Multivariate Normal Approximation by Stein’s Method: The Concentration Inequality Approach

Louis H.Y. Chen* and Xiao Fang*

Department of Mathematics
National University of Singapore
10, Lower Kent Ridge Road
Singapore 119076
Republic of Singapore
e-mail: matchyl@nus.edu.sg

Department of Statistics and Applied Probability
National University of Singapore
6 Science Drive 2
Singapore 117546
Republic of Singapore
e-mail: fangwd2003@gmail.com

Abstract: The concentration inequality approach for normal approximation by Stein’s method is generalized to the multivariate setting. This approach is used to prove a multivariate normal approximation theorem for standardized sums of independent random vectors with an error bound of the order $k^{1/2}\gamma$, where $k$ is the dimension of the random vectors and $\gamma$ is the sum of absolute third moments of the random vectors.

AMS 2000 subject classifications: Primary 60F05; secondary 60B10.

Keywords and phrases: Concentration inequality, Multivariate normal approximation, Stein’s method.

1. Introduction

Since Stein introduced his method for normal approximation in 1972, much has been developed for normal approximation in one dimension for dependent random variables for both smooth and non-smooth functions. A typical non-
smooth function is the indicator of a half line. Three approaches have been developed to deal with non-smooth functions: the induction approach popularized by Bolthausen (1984), the recursive approach of Raïč (2003) and the concentration inequality approach developed by Chen (1986), Chen (1998), Chen and Shao (2001) and Chen and Shao (2004).

Although Stein’s method has been extended to multivariate normal approximation (see, for example, Barbour (1990), Götze (1991), Goldstein and Rinott (1996), Chatterjee and Meckes (2008), Reinert and Röllin (2009)), relatively few results have been obtained for non-smooth functions, typically for indicators of convex sets in finite dimensional Euclidean spaces. In general, it is much harder to obtain optimal bounds for non-smooth functions than for smooth functions. As far as we know, results for non-smooth functions are those of Götze (1991), Reinert and Rotar (1996) and Bhattacharya and Holmes (2010), which is an exposition of Götze’s result. While the result of Reinert and Rotar (1996) is for bounded locally dependent random vectors, those of Götze (1991) and of Bhattacharya and Holmes (2010) are for independent random vectors with finite third moments. The approach of Götze (1991) and of Bhattacharya and Holmes (2010) is by induