &= \mathbb{P}\big\{X_k' \leq \rho_{i,k}X_i', \forall k \in [d]_i\big\}\phi_i(x_i). \end{aligned} \tag{2.7}$$

Let $\pi_i = \mathbb{P}\{X_i' > 0\}G_i(x)/\phi_i(x_i)$, $\nu_i(x) = (x_i - \mu_i)/\sigma_i$ and $R_i = \sum_{k\in[d]} I\{X_k' \ge X_i'\}$ be the rank of X_i'. Due to the independence of X_i' and $\{X_k' - \rho_{i,k}X_i', k \in [d]_i\}$ and the fact that $\{X_k' \le \rho_{i,k}X_i', k \in [d]_i, X_i' > 0\} \subseteq \{R_i = 1\}$,

$$\pi_i = \mathbb{P}\{X_k' \le \rho_{i,k}X_i', \forall k \in [d]_i\}\Phi(-\nu_i(x)) \le \mathbb{P}\{R_i = 1\}, \tag{2.8}$$

so that $\sum_{i\in[d]} \pi_i \le 1$. Let $\psi_1(t) = \Phi(-t)/\varphi(t)$. We have $h(\nu_i(x))/\psi_1(\nu_i(x)) \le h(a_1^*)/\psi_1(a_1^*)$ because $a_1^* \ge t_0$, $h(t)/\psi_1(t) \le h(t_0)/\psi_1(t_0)$ for $t \le t_0$ and $h(t)/\psi_1(t)$ is nondecreasing for $t \ge t_0$. Because $\sigma_i G_i(x) = \pi_i/\psi_1(\nu_i(x))$ by (2.7) and (2.8), the first upper bound in (2.6) follows from $1/\psi_1(t) \le t + \psi_1(t)$ and $h(t) \wedge h'(t) \ge 0$ for $t \ge t_0$. For the second upper bound in (2.6), (2.7) and (2.8) yield

$$\begin{aligned}
&\sum_{i=1}^d \sigma_i G_i(x) h(\nu_i(x)) \\
&\quad \le \max_{\nu_i,\pi_i,i\in[d]} \left\{ \sum_{\nu_i \le t_0} \frac{h(t_0)\varphi(t_0)\pi_i}{\Phi(-\nu_i)} + \sum_{\nu_1 > t_0} \frac{h(\nu_i)\pi_i}{\psi_1(\nu_i)} : \sum_{i=1}^d \pi_i \le 1, 0 \le \pi_i \le \Phi(-\nu_i) \right\}. \\
&\quad = \max_{\nu_i \ge t_0,\pi_i,i\in[d]} \left\{ \sum_{i=1}^d h(\nu_i)\varphi(\nu_i) : \sum_{i=1}^d \Phi(-\nu_i) \le 1 \right\}
\end{aligned} \tag{2.9}$$

due to $h(t)\varphi(t) \le h(t_0)\varphi(t_0)$ in $(\infty, t_0]$ and the monotonicity of $\Phi(t)$ in $\mathbb{R}$ and $h(t)/\psi_1(t)$ in $[t_0, \infty)$. The global maximum on the right-hand side of (2.9) must be attained when $\nu_i h(\nu_i) - h'(\nu_i) = \lambda$ for all i with a Lagrange multiplier λ. As $th(t) - h'(t) = \lambda$ has one solution in $[t_0, \infty)$, the global maximum is attained at $\nu_i = t$ for all $i \in [d]$ and some $t \ge t_0$. As $(d/dt)\{h(t)\varphi(t)\} \le 0$ for $t \ge t_0$, the maximum is attained at $t = t_0 \vee t_1$ and given by $dh(t_0 \vee t_1)\varphi(t_0 \vee t_1)$, where t_1 is the solution of $\Phi(-t_1) = 1/d$. This gives the second upper bound in (2.6) because $t_1 \le \sqrt{2\log d}$ and $d\varphi(t_1) = 1/\psi_1(t_1) \le \sqrt{2\log d}$ for $d \ge 2$. ∎

To extend Theorem 2 to the second order, we need to define certain quantities $\alpha_{i,j}$ as an extension of the weights σ_i. Let $\Sigma^{i,j}$ and $\phi_{i,j}(\cdot)$ be respectively the covariance matrix and joint density of $(X_i, X_j)^\top$. As $1/\sigma_i = \max_t \sqrt{2\pi}\phi_i(t)$, $\alpha_{i,j}$ is expected to involve $|\det(\Sigma^{i,j})|^{1/2}$ as the Jacobian in the denominator of $\phi_{i,j}(\cdot)$. However, $\alpha_{i,j}$ also involves a certain threshold level t_i for a two-dimensional extension of (2.8). Let $\rho_{i,j} = \mathrm{Corr}(X_i, X_j)$. The threshold level t_i is defined as

$$t_i = \min\left\{\sqrt{(1-\rho_{i,j})/(1+\rho_{i,j})} : j \in [d]_i\right\}, \tag{2.10}$$

which can be viewed as the tangent of the minimum half-angle between standardized X_i and X_j in $L_2(\mathbb{P})$. Let $Y_i = (X_i - \mu_i)/\sigma_i$ be the standardized X_i, $\theta_{i,j} = \arccos(\rho_{i,j}) \in [0, \pi]$ be the $L_2(\mathbb{P})$ angle between Y_i and Y_j, and $\theta_{i,\min} = \min\{\theta_{i,j} : j \in [d]_i\}$ be the angle between Y_i and its nearest neighbor. The threshold level in (2.10) can be written as

$$t_i = \tan(\theta_{i,\min}/2).$$

The quantity $\alpha_{i,j}$ is then defined as

$$\begin{aligned}\alpha_{i,j} &= 2\tan(\theta_{i,\min}/4)\big|\det\big(\Sigma^{i,j}\big)\big|^{1/2} \\ &= 2\tan\big(\arctan(t_i)/2\big)\sigma_i\sigma_j\big(1-\rho_{i,j}^2\big)^{1/2}\end{aligned} \tag{2.11}$$

with $\tan(\theta_{i,\min}/4) \in [0, 1]$ and t_i as in (2.10). We note that $2\tan(\theta_{i,\min}/4) \approx t_i$ when t_i is small. We also consider quantities

$$\tilde{\nu}_{i,j} = \tilde{\nu}_{i,j}(x) = \big(\nu_i^2 + \nu_{j|i}^2\big)^{1/2} \wedge \big(\nu_i + t_i(\nu_{j|i})_+\big)_+ \tag{2.12}$$

as signed versions of

$$\nu_{i,j} = \nu_{i,j}(x) = \big(\nu_i^2 + \nu_{j|i}^2\big)^{1/2}, \tag{2.13}$$

where $\nu_i = \nu_i(x) = (x_i - \mu_i)/\sigma_i$ and $\nu_{j|i} = \nu_{j|i}(x) = (\nu_j - \rho_{i,j}\nu_i)/(1-\rho_{i,j}^2)^{1/2}$. We are now ready to state a second order Gaussian anticoncentration theorem.

Theorem 3. *Let $d \geq 2$ and $X = (X_1, \dots, X_d)^\top$ be a Gaussian vector with a joint distribution function $G(x)$. Let $G_{i,j}(x) = (\partial/\partial x_i)(\partial/\partial x_j)G(x)$, $\alpha_{i,j}$ as in (2.11), $\rho_{i,j} = \mathrm{Corr}(X_i, X_j)$, and $a_2^* = a_2^*(x) = \sqrt{2} \vee \max_{i,j} \tilde{\nu}_{i,j}(x)$ with $\tilde{\nu}_{i,j}(x)$ as in (2.12). Then,*

$$\sum_{(i,j)\in[d]_{\neq}^2} \alpha_{i,j} G_{i,j}(x) \leq \min\big\{(1/\pi) \vee \big(2\log\big(d(d-1)/2\big)\big), \big(a_2^* + \sqrt{2}\big)^2\big\}. \tag{2.14}$$

Moreover, with $a_1^* = 1 \vee \max_{i\in[d]}(x_i - \mu_i)/\sigma_i$,

$$\sum_{i=1}^d \Big|\sigma_i^2 G_{i,i}(x) + \sum_{j\in[d]_i} \rho_{i,j}\sigma_i\sigma_j G_{i,j}(x)\Big| \leq \min\big\{2\log d, \big(a_1^*\big)^2 + 1\big\}. \tag{2.15}$$

Before we move ahead to proving Theorem 3, we state in the following corollary a scaled ℓ_1-bound for the Hessian of the joint distribution function $G(x)$ as a direct consequence of the theorem using $\tan(\theta_{i,\min}/4) \geq \sqrt{(1-\max_{k\neq i}\rho_{i,k})/8}$ in (2.11).

Corollary 2. *With $\sigma_i = \mathrm{Var}^{1/2}(X_i)$ and $\rho_{i,j} = \mathrm{Corr}(X_i, X_j)$,*

$$\sum_{i=1}^d\sum_{j=1}^d \sigma_i\sigma_j\big|G_{i,j}(x)\big| \leq \max_{(i,j,k)\in[d]_{\neq}^3} \frac{8\log d}{\sqrt{(1-|\rho_{i,j}|)(1-\rho_{j,k})}} + 2\log d, \quad d \geq 2.$$

Proof of Theorem 3. To prove (2.14), we define

$$X_i' = \frac{X_i - x_i}{\sigma_i}, \quad X_{j|i}' = \frac{X_j' - \rho_{i,j}X_i'}{(1-\rho_{i,j}^2)^{1/2}}, \quad \rho_{(j,k)|i} = \mathrm{Corr}\big(X_j', X_k' | X_i'\big). \tag{2.16}$$

Let $\phi_{i,j}(\cdot)$ be the joint density of $(X_i, X_j)^\top$. As in (2.7), it holds for all $(i,j) \in [d]_{\neq}^2$ that

$$\begin{aligned}G_{i,j}(x) &= \mathbb{P}\big\{X_k' < 0,\ \forall k \in [d]_{i,j} | X_i' = X_j' = 0\big\}\phi_{i,j}(x_i, x_j) \\ &= \mathbb{P}\big\{X_{k|i}' - \rho_{(j,k)|i}X_{j|i}' < 0,\ \forall k \in [d]_{i,j}\big\}\phi_{i,j}(x_i, x_j).\end{aligned} \tag{2.17}$$

For the second step above, we note that $X'_{k|i} - \rho_{(j,k)|i} X'_{j|i}$ is independent of $(X'_i, X'_j)^\top$. Similar to the proof leading to (2.9), we set $\mathscr{C}_{i,j} = \{0 < X'_{j|i} < t_i X'_i\}$ with the threshold level t_i in (2.10), and define

$$\pi_{i,j} = \mathbb{P}\{\mathscr{C}_{i,j}\} G_{i,j}(x)/\phi_{i,j}(x_i, x_j). \tag{2.18}$$

Let $R_i = \sum_{j=1}^d I\{X'_j \geq X'_i\}$ and $R_{j|i} = \sum_{k\in[d]_i} I\{X'_{k|i} \geq X'_{j|i}\}$ be respectively the marginal and conditional ranks of X'_i and $X'_{j|i}$ in (2.16). By the definition of t_i, $t_i \leq \sqrt{(1-\rho_{i,k})/(1+\rho_{i,k})}$, so that $\rho_{i,k} + (1-\rho_{i,k}^2)^{1/2} t_i \leq 1$ for all $k \in [d]_i$. It follows that

$$\begin{aligned}
\pi_{i,j} &= \mathbb{P}\big\{X'_{k|i} - \rho_{(j,k)|i} X'_{j|i} < 0 \ \forall k \in [d]_{i,j}\big\}\mathbb{P}\{\mathscr{C}_{i,j}\} \\
&= \mathbb{P}\big\{X'_{k|i} < \rho_{(j,k)|i} X'_{j|i}, \forall k \in [d]_{i,j}, 0 \leq X'_{j|i} < t_i X'_i, X'_i > 0\big\} \\
&\leq \mathbb{P}\big\{R_{j|i} = 1,\ X'_{k|i} < t_i X'_i,\ \forall k \in [d]_i, X'_i > 0\big\} \\
&= \mathbb{P}\big\{R_{j|i} = 1,\ X'_k < \big(\rho_{i,k} + \big(1-\rho_{i,k}^2\big)^{1/2} t_i\big) X'_i \leq X'_i,\ \forall k \in [d]_i\big\} \\
&\leq \mathbb{P}\{R_{j|i} = 1, R_i = 1\}.
\end{aligned}$$

Consequently,

$$\sum_{(i,j)\in[d]^2_{\neq}} \pi_{i,j} \leq 1. \tag{2.19}$$

We still need a lower bound for $\mathbb{P}\{\mathscr{C}_{i,j}\}$ to use (2.19). To this end, we prove

$$\begin{aligned}
\mathbb{P}\{\mathscr{C}_{i,j}\} &= \pi_{i,j}\phi_{i,j}(x_i, x_j)/G_{i,j}(x) \\
&\geq 2\tan(\theta_{i,\min}/4)\varphi(\nu_{i,j})\varphi(0)\psi_2(\tilde{\nu}_{i,j}) \\
&= \alpha_{i,j}\phi_{i,j}(x_i, x_j)\psi_2(\tilde{\nu}_{i,j}),
\end{aligned} \tag{2.20}$$

where $\nu_{i,j} = \nu_{i,j}(x)$ are as in (2.13), $\tilde{\nu}_{i,j} = \tilde{\nu}_{i,j}(x) = \min\{\nu_{i,j}, (\nu_i + t_i(\nu_{j|i})_+)_+\}$ are as in (2.12), $\alpha_{i,j}$ and $\theta_{i,\min}$ are as in (2.11), and

$$\psi_2(t) = \int_0^\infty \int_0^{y_1} e^{-y_2^2/2 - t y_1 - y_1^2/2} dy_2 dy_1. \tag{2.21}$$

Moreover, with $\psi_1(t) = \Phi(-t)/\varphi(t)$ as in (2.9), we prove that for all $t \geq 0$,

$$1/\psi_2(t) \leq 1/\psi_1^2(t) + 1 + 2/\big(1 + \psi_1^2(t)\big). \tag{2.22}$$

The first equality in (2.20) is from the definition of $\pi_{i,j}$ in (2.18), and the last follows from $\varphi(\nu_i)\varphi(\nu_{j|i}) = |\det(\Sigma^{i,j})|^{1/2}\phi_{i,j}(x_i, x_j)$ and the definition of $\alpha_{i,j}$ in (2.11). We note that $\nu_{i,j} = (\nu_i^2 + \nu_{j|i}^2)^{1/2}$ and the variables $X'_i \sim N(-\nu_i, 1)$ and $X'_{j|i} \sim N(-\nu_{j|i}, 1)$ are independent by (2.16). It follows that

$$\mathbb{P}\{\mathscr{C}_{i,j}\} = \int_0^\infty \int_0^{t_i y_1} \varphi(y_1 + \nu_i)\varphi(y_2 + \nu_{j|i}) dy_2 dy_1 \tag{2.23}$$

with $t_i = \tan(\theta_{i,\min}/2)$. Given $\nu_{i,j} = (\nu_i^2 + \nu_{j|i}^2)^{1/2}$, the above integral is minimized when $\nu_i \wedge \nu_{j|i} \geq 0$ and $\nu_{j|i}/\nu_i = \tan(\theta_{i,\min}/4)$. Thus, after proper rotation

$$\mathbb{P}\{\mathscr{C}_{i,j}\} \geq \int_0^\infty \int_{|y_2| \leq \tan(\theta_{i,\min}/4) y_1} \varphi(y_1 - \nu_{i,j})\varphi(y_2) dy_2 dy_1,$$

which implies the inequality in (2.20) for $\nu_{i,j} = \tilde{\nu}_{i,j}$. For $\nu_{i,j} > \tilde{\nu}_{i,j}$ and $0 < y_2 \le t_i y_1$, $\nu_i y_1 + \nu_{j|i} y_2 \le \nu_i y_1 + t_i(\nu_{j|i})_+ y_1 \le \tilde{\nu}_{i,j} y_1$, so that by (2.23)

$$\mathbb{P}\{\mathscr{C}_{i,j}\} \ge \frac{\varphi(\nu_{i,j})}{\sqrt{2\pi}} \int_0^\infty \int_0^{t_i y_1} e^{-y_1^2/2 - y_2^2/2 - \tilde{\nu}_{i,j} y_1} dy_2 dy_1,$$

which again implies the inequality in (2.20). For (2.22), we note that by (2.21)

$$\psi_2(t) = \int_0^\infty \{\psi_1\big((t+y_1)/\sqrt{2}\big)/\sqrt{2}\} e^{-ty_1 - y_1^2/2} dy_1.$$

As in Lemma 9 of **[13]**, $1/(t + \psi_1(t)) < \psi_1(t) < 1/t$, so that

$$\int_0^\infty y_1 e^{-ty_1 - y_1^2/2} dy_1 = 1 - t\psi_1(t) \le \psi_1^2(t).$$

As $\psi_1(\cdot)$ is convex and decreasing in $[0, \infty)$, an application of Jensen's inequality yields $\psi_2(t) \ge \{\psi_1(t)/\sqrt{2}\}\psi_1((t + \psi_1(t))/\sqrt{2})$. Thus, as $(1/t)/(1 + 1/t^2) < \psi_1(t) < 1/t$,

$$\psi_2(t) \ge \frac{\psi_1(t)}{\sqrt{2}} \psi_1\left(\frac{1 + \psi_1^2(t)}{\sqrt{2}\psi_1(t)}\right) \ge \frac{\psi_1^2(t)/(1 + \psi_1^2(t))}{1 + 2\psi_1^2(t)/(1 + \psi_1^2(t))^2},$$

which gives (2.22).

Let $\pi'_{i,j} = \alpha_{i,j} G_{i,j}(x)\psi_2(\tilde{\nu}_{i,j})$. It follows from (2.20) that $\pi'_{i,j} \le \pi_{i,j}$. By (2.11) and (2.17),

$$\pi'_{i,j} \le 2\big|\det(\Sigma^{i,j})\big|^{1/2} \phi_{i,j}(x_i, x_j)\psi_2(\tilde{\nu}_{i,j}) = \sqrt{2/\pi}\varphi(\nu_{i,j})\psi_2(\tilde{\nu}_{i,j}). \tag{2.24}$$

This gives (2.14) for $d = 2$ as $\alpha_{1,2} G_{1,2}(x) \le 1/\pi$. By (2.19) and (2.22),

$$\sum_{(i,j)\in[d]^2_{\neq}} \alpha_{i,j} G_{i,j}(x) = \sum_{(i,j)\in[d]^2_{\neq}} \frac{\pi'_{i,j}}{\psi_2(\tilde{\nu}_{i,j})} \le \big(a_2^* + \sqrt{2}\big)^2$$

due to $a_2^* = \sqrt{2} \vee \max_{(i,j)\in[d]^2_{\neq}} \tilde{\nu}_{i,j}$ and $1/\psi_1(t) \le t + 1/t$. In general, (2.19) and (2.24) yield

$$\begin{aligned}
&\sum_{(i,j)\in[d]^2_{\neq}} \alpha_{i,j} G_{i,j}(x) \\
&\le \max_{\nu_{i,j}\ge 0, \pi'_{i,j}} \left\{ \sum_{(i,j)\in[d]^2_{\neq}} \frac{\pi'_{i,j}}{\psi_2(\nu_{i,j})} : \sum_{(i,j)\in[d]^2_{\neq}} \pi'_{i,j} \le 1, \pi'_{i,j} \le \sqrt{2/\pi}\varphi(\nu_{i,j})\psi_2(\nu_{i,j}) \right\} \\
&= \max_{\nu_{i,j}\ge 0} \left\{ \sum_{(i,j)\in[d]^2_{\neq}} \sqrt{2/\pi}\varphi(\nu_{i,j}) : \sum_{(i,j)\in[d]^2_{\neq}} \sqrt{2/\pi}\varphi(\nu_{i,j})\psi_2(\nu_{i,j}) \le 1 \right\}
\end{aligned} \tag{2.25}$$

because $\psi_2(t)$ and $\varphi(t)$ are both decreasing in $[0, \infty)$. Let $d_2 = d(d-1)/2$. By (2.21), $\psi_2(t) - \psi_2'(t)/t$ is decreasing in t in $[0, \infty)$, so that the optimization problem is solved by $\nu_{i,j} = t_2$ with a Lagrange multiplier, where t_2 is the solution of $\sqrt{2/\pi}\varphi(t_2)\psi_2(t_2) = 1/(2d_2)$. For $d \ge 3$ and $t = \sqrt{(2\log(2d_2/(2\pi\log d_2)))_+}$, we have $1/\psi_2(t) \le 2\log d_2$ via (2.22). Thus, the right-hand side of (2.25) is no greater than $2\log d_2$.

Finally, it follows from (2.16) and (2.7) that

$$G_{i,i}(x) = -G_i(x)\nu_i(x)/\sigma_i - \sum_{j\in[d]_i} G_{i,j}(x)\rho_{i,j}\sigma_j/\sigma_i \tag{2.26}$$

with $\nu_i(x) = (x_i - \mu_i)/\sigma_i$, so that (2.15) follows from Theorem 2 with $h(t) = |t|$. ∎

3. COMPARISON OF GAUSSIAN DISTRIBUTION FUNCTIONS

The Gaussian anticoncentration theorem in Section 2 yields the following error bounds in the comparison of Gaussian joint distribution functions.

Let $X = (X_1, \ldots, X_d)^\top$ and $Y = (Y_1, \ldots, Y_d)^\top$ be two Gaussian vectors with common mean μ and respective covariance matrices Σ^X and Σ^Y and joint distribution functions

$$G^X(x) = \mathbb{P}\{X_k \le x_k \forall k \in [d]\}, \quad G^Y(y) = \mathbb{P}\{Y_k \le y_k \forall k \in [d]\}. \tag{3.1}$$

For $0 \le s \le 1$, let $\Sigma_{i,j}(s)$ be the elements of $\Sigma(s) = (1-s)\Sigma^X + s\Sigma^Y$,

$$\nu_i(x;s) = (x_i - \mu_i)/\sqrt{\Sigma_{i,i}(s)}, \tag{3.2}$$

$$\Delta_{i,j}(s) = \frac{\Sigma^Y_{i,j} - \Sigma^X_{i,j}}{\sqrt{\Sigma_{i,i}(s)\Sigma_{j,j}(s)}}, \quad \rho_{i,j}(s) = \frac{\Sigma_{i,j}(s)}{\sqrt{\Sigma_{i,i}(s)\Sigma_{j,j}(s)}}, \tag{3.3}$$

and

$$\Delta_{i,j,\pm}(s) = \max_{k \ne i, \ell \ne j} \frac{(2\Delta_{i,j}(s))_\pm \vee |\Delta_{i,i}(s) + \Delta_{j,j}(s)|}{\sqrt{(1-|\rho_{i,j}(s)|)}(\sqrt{1-\rho_{i,k}(s)} + \sqrt{1-\rho_{j,\ell}(s)})}. \tag{3.4}$$

Theorem 4. *Let $G^X(x)$ and $G^Y(y)$ be as in (3.1), $\nu_*(x;s) = 1 \vee \max_{i]\in[d]} |\nu_i(x;s)|$, $\Delta^*_+(s) = \max_{(i,j)\in[d]^2_{\ne}} \Delta_{i,j,+}(s)$ and $\Delta^{\mathrm{diag}}(s) = \max_{i\in[d]} |\Delta_{i,i}(s)|$, where $\nu_i(x;s)$, $\Delta_{i,i}(s)$, and $\Delta_{i,j,\pm}(s)$ are as in (3.2). (3.3), and (3.4), respectively. Then, for $d \ge 2$,*

$$G^Y(x) - G^X(x) \le \int_0^1 \big(2\Delta^*_+(s) + \Delta^{\mathrm{diag}}(s)/2\big) \min\big\{2\log d, \big(\nu_*(x,s)+1\big)^2\big\} ds. \tag{3.5}$$

Because $G^X_{\max}(t) = G^X(t, \ldots, t)$ and $G^Y_{\max}(t) = G^Y(t, \ldots, t)$, Theorem 1 is an immediate consequence of Theorem 4. Conversely, as t can be absorbed into the mean, Theorem 1 is a simplified version of Theorem 4.

Assume, without loss of generality, that X and Y are independent as the theorem does not involve the joint distribution of X and Y. With $\mu = \mathbb{E}[X]$, write $X(s) = \sqrt{1-s}(X-\mu) + \sqrt{s}(Y-\mu) + \mu$, $s \in [0,1]$, as Slepian's interpolation and

$$\phi(x;s) = (2\pi)^{-d} \int_{\mathbb{R}^d} \exp\big[\sqrt{-1}(\mu - x)^\top u - u^\top\big((1-s)\Sigma^X + s\Sigma^Y\big)u/2\big] du$$

as the joint density of $X(s)$. Slepian's inequality was proved by passing the differentiation of $\mathbb{E}[f(X(s))]$ to twice differentiation of f through the above formula,

$$\frac{d}{ds}\mathbb{E}\big[f\big(X(s)\big)\big] = \frac{1}{2}\sum_{i=1}^d \sum_{j=1}^d (\Sigma^Y_{i,j} - \Sigma^X_{i,j}) \int \bigg(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\bigg)\phi(x;s)dx, \tag{3.6}$$

provided the twice differentiability of $f(x)$. However, for comparison of distribution functions, this is not feasible as f is an indicator function. Instead, with $y = (y_1, \ldots, y_d)^\top$, we may exchange the differentiation and integration in (3.6) and write

$$\frac{d}{ds}\mathbb{E}\big[f\big(X(s)\big)\big] = \frac{1}{2}\sum_{i=1}^d \sum_{j=1}^d (\Sigma^Y_{i,j} - \Sigma^X_{i,j}) \frac{\partial^2 F(y;s)}{\partial y_i \partial y_j}\bigg|_{y=0} \tag{3.7}$$

with $F(y;s) = \int f(x)\phi(x+y;s)dx = \mathbb{E}[f(X(s)-y)]$. In the proof of Theorem 4, we directly apply the weighted anticoncentration inequality in Theorem 3 to (3.7).

Proof of Theorem 4. Let $\sigma_i(s) = \Sigma_{i,i}^{1/2}(s)$ and

$$G(x;s) = \mathbb{P}\{X_k(s) \le x_k, k \in [d]\},$$

so that (3.7) becomes

$$(\partial/\partial s)G(x;s) = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}(\Sigma_{i,j}^{Y} - \Sigma_{i,j}^{X})G_{i,j}(x;s), \tag{3.8}$$

where $G_{i,j}(x;s) = (\partial/\partial x_i)(\partial/\partial x_j)G(x;s)$. As in (2.11), let

$$\alpha_{i,j}(s) = 2\sigma_i(s)\sigma_j(s)\big(1-\rho_{i,j}^2(s)\big)^{1/2}\tan\big(\theta_{i,\min}(s)/4\big),$$

with $\theta_{i,\min}(s) = \min\{\arccos(\rho_{i,k})(s), k \in [d]_i\} \in [0,\pi]$. We have

$$\frac{\sigma_i(s)\sigma_j(s)(1+|\rho_{i,j}(s)|)}{\alpha_{i,j}(s)+\alpha_{j,i}(s)} \le \max_{k\ne i,\ell\ne j}\frac{\sqrt{2(1+|\rho_{i,j}(s)|)}}{\sqrt{1-|\rho_{i,j}(s)|}\big(\sqrt{(1-\rho_{i,k})}+\sqrt{(1-\rho_{j,\ell})}\big)}$$

due to $\tan(\theta_{i,\min}/4) \ge \sqrt{(1-\max_{k\ne i}\rho_{i,k})/8}$. Let $\nu_i(x;s) = (x_i-\mu_i)/\sigma_i(s)$. We use $\sigma_i(s)$ to scale (3.8) and apply (2.26) and Theorem 3 as follows:

$$\begin{aligned}
&(\partial/\partial s)G(x;s)\\
&= \frac{1}{2}\sum_{(i,j)\in[d]_{\ne}^2}\Delta_{i,j}(s)\sigma_i(s)\sigma_j(s)G_{i,j}(x;s) - \frac{1}{2}\sum_{i=1}^{d}\Delta_{i,i}(s)\sigma_i(s)G_i(x;s)\nu_i(x;s)\\
&\quad - \frac{1}{4}\sum_{(i,j)\in[d]_{\ne}^2}\big(\Delta_{i,i}(s)+\Delta_{j,j}(s)\big)\rho_{i,j}(s)\sigma_i(s)\sigma_j(s)G_{i,j}(x;s)\\
&\le \sum_{(i,j)\in[d]_{\ne}^2}\Delta_{i,j,+}(s)\left(\frac{\alpha_{i,j}(s)+\alpha_{j,i}(s)}{2}\right)G_{i,j}(x;s)\\
&\quad + \frac{1}{2}\sum_{i=1}^{d}\big|\Delta_{i,i}(s)\nu_i(x;s)\big|\sigma_i(s)G_i(x;s)\\
&\le \max_{(i,j)\in[d]_{\ne}^2}\Delta_{i,j,+}(s)\{(\sqrt{2}\nu_* + \sqrt{2})^2 \wedge (4\log d)\}\\
&\quad + \max_{i\in[d]}\big|\Delta_{i,i}(s)/2\big|\{(\nu_*+1)^2 \wedge (2\log d)\}.
\end{aligned}$$

This gives (3.5) by integrating over $s \in [0,1]$. ■

In the rest of this section we prove Theorem 1.

Proof of Theorem 1. Let $\mathrm{Err}_t = \mathbb{P}\{\max_{1\le i\le d} Y_i \le t\} - \mathbb{P}\{\max_{1\le i\le d} X_i \le t\}$ and write

$$\mathrm{Err}_t = \mathbb{P}\Big\{\max_{1\le i\le d} Y_i' \le 0\Big\} - \mathbb{P}\Big\{\max_{1\le i\le d} X_i' \le 0\Big\},$$

with $X_i' = (X_i - t)/\sigma_i$ and $Y_i' = (Y_i - t)/\sigma_i$. Let $\varepsilon \ge \varepsilon' > 0$, $\beta = (\log d)/(2\varepsilon')$, $g(x) = \beta^{-1}\log(\sum_{i=1}^{d} e^{\beta x_i})$ for $x = (x_1,\dots,x_d)^\top$, and $f_\varepsilon(t)$ be the nonincreasing function with $f_\varepsilon(\varepsilon) = 0$ and derivative $f'_\varepsilon(t) = -\varepsilon^{-1}(1 - |t|/\varepsilon)_+$. Let $x_{\max} = \max_{1\le i\le d} x_i$. Similar to [5], we approximate $I\{x_{\max} \le 0\}$ by $f_\varepsilon(g(x) - \varepsilon')$. Because $x_{\max} \le g(x) \le x_{\max} + 2\varepsilon'$,

$$\begin{aligned} I\{y_{\max} \le 0\} - I\{x_{\max} \le 0\} - f_\varepsilon\big(g(y) - \varepsilon'\big) + f_\varepsilon\big(g(x) - \varepsilon'\big) \\ \le I\{y_{\max} \le 0\}\big\{1 - f_\varepsilon\big(y_{\max} + \varepsilon'\big)\big\} + I\{x_{\max} > 0\} f_\varepsilon\big(x_{\max} - \varepsilon'\big), \end{aligned}$$

for any x and $y = (y_1, \dots, y_d)^\top$, where $y_{\max} = \max_{1\le i\le d} y_i$. Set $\varepsilon'/\varepsilon = 3/10$. As $\mathrm{Var}(X'_i) \wedge \mathrm{Var}(Y'_i) \ge 1$ and $f(t) + f(-t) = 1$, Corollary 1 provides

$$\begin{aligned} &\mathrm{Err}_t - \mathbb{E}\big[f\big(g(Y)\big)\big] + \mathbb{E}\big[f\big(g(X)\big)\big] \\ &\le \sqrt{2\log d}\left(\int_{-\varepsilon-\varepsilon'}^{0} \big(1 - f_\varepsilon\big(t + \varepsilon'\big)\big)dt + \int_0^{\varepsilon+\varepsilon'} f_\varepsilon\big(t - \varepsilon'\big)dt\right) \\ &= \sqrt{2\log d}\,\big\{2\varepsilon' + \varepsilon\big(1 - \varepsilon'/\varepsilon\big)^3/3\big\} \\ &\le (3\varepsilon/4)\sqrt{2\log d}. \end{aligned} \tag{3.9}$$

The approximation allows us to apply (3.6) to X' and Y'. Let $p_i = p_i(x) = e^{\beta x_i}/\sum_{j=1}^d e^{\beta x_j}$. We have $\partial g(x)/\partial x_i = p_i$ and $\partial p_i/\partial x_j = \beta I_{\{i=j\}} p_i - \beta p_i p_i$. It follows that

$$\begin{aligned} |\mathrm{Err}_t| &\le (3\varepsilon/4)\sqrt{2\log d} + \frac{1}{2}\int_{\mathbb{R}^d} \beta f'_\varepsilon\big(g(x)\big)\sum_{i=1}^d p_i \Delta_{i,i}\phi(x; s)dx \\ &\quad + \frac{1}{2}\int_{\mathbb{R}^d} \big\{f''_\varepsilon\big(g(x)\big) - \beta f'_\varepsilon\big(g(x)\big)\big\}\sum_{i=1}^d\sum_{j=1}^d p_i p_j \Delta_{i,j}\phi(x; s)dx \\ &\le (3\varepsilon/4)\sqrt{2\log d} + \frac{\Delta}{2}\int_{\mathbb{R}^d} \big|f''_\varepsilon\big(g(x)\big)\big|\phi(x; s)dx \\ &\quad + \frac{(\Delta^{\mathrm{diag}} + \Delta^{\mathrm{cross}}_+)\log d}{4\varepsilon'}\int_{\mathbb{R}^d} \big|f'_\varepsilon\big(g(x)\big)\big|\phi(x; s)dx, \end{aligned} \tag{3.10}$$

due to $f'_\varepsilon(t) \le 0$, where $\Delta = \max_{1\le i\le j\le d} |\Delta_{i,j}|$. Similar to (3.9), we have

$$\frac{1}{4\varepsilon'\sqrt{2\log d}}\int_{\mathbb{R}^d} \big|f'_\varepsilon\big(g(x)\big)\big|\phi(x; s)dx \le \frac{1 + 2\varepsilon'/\varepsilon}{4\varepsilon'} = \frac{4}{3\varepsilon}$$

and $\int_{\mathbb{R}^d} |f''_\varepsilon(g(x))|\phi(x; s)dx/(2\sqrt{2\log d}) \le (\varepsilon'/\varepsilon + 1)/\varepsilon \le 4/(3\varepsilon)$. Inserting the above bounds for the integrals into (3.10), we find that

$$\frac{|\mathrm{Err}_t|}{\sqrt{2\log d}} \le \frac{3\varepsilon}{4} + \frac{4}{3\varepsilon}\big\{\big(\Delta^{\mathrm{diag}} + \Delta^{\mathrm{cross}}_+\big)\log d + \Delta\big\}.$$

This gives (1.4) with ε minimizing the right-hand side. Theorem 4 implies (1.5) due to $|\rho_{i,j}(s)| \le \rho^*\{(1 - s)\sqrt{\sigma_i(0)\sigma_j(0)} + s\sqrt{\sigma_i(1)\sigma_j(1)}\}/\{\sigma_i(s)\sigma_j(s)\} \le \rho^*$ in (3.4). ■

4. HIGHER-ORDER ANTICONCENTRATION

In this section we extend Theorem 3 to higher order by developing upper bounds for weighted sums of the absolute values of the derivatives

$$G_{i_1,\dots,i_m}(x) = \frac{\partial^m \mathbb{P}\{X_k \le x_k\ \forall k \in [d]\}}{\partial x_{i_1}\cdots\partial x_{i_m}} \tag{4.1}$$

for Gaussian vectors $X = (X_1, \dots, X_d)^\top$, where $x = (x_1, \dots, x_d)^\top$. We shall defer proofs to the end of the section after the statement and discussion of these extensions.

We first present an mth-order anticoncentration inequality in terms of partial correlations between the components of the Gaussian vector X. For $i_{1:m} = (i_1, \dots, i_m) \in [d]^m_{\neq}$ and $(j,k) \in [d]^2_{i_{1:m},\neq}$, the partial correlation of X_j and X_k given $X_{i_{1:m}} = (X_{i_1}, \dots, X_{i_m})^\top$ is

$$\rho_{j,k|i_{1:m}} = \mathrm{Corr}(X_j, X_k | X_{i_{1:m}}) \tag{4.2}$$

with the convention $\rho_{(j,k)|i_{1:0}} = \rho_{j,k} = \mathrm{Corr}(X_j, X_k)$. Define threshold levels

$$t_{i_{1:j}} = 1 \wedge \min\left\{ \frac{\sqrt{1 - \rho_{i_j,k|i_{1:(j-1)}}}}{\sqrt{1 + \rho_{i_j,k|i_{1:(j-1)}}}}, k \in [d]_{i_{1:j}} \right\}, \quad i_{1:j} \in [d]^j_{\neq}, \tag{4.3}$$

and an extension of a simplification of (2.11) as

$$\alpha'_{i_{1:m}} = \left|\det\left(\Sigma^{i_{1:m}}\right)\right|^{1/2} \prod_{j=1}^{m-1} t_{i_{1:j}}, \quad i_{1:m} \in [d]^m_{\neq}, \tag{4.4}$$

with $\alpha'_i = \sigma_i = \Sigma^{1/2}_{i,i}$, where $\Sigma^{i_{1:m}}$ is the $m \times m$ covariance matrix of $X_{i_{1:m}}$. Compared with (2.11) where $\cos(\theta_{i,\min}) = \rho_{i,\max} = \max\{\rho_{i,k} : k \in [d]_i\}$, $t_i = 1 \wedge \tan(\theta_{i,\min}/2)$ for $i_1 = i$ in (4.3), so that $\alpha'_{i,j}$ in (4.4) and $\alpha_{i,j}$ in (2.11) are within a factor of 2 of each other.

Theorem 5. *For any positive integer $m < d$, there exists a finite numerical constant C_m depending on m only such that for any set of positive constants $\{b_{i_{1:m}} : i_{1:m} \in [d]^m_<\}$ with ordered values $b_{(1)} \le b_{(2)} \le \cdots$, the mth-order derivatives in* (4.1) *are bounded by*

$$\sup_x \sum_{i_{1:m} \in [d]^m_<} \frac{\alpha'_{i_{1:m}}}{b_{i_{1:m}}} \left|G_{i_{1:m}}(x)\right| \le C_m \max_{1 \le k \le d} \frac{(1 + \sqrt{2\log k})^m}{b_{(k)}}, \tag{4.5}$$

where $\alpha'_{i_{1:m}}$ are as in (4.4) *for $i_{1:m} \in [d]^m_<$.*

As mentioned in the discussion of (2.5), the upper bound in our anticoncentration inequality can be expressed in terms of the minimum eigenvalue of the correlation matrix of no more than m components of X. For $i_{1:m} = (i_1, \dots, i_m) \in [d]^m_{\neq}$, let $\rho^{i_{1:m}}$ be the $m \times m$ correlation matrix of $X_{i_{1:m}} = (X_{i_1}, \dots, X_{i_m})^\top$ and define the corresponding minimum eigenvalue as

$$\lambda^{i_{1:m}}_{\min} = \min\left\{u^\top \rho^{i_{1:m}} u : u \in \mathbb{R}^m, \|u\|_2 = 1\right\}. \tag{4.6}$$

The following theorem asserts that the quantity $\alpha'_{i_{1:m}}$ in Theorem 5 can be replaced by

$$\alpha''_{i_{1:m}} = (\sigma_{i_1} \cdots \sigma_{i_m}) \left(\lambda^{i_{1:m}}_{\min} \prod_{j=1}^{m-1} \min\{\lambda^{i_1,\dots,i_j,k}_{\min} : k \in [d]_{i_1,\dots,i_j}\} \right)^{1/2}. \tag{4.7}$$

For $m = 2$, $\min\{\lambda^{i,k}_{\min} : k \in [d]_i\} = 1 - \max\{|\rho_{i,k}| : k \in [d]_i\}$ in (4.7) while the sharper one-sided $t_i = 1 \wedge \tan(\theta_{i,\min}/2)$ and $2\tan(\theta_{i,\min}/4)$ are respectively used in (4.4) and (2.11), where $\cos(\theta_{i,\min}) = \max\{\rho_{i,k} : k \in [d]_i\}$.

Theorem 6. *For any positive integer $m < d$, there exists a finite numerical constant C_m depending on m only such that (4.5) holds with the quantity $\alpha'_{i_{1:m}}$ replaced by the quantity $\alpha''_{i_{1:m}}$ in (4.7). In particular,*

$$\sup_x \sum_{i_{1:m}\in[d]^m_<} \left|G_{i_{1:m}}(x)\right| \le \frac{C_m(1+\sqrt{2\log d})^m}{\min\{\alpha''_{i_{1:m}} : i_{1:m}\in[d]^m_<\}}, \tag{4.8}$$

and in terms of the sparse eigenvalue $\lambda_{\min,j} = \min\{\lambda^{i_{1:j}}_{\min} : i_{1:j}\in[d]^j_{\neq}\}$ with the $\lambda^{i_{1:j}}_{\min}$ in (4.6)

$$\sup_x \sum_{i_{1:m}\in[d]^m_<} \left(\prod_{j=1}^m \sigma_{i_j}\right)\left|G_{i_{1:m}}(x)\right| \le \frac{C_m(1+\sqrt{2\log d})^m}{\lambda_{\min,m}\sqrt{\lambda_{\min,m-1}\cdots\lambda_{\min,2}}}. \tag{4.9}$$

While the quantity $\alpha''_{i_{1:m}}$ in (4.7) is expressed in terms of the more familiar minimum eigenvalues, it is bounded from the above by the quantity $\alpha'_{i_{1:m}}$ in (4.4) up to a constant factor. Moreover, compared with $\alpha''_{i_{1:m}}$, the quantity $\alpha'_{i_{1:m}}$ is potentially of larger order as it involves one-sided threshold levels $t_{i_{1:j}}$ in (4.3). Thus, Theorem 5 is slightly sharper than Theorem 6. We present next an upper bound of the ratio $\alpha''_{i_{1:m}}/\alpha'_{i_{1:m}}$ through a Cholesky decomposition of correlation matrices, and thus the validity of Theorem 6 as a corollary of Theorem 5.

Because the quantity $\alpha'_{i_{1:m}}$ involves partial correlations in (4.3), we construct the Cholesky decomposition through a Gram–Schmidt orthogonalization process. Let $Y_i = (X_i - \mu_i)/\sigma_i$. In the Gram–Schmidt orthogonalization process, we write

$$Y_{k|i_{1:j}} = \frac{Y_{k|i_{1:(j-1)}} - \rho_{i_j,k|i_{1:(j-1)}} Y_{i_j|i_{1:(j-1)}}}{(1-\rho^2_{i_j,k|i_{1:(j-1)}})^{1/2}}, \quad k\in[d]_{i_{1:j}},\ j=0,\dots,m-1, \tag{4.10}$$

with the convention $Y_{k|i_{1:0}} = Y_k$. Let $A^{i_{1:m}}$ be the matrix satisfying

$$\begin{pmatrix} Y_{i_1} \\ Y_{i_2|i_1} \\ \vdots \\ Y_{i_m|i_{1:(m-1)}} \end{pmatrix} = A^{i_{1:m}} \begin{pmatrix} Y_{i_1} \\ Y_{i_2} \\ \vdots \\ Y_{i_m} \end{pmatrix}. \tag{4.11}$$

Because $\{Y_{k|i_{1:j}}, k\in[d]_{i_{1:j}}\}$ and $Y_{i_{1:j}}$ are independent, $Y_{i_1}, Y_{i_2|i_1}, \dots, Y_{i_m|i_{1:(m-1)}}$ are iid $N(0,1)$ variables, so that $A^{i_{1:m}}$ gives a Cholesky decomposition of $\rho^{i_{1:m}}$ in the sense of

$$I_{m\times m} = A^{i_{1:m}}\rho^{i_{1:m}}\left(A^{i_{1:m}}\right)^\top. \tag{4.12}$$

As the spectrum norm of $A^{i_{1:m}}$ is bounded by $(\lambda^{i_{1:m}}_{\min})^{-1/2}$ and the elements of $A^{i_{1:m}}$ are expressed in terms of partial correlations, (4.10), (4.11), and (4.12) lead to the following lemma.

Lemma 1. *For $i_{1:m}\in[d]^m_{\neq}$, let $\rho^{i_{1:m}}$ be the $m\times m$ correlation matrix of the Gaussian vector $X_{i_{1:m}} = (X_{i_1},\dots,X_{i_m})^\top$. For $j\in[d]_{i_{1:m}}$, let $\rho_{(i_m,j)|i_{1:(m-1)}}$ be the partial correlation as defined in (4.2). Then, the determinant of $\rho^{i_{1:m}}$ is given by*

$$\det\left(\rho^{i_{1:m}}\right) = \prod_{k=2}^m\prod_{j=1}^{k-1}\left(1-\rho^2_{i_j,i_k|i_{1:(j-1)}}\right) \tag{4.13}$$

with $\rho_{i_k,i_1|i_{1:0}} = \rho_{i_k,i_1}$. *Consequently, with* $\lambda_{\min}^{i_{1:m}}$ *being the smallest eigenvalue of* $\rho^{i_{1:m}}$,

$$\det(\rho^{i_{1:m}}) \prod_{k=1}^{m-1} \min\left(1, \frac{1-\rho_{\ell_{k+1},i_k|i_{1:(k-1)}}}{1+\rho_{\ell_{k+1},i_k|i_{1:(k-1)}}}\right) \geq \lambda_{\min}^{i_{1:m}} \prod_{k=1}^{m-1} (\lambda_{\min}^{i_{1:k},\ell_{k+1}}/5) \tag{4.14}$$

for all $\ell_{k+1} \in [d]_{i_{1:k}}$, $k = 1, \ldots, m-1$.

It follows from Lemma 1 that $\alpha''_{i_{1:m}} \leq 5^{(m-1)/2}\alpha'_{i_{1:m}}$ for the quantities in (4.7) and (4.4), respectively, so that Theorem 6 is a consequence of Theorem 5.

We still need to consider the case where the differentiation is taken multiple times in some of the directions. As a general study of such results is beyond the scope of this paper, we present here an upper bound for the third derivative and discuss the main difficulties in the higher-order cases.

Theorem 7. *Let* $G_{i,j,k}(x)$ *be as in* (4.1) *for a Gaussian vector* $X_{1:d}$ *with marginal distributions* $X_i \sim N(\mu_i, \sigma_i^2)$. *Let* $\lambda_{\min,j}$ *be the lower sparse eigenvalue as in Theorem* 6 *for the correlation matrices of* j*-components of* $X_{1:d}$. *Then, for some numeric constant* C_3,

$$\sup_x \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \sigma_i\sigma_j\sigma_k |G_{i,j,k}(x)| \leq \frac{C_3(1+\sqrt{2\log d})^3}{\lambda_{\min,3}\sqrt{\lambda_{\min,2}}}. \tag{4.15}$$

In our approach, the proof of Theorem 7 and the analysis in higher-order cases involve factors which can be expressed as regression coefficients. Let $Y_i = (X_i - \mu_i)/\sigma_i$ as in (4.10). Given $i_{1:m} \in [d]^m_{\neq}$ and $k \in [d]_{i_{1:m}}$, the linear regression of Y_k against $Y_{i_{1:m}}$ is given by

$$\mathbb{E}[Y_k|Y_{i_{1:m}}] = \sum_{j=1}^m \beta_{i_j}^{k|i_{1:m}} Y_{i_j}. \tag{4.16}$$

These regression coefficients $\beta_{i_j}^{k|i_{1:m}}$ appear in the derivatives (4.1) as follows. Let

$$P_{i_{1:m}}(x) = \mathbb{P}\{X_k \leq x_k \ \forall\, k \in [d]_{i_{1:m}} | X_{i_{1:m}} = x_{i_{1:m}}\} \tag{4.17}$$

and $\phi_{i_{1:m}}(x)$ be the joint density of $X_{i_{1:m}}$. As in (2.7), we have

$$G_{i_{1:m}}(x) = P_{i_{1:m}}(x)\phi_{i_{1:m}}(x). \tag{4.18}$$

As $Y_k - \mathbb{E}[Y_k|Y_{i_{1:m}}]$ is independent of $Y_{i_{1:m}}$ and $Y_{i_{1:m}}$ is linear in $X_{i_{1:m}}$, the conditional probability in (4.17) can be written as

$$P_{i_{1:m}}(x) = \mathbb{P}\left\{Y_k - \mathbb{E}[Y_k|Y_{i_{1:m}}] \leq \frac{x_k-\mu_k}{\sigma_k} - \sum_{j=1}^m \beta_{i_j}^{k|i_{1:m}} \frac{x_{i_j}-\mu_{i_j}}{\sigma_{i_j}}, \ \forall k \in [d]_{i_{1:m}}\right\}.$$

Thus, for $a \in [m]$,

$$\begin{aligned}(\partial/\partial x_{i_a})G_{i_{1:m}}(x) &= P_{i_{1:m}}(x)(\partial/\partial x_{i_a})\phi_{i_{1:m}}(x) + \phi_{i_{1:m}}(x)(\partial/\partial x_{i_a})P_{i_{1:m}}(x) \\ &= G_{i_{1:m}}(x)(\partial/\partial x_{i_a})\log\phi_{i_{1:m}}(x) \\ &\quad - \sum_{i_{m+1}\in[d]_{i_{1:m}}} G_{i_{1:m+1}}\beta_{i_a}^{i_{m+1}|i_{1:m}}\sigma_{i_{m+1}}/\sigma_{i_a}.\end{aligned} \tag{4.19}$$

In general, the scaled mth partial derivative $(\sigma_{i_1}\partial/\partial x_{i_1})^{j_1}\cdots(\sigma_{i_k}\partial/\partial x_{i_k})^{j_k}G(x)$ with $j_1+\cdots+j_k=m$ would involve a term of the form

$$(-1)^{m-k}\sigma_{i_1}\cdots\sigma_{i_m}G_{i_1,\dots,i_m}\beta_{\ell_{k+1}}^{i_{k+1}|i_{1:k}}\cdots\beta_{\ell_m}^{i_m|i_{1:(m-1)}}$$

such that i_a appears j_a-1 times in $\ell_{k+1},\dots,\ell_m$, optionally in the order of $a=1,\dots,k$. While $\beta_{\ell_{k+1}}^{i_{k+1}|i_{1:k}}\lesssim 1/\lambda_{\min}^{i_{1:k}}$, a difficulty is to find sharper bounds for

$$\sigma_{i_1}\cdots\sigma_{i_m}\beta_{\ell_{k+1}}^{i_{k+1}|i_{1:k}}\cdots\beta_{\ell_m}^{i_m|i_{1:(m-1)}}/\alpha'_{i_{1:m}}$$

to extend Theorem 7 to higher order in the same form as that of (4.15) and (4.9).

Proof of Theorem 5. Consider a fixed sequence of integers $i_{1:m}\in[d]_{\neq}^m$. Define

$$X'_{j|i_{1:m}}=\frac{X'_{j|i_{1:(m-1)}}-\rho_{(i_m,j)|i_{1:(m-1)}}X'_{i_m|i_{1:(m-1)}}}{(1-\rho^2_{(i_m,j)|i_{1:(m-1)}})^{1/2}} \tag{4.20}$$

as in (2.16) with the partial correlation $\rho_{(i_m,j)|i_{1:(m-1)}}$ in (4.2) and initialization $X'_{j|i_{1:0}}=X'_j=(X_j-x_j)/\sigma_j$. This is the same Gram–Schmidt orthogonalization process as in (4.10) but the X'_j are not centered to have mean zero at the initialization. Still the covariance structure of $X'_{j|i_{1:m}}$ is the same as that of $Y_{j|i_{1:m}}$. Because $X'_{k|i_{1:m}}, k\in[d]_{i_{1:m}}$ are independent of $X_{i_{1:m}}$,

$$\begin{aligned}G_{i_{1:m}}(x)&=\int_{y_k\le x_k,\forall k\in[d]_{i_{1:m}}}\phi_{[d]}(y)\prod_{k\in[d]_{i_{1:m}}}dy_k|_{y_{i_{1:m}}=x_{i_{1:m}}}\\&=\mathbb{P}\{X'_k<0\ \forall k\in[d]_{i_{1:m}}|X'_{i_1}=\cdots=X'_{i_m}=0\}\phi_{i_{1:m}}(x_{i_{1:m}})\\&=\mathbb{P}\{X'_{k|i_{1:m}}<0\ \forall k\in[d]_{i_{1:m}}\}\phi_{i_{1:m}}(x_{i_{1:m}})\end{aligned} \tag{4.21}$$

as in (2.17) and (4.18). To bound the probability $\mathbb{P}\{X'_{k|i_{1:m}}<0\ \forall k\in[d]_{i_{1:m}}\}$, we define

$$\pi_{i_{1:m}}=\mathbb{P}\{\mathscr{C}_{i_{1:m}}\}G_{i_{1:m}}(x)/\phi_{i_{1:m}}(x_{i_{1:m}}), \tag{4.22}$$

where $\mathscr{C}_{i_{1:m}}$ is defined with the threshold levels $t_{i_{1:j}}$ in (4.3) as

$$\mathscr{C}_{i_{1:m}}=\{0<X'_{i_{j+1}|i_{1:j}}\le t_{i_{1:j}}X'_{i_j|i_{1:(j-1)}},1\le j<m,X_{i_1}>0\}.$$

Given integers $j\ge 0$ and $i_{1:j}$, define the rank of $X'_{k|i_{1:j}}$ as

$$R_{k|i_{1:j}}=\sum_{\ell\in[d]_{i_{1:j}}}I\{X'_{\ell|i_{1:j}}\ge X'_{k|i_{1:j}}\},\quad k\in[d]_{i_{1:j}}.$$

Here $R_{k|i_{1:0}}=\sum_{\ell\in[d]}I\{X'_\ell\ge X'_k\}$ is the marginal rank of X'_k as $X'_{\ell|i_{1:0}}=X'_\ell$ in (4.20). In the event $\{X'_{k|i_{1:m}}<0\ \forall k\in[d]_{i_{1:m}}\}\cap\mathscr{C}_{i_{1:m}}$, we have $R_{i_m|i_1,\dots,i_{m-1}}=1$ due to

$$X'_{k|i_{1:(m-1)}}\le\rho_{i_m,k|i_{1:(m-1)}}X'_{i_m|i_{1:(m-1)}}\le X'_{i_m|i_{1:(m-1)}},$$

and by induction $R_{i_j|i_{1:(j-1)}}=1$ given $R_{i_{j+1}|i_{1:j}}=1$ for $j=m-1,\dots,1$ due to

$$\begin{aligned}X'_{k|i_{1:(j-1)}}&=\rho_{i_j,k|i_{1:(j-1)}}X'_{i_j|i_{1:(j-1)}}+\{1-\rho^2_{i_j,k|i_{1:(j-1)}}\}^{1/2}X'_{k|i_{1:j}}\\&\le\rho_{i_j,k|i_{1:(j-1)}}X'_{i_j|i_{1:(j-1)}}+\{1-\rho^2_{i_j,k|i_{1:(j-1)}}\}^{1/2}X'_{i_{j+1}|i_{1:j}}\\&\le\{\rho_{i_j,k|i_{1:(j-1)}}+(1-\rho^2_{i_j,k|i_{1:(j-1)}})^{1/2}t_{i_{1:j}}\}X'_{i_j|i_{1:(j-1)}}\\&\le X'_{i_j|i_{1:(j-1)}},\end{aligned}$$

by the choice of $t_{i_{1:j}}$ in (4.3). Thus, due to the independence between the event $\mathscr{C}_{i_{1:m}}$ and the set of random variables $\{X'_{k|i_{1:m}}, k \in [d]_{i_{1:m}}\}$,

$$\pi_{i_{1:m}} = \mathbb{P}\{X'_{k|i_{1:m}} < 0 \ \forall k \in [d]_{i_{1:m}}, \mathscr{C}_{i_{1:m}}\} \le \mathbb{P}\{R_{i_j|i_{1:(j-1)}}, 1 \le j \le m\}.$$

Consequently,

$$\sum_{i_{1:m} \in [d]^m_{\neq}} \pi_{i_{1:m}} \le 1. \tag{4.23}$$

We still need to find a suitable lower bound for $\mathbb{P}\{\mathscr{C}_{i_{1:m}}\}$ to use (4.23). Let $\nu_i = \mathbb{E}[X'_i]$,

$$\nu_{i_{1:m}} = \{(\nu_{i_1}, \dots, \nu_{i_m})(\Sigma^{i_{1:m}})^{-1}(\nu_{i_1}, \dots, \nu_{i_m})^\top\}^{1/2}, \quad \nu_{i_j|i_{1:(j-1)}} = \mathbb{E}[X'_{i_j|i_{1:(j-1)}}],$$

and $\varphi_{i_j|i_{1:(j-1)}}$ be the $N(\nu_{i_j|i_{1:(j-1)}}, 1)$ density. We shall prove that

$$\mathbb{P}\{\mathscr{C}_{i_{1:m}}\} \ge \alpha'_{i_{1:m}} \phi_{i_{1:m}}(x) C'_m J_m(\nu_{i_{1:m}})/\nu^m_{i_{1:m}}, \tag{4.24}$$

with $J_m(t) = \int_0^\infty y^{m-1} e^{-y-y^2/(2t^2)} dy$ and $C'_m = 2\pi^{m/2}/\{2^m \Gamma(m/2) m!\}$, and that

$$\big|\det(\Sigma^{i_{1:m}})\big|^{1/2} \phi_{i_{1:m}}(x) = (2\pi)^{-m/2} \exp(-\nu^2_{i_{1:m}}/2). \tag{4.25}$$

Because $X'_{i_j|i_{1:(j-1)}}$ are defined by the Gram–Schmidt process, they are independent $N(\nu_{i_j|i_{1:(j-1)}}, 1)$ variables. Thus, as the Jacobian of a linear transformation of $X'_{i_1}, \dots, X'_{i_m}$ is a constant, $|\det(\Sigma^{i_{1:m}})|^{1/2} \phi_{i_{1:m}}(x) = \prod_{j=1}^m \varphi_{i_j|i_{1:(j-1)}}(0)$ and $\sum_{j=1}^m \nu^2_{i_j|i_{1:(j-1)}} = \nu^2_{i_{1:m}}$. This gives (4.25). Because $t_{i_{1:j}} \le 1$ for all j, it follows that

$$\begin{aligned}
\mathbb{P}\{\mathscr{C}_{i_{1:m}}\} &= \mathbb{P}\{0 < X'_{i_{j+1}|i_{1:j}} \le t_{i_{1:j}} X'_{i_j|i_{1:(j-1)}}, 1 \le j < m, X_{i_1} > 0\} \\
&= \int_0^\infty \int_0^{t_{i_1} x_1} \cdots \int_0^{t_{i_{1:(m-1)}} x_{m-1}} \prod_{j=1}^m \varphi(x_j - \nu_{i_j|i_1,\dots,i_{j-1}}) dx_j \\
&\ge \int_0^\infty \int_0^{t_{i_1} x_1} \cdots \int_0^{t_{i_{1:(m-1)}} x_{m-1}} \frac{\exp(-\nu^2_{i_{1:m}}/2 - \nu_{i_{1:m}}\|x\|_2 - \|x\|_2^2/2)}{(2\pi)^{m/2}} dx \\
&\ge \left(\prod_{j=1}^{m-1} t_{i_{1:j}}\right) \int_0^\infty \int_0^{x_1} \cdots \int_0^{x_{m-1}} \frac{\exp(-\nu^2_{i_{1:m}}/2 - \nu_{i_{1:m}}\|x\|_2 - \|x\|_2^2/2)}{(2\pi)^{m/2}} dx \\
&= \left(\prod_{j=1}^{m-1} t_{i_{1:j}}\right) \frac{2}{m! 2^m \Gamma(m/2) 2^{m/2}} \int_0^\infty y^{m-1} e^{-\nu^2_{i:m}/2 - \nu_{i_{1:m}} y - y^2/2} dy \\
&= \left(\prod_{j=1}^{m-1} t_{i_{1:j}}\right) \frac{2\pi^{m/2} |\det(\Sigma^{i_{1:m}})|^{1/2} \phi_{i_{1:m}}(x)}{2^m \Gamma(m/2) m!} \frac{J_m(\nu_{i_{1:m}})}{\nu^m_{i_{1:m}}} \\
&= \alpha'_{i_{1:m}} \phi_{i_{1:m}}(x) C'_m J_m(\nu_{i_{1:m}})/\nu^m_{i_{1:m}}.
\end{aligned}$$

Putting together (4.4), (4.21), (4.22), (4.23), (4.24), and (4.25), we find that

$$\begin{aligned}
\sum_{i_{1:m}\in[d]^m_<} \frac{\alpha'_{i_{1:m}}}{b_{i_{1:m}}} G_{i_{1:m}}(x) &= \sum_{i_{1:m}\in[d]^m_<} \min\left\{\frac{e^{-\nu^2_{i_{1:m}}/2}}{(2\pi)^{m/2} b_{i_{1:m}}}, \frac{\alpha'_{i_{1:m}}}{b_{i_{1:m}}} \frac{\pi_{i_{1:m}}\phi_{i_{1:m}}(x_{i_{1:m}})}{\mathbb{P}\{\mathscr{C}_{i_{1:m}}\}}\right\} \\
&\le \sum_{i_{1:m}\in[d]^m_<} \min\left\{\frac{e^{-\nu^2_{i_{1:m}}/2}}{(2\pi)^{m/2} b_{i_{1:m}}}, \frac{\pi_{i_{1:m}}\nu^m_{i_{1:m}}}{b_{i_{1:m}} C'_m J_m(\nu_{i_{1:m}})}\right\} \\
&\le \sum_{k\notin K} \frac{e^{-L_k^2/2}}{(2\pi)^{m/2} b_{(1)}} + \max_{k\in K} \frac{m! L_k^m}{b_{(k)} C'_m J_m(L_k)} \\
&\le C_m \max_{k\in[d]} \frac{(1+\sqrt{2\log k})^m}{b_{(k)}}, \qquad (4.26)
\end{aligned}$$

with $L_k = 1 + \sqrt{2\log k}$ and $K = \{k : \nu_{(k)} \le L_k\}$ due to the monotonicity $J_m(t)/t^m \uparrow$ in $(0,\infty)$ and $J_m(L_k) \ge J_m(L_1) = J_m(1)$. This completes the proof of Theorem 5. ■

Proof of Theorem 6. In view of the definitions of $\alpha'_{i_{1:m}}$ and $\alpha''_{i_{1:m}}$ in (4.4) and (4.7), respectively, Theorem 6 follows directly from Theorem 5 and Lemma 1. ■

Proof of Lemma 1. Let $i_{1:m} = 1 : m$, as a permutation of labels does not change the conclusions. It follows from (4.12) that

$$\det(\rho^{1:m})\big(\det(A^{1:m})\big)^2 = 1.$$

Because $A^{1:m}$ is a lower-triangular matrix with diagonal elements $A^{1:m}_{1,1} = 1$ and $A^{1:m}_{k,k} = \prod_{j=1}^{k-1}(1-\rho^2_{k,j|1:(j-1)})^{-1/2}$ for $2 \le k \le m$,

$$\det(\rho^{1:m}) = \frac{1}{\det^2(A^{1:m})} = \prod_{k=2}^m \frac{1}{(A^{1:m}_{k,k})^2} = \prod_{k=2}^m\prod_{j=1}^{k-1}\big(1-\rho^2_{k,j|1:(j-1)}\big).$$

This gives (4.13). Now we write for $k < \ell_{k+1} \le d$

$$\begin{aligned}
&\det(\rho^{1:m}) \prod_{k=1}^{m-1} \min\left(1, \frac{1-\rho_{\ell_{k+1},k|1:(k-1)}}{1+\rho_{\ell_{k+1},k|1:(k-1)}}\right) \\
&= \left(\prod_{j=1}^{m-1}\big(1-\rho^2_{m,j|1:(j-1)}\big)\right) \prod_{k=1}^{m-1}\left\{\min\left(1, \frac{1-\rho_{\ell_{k+1},k|1:(k-1)}}{1+\rho_{\ell_{k+1},k|1:(k-1)}}\right) \prod_{j=1}^{k-1}\big(1-\rho^2_{k,j|1:(j-1)}\big)\right\}, \qquad (4.27)
\end{aligned}$$

with the convention $\prod_{j=1}^{k-1}(1-\rho^2_{k,j|1:(j-1)}) = 1$ for $k = 1$. By (4.12),

$$\left(\prod_{j=1}^{m-1}\big(1-\rho^2_{m,j|1:(j-1)}\big)\right)^{-1} = \big(A^{1:m}_{m,m}\big)^2 \le \frac{1}{\lambda_{\min}(\rho^{1:m})},$$

as the spectral norm of $A^{1:m}$ is no greater than $1/\lambda_{\min}^{1/2}(\rho^{1:m})$. For $1 \le k \le m-1$,

$$
\begin{aligned}
&\max\left(1, \frac{\sqrt{1+\rho_{k+1,k|1:(k-1)}}}{\sqrt{1-\rho_{k+1,k|1:(k-1)}}}\right)\prod_{j=1}^{k-1}\left(1-\rho_{k,j|1:(j-1)}^2\right)^{-1/2} \\
&\le \left(1 + 2\frac{|\rho_{k+1,k|1:(k-1)}|}{\sqrt{1-\rho_{k+1,k|1:(k-1)}^2}}\right)\prod_{j=1}^{k-1}\left(1-\rho_{k,j|1:(j-1)}^2\right)^{-1/2} \\
&= A_{k.k}^{1:(k+1)} + 2\left|A_{k+1,k}^{1:(k+1)}\right| \\
&\le \sqrt{5/\lambda_{\min}\left(\rho^{1:(k+1)}\right)},
\end{aligned}
$$

due to $\sqrt{(1+t)/(1-t)} \le 1 + 2|t|/\sqrt{1-t^2}$ or, equivalently, $1 + t \le \sqrt{1-t^2} + 2|t|$ for $|t| < 1$. This and (4.27) give (4.14) because labels do not matter. ■

Proof of Theorem 7. Let $\phi_{i,j}(x)$ be the joint density of $(X_i, X_j)^\top$, $\nu_i(x) = (x_i - \mu_i)/\sigma_i$, and $\nu_{i,j}(x)$ be given by $-\nu_{i,j}^2(x)/2 = \log(2\pi \det^{1/2}(\Sigma^{i,j})\phi_{i,j}(x))$ as in (4.25). As in (4.19) and similar to (2.26), for $i \ne j$,

$$
G_{i,j,j}(x) = G_{i,j}(x)(\partial/\partial x_j)\log\phi_{i,j}(x) - \sum_{k\in[d]_{i,j}} G_{i,j,k}(x)\beta_j^{k|i,j}\sigma_k/\sigma_j,
$$

with $|(\partial/\partial x_j)\log\phi_{i,j}(x)| = |e_j^\top(\rho^{i,j})^{-1}(\nu_i(x),\nu_j(x))^\top|/\sigma_j \le (\lambda_{\min}^{i,j})^{-1/2}\nu_{i,j}(x)/\sigma_j$ and

$$
\beta_j^{k|i,j} = \frac{(1-\rho_{k,j|i}^2)^{-1/2}\rho_{k,j|i}(1-\rho_{j,i}^2)^{-1/2}}{(1-\rho_{k,j|i}^2)^{-1/2}(1-\rho_{k,i}^2)^{-1/2}} = \frac{\rho_{k,j|i}(1-\rho_{k,i}^2)^{1/2}}{(1-\rho_{i,j}^2)^{1/2}}.
$$

The formula for the regression coefficient is obtained by noticing that $\beta_j^{k|i,j} = -A_{k,j}^{i,j,k}/A_{k,k}^{i,j,k}$ in the Cholesky decomposition (4.11) with the matrix elements $A_{k,j}^{i,j,k}$ and $A_{k,k}^{i,j,k}$ determined by the Gram–Schmidt formula (4.10). By (4.13),

$$
\det\left(\rho^{i,j,k}\right) = \left(1-\rho_{k,j|i}^2\right)\left(1-\rho_{k,i}^2\right)\left(1-\rho_{i,j}^2\right).
$$

As in the proof of Lemma 1, we have, by (4.4) and (4.3),

$$
\begin{aligned}
\left(\frac{\sigma_i\sigma_j\sigma_k\beta_j^{k|i,j}}{\alpha'_{i,j,k}}\right)^2 &= \frac{\rho_{k,j|i}^2(1-\rho_{k,i}^2)(1-\rho_{i,j}^2)^{-1}}{t_{i,j}^2 t_i^2(1-\rho_{k,j|i}^2)(1-\rho_{k,i}^2)(1-\rho_{i,j}^2)} \\
&\le \frac{1}{t_i^2} \times \frac{1}{t_{i,j}^2(1-\rho_{i,j}^2)} \times \frac{1}{(1-\rho_{j,k|i}^2)(1-\rho_{i,j}^2)} \\
&\le \frac{5}{\lambda_{\min}^{i,\ell_j}} \times \frac{5}{\lambda_{\min}^{i,j,\ell_k}} \times \frac{1}{\lambda_{\min}^{i,j,k}},
\end{aligned}
$$

for some $\ell_j \in [d]_i$ and $\ell_k \in [d]_{i,j}$. It follows that

$$
\begin{aligned}
\sum_{(i,j)\in[d]_{\ne}^2} \sigma_i\sigma_j^2\left|G_{i,j,j}(x)\right| \le{}& \sum_{(i,j)\in[d]_{\ne}^2} \sigma_i\sigma_j G_{i,j}(x)\left(\lambda_{\min}^{i,j}\right)^{-1/2}\nu_{i,j}(x) \\
&+ \sum_{(i,j,k)\in\in[d]_{\ne}^3} \frac{5\alpha'_{i,k,j}|G_{i,k,j}(x)|}{\lambda_{\min,3}\sqrt{\lambda_{\min,2}}}.
\end{aligned}
\tag{4.28}
$$

Similar to (4.26) in the proof of Theorem 5, the first term on the right-hand side above is bounded by

$$\sum_{(i,j)\in[d]^2_{\neq}} \frac{\sigma_i\sigma_j G_{i,j}(x)\nu_{i,j}(x)}{\sqrt{\lambda_{\min,2}}} \le \frac{C_3'(1+\sqrt{2\log d})^3}{\lambda_{\min,2}}.$$

By Theorem 5, $\sum_{(i,j,k)\in\in[d]^3_{\neq}} \alpha'_{i,j,k}|G_{i,j,k}(x)| \le 6C_3(1+\sqrt{2\log d})^3$. Thus, the right-hand side of (4.28) is bounded by $C_3''(1+\sqrt{2\log d})^3/(\lambda_{\min,3}\sqrt{\lambda_{\min,2}})$. Similarly,

$$\sum_{i=1}^{2} \sigma_i^3 \left|G_{i,i,i}(x)\right| \le \frac{C_3''(1+\sqrt{2\log d})^3}{\lambda_{\min,3}\sqrt{\lambda_{\min,2}}}$$

by differentiating the identity

$$G_{i,i}(x) = G_i(x)\nu_i(x)/\sigma_i - \sum_{j:i\neq j\in[d]} G_{i,j}(x)\rho_{i,j}\sigma_j/\sigma_i$$

in (2.26). The conclusion follows as the sum over $[d]^3_{\neq}$ is bounded in Theorem 6. ■

FUNDING

This work was partially supported by NSF grants IIS-1741390, CCF-1934924, and DMS-2052949.

REFERENCES

[1] R. J. Adler, An introduction to continuity, extrema, and related topics for general gaussian processes. *IMS* (1990).

[2] N. H. Anderson, P. Hall, and D. Titterington, Edgeworth expansions in very-high-dimensional problems. *J. Statist. Plann. Inference* **70** (1998), no. 1, 1–18.

[3] V. Bentkus, Smooth approximations of the norm and differentiable functions with bounded support in Banach space l_k^∞. *Lith. Math. J.* **30** (1990), no. 3, 223–230.

[4] V. Bentkus, On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** (2003), no. 2, 385–402.

[5] S. Chatterjee, An error bound in the Sudakov–Fernique inequality. 2005, arXiv:math/0510424.

[6] V. Chernozhukov, D. Chetverikov, and K. Kato, Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** (2013), no. 6, 2786–2819.

[7] V. Chernozhukov, D. Chetverikov, and K. Kato, Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probab. Theory Related Fields* **162** (2015), no. 1, 47–70.

[8] V. Chernozhukov, D. Chetverikov, and K. Kato, Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** (2017), no. 4, 2309–2352.

[9] V. Chernozhukov, D. Chetverikov, K. Kato, and Y. Koike, Improved central limit theorem and bootstrap approximations in high dimensions. 2019, arXiv:1912.10529.

[10] V. Chernozhukov, D. Chetverikov, and Y. Koike, Nearly optimal central limit theorem and bootstrap approximations in high dimensions. 2020, arXiv:2012.09513.

[11] D. Das and S. Lahiri, Central limit theorem in high dimensions: The optimal bound on dimension growth rate. *Trans. Amer. Math. Soc.* **374** (2021), no. 10, 6991–7009.

[12] H. Deng, Slightly conservative bootstrap for maxima of sums. 2020, arXiv:2007.15877.

[13] H. Deng and C.-H. Zhang, Beyond gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Ann. Statist.* **48** (2020), no. 6, 3643–3671.

[14] X. Fang and Y. Koike, High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.* **31** (2021), no. 4, 1660–1686.

[15] X. Fernique, Des resultats nouveaux sur les processus gaussiens. *C. R. Acad. Sci., Sér. A–B* **278** (1974), A363–A365.

[16] Y. Gordon, Some inequalities for Gaussian processes and applications. *Israel J. Math.* **50** (1985), no. 4, 265–289.

[17] F. Gotze, On the rate of convergence in the multivariate CLT. *Ann. Probab.* (1991), 724–739.

[18] F. Götze, A. Naumov, V. Spokoiny, and V. Ulyanov, Gaussian comparison and anti-concentration inequalities for norms of Gaussian random elements. 2017, arXiv:1708.08663.

[19] Y. Koike, Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* (2020), 1–41.

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

[21] M. Ledoux and M. Talagrand, *Probability in Banach spaces: isoperimetry and processes* 23. Springer, 1991.

[22] W. V. Li and Q.-M. Shao, Gaussian processes: inequalities, small ball probabilities and applications. *Handb. Statist.* **19** (2001), 533–597.

[23] W. V. Li and Q.-M. Shao, A normal comparison inequality and its applications. *Probab. Theory Related Fields* **122** (2002), no. 4, 494–508.

[24] J. W. Lindeberg, Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15** (1922), no. 1, 211–225.

[25] M. E. Lopes, Central limit theorem and bootstrap approximation in high dimensions with near $1/\sqrt{n}$ rates. 2020, arXiv:2009.06004.

[26] F. Nazarov, On the maximal perimeter of a convex set in r^n with respect to a Gaussian measure. In *Geometric aspects of functional analysis*, pp. 169–187, Springer, 2003.

[27] I. Nourdin and F. Viens, Density formula and concentration inequalities with Malliavin calculus. *Electron. J. Probab.* **14** (2009), 2287–2309.

[28] R. O'Donnell, R. A. Servedio, and L.-Y. Tan, Fooling polytopes. In *Proceedings of the 51st annual ACM SIGACT symposium on theory of computing*, pp. 614–625, ACM, 2019.

[29] M. Rudelson and R. Vershynin, The Littlewood–Offord problem and invertibility of random matrices. *Adv. Math.* **218** (2008), no. 2, 600–633.

[30] M. Rudelson and R. Vershynin, Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.* **62** (2009), no. 12, 1707–1739.

[31] D. Slepian, The one-sided barrier problem for Gaussian noise. *Bell Syst. Tech. J.* **41** (1962), no. 2, 463–501.

[32] V. N. Sudakov, Gaussian random processes and measures of solid angles in Hilbert space. *Dokl. Akad. Nauk* **197** (1971), 43–45.

[33] V. N. Sudakov, Geometric problems of the theory of infinite-dimensional probability distributions. *Tr. Mat. Inst. Steklova* **141** (1976), 3–191.

[34] M. Talagrand, *Spin glasses: a challenge for mathematicians: cavity and mean field models* 46, Springer, 2003.

[35] D. Zhang and W. B. Wu, Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** (2017), no. 5, 1895–1919.

[36] X. Zhang and G. Cheng, Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli* **24** (2018), 2640–2675.

[37] M. Zhilova, Nonclassical Berry–Esseen inequalities and accuracy of the bootstrap. *Ann. Statist.* **48** (2020), no. 4, 1922–1939.

CUN-HUI ZHANG

Department of Statistics, Hill Center, Busch Campus, Rutgers University, Piscataway, NJ 08854, USA, czhang@stat.rutgers.edu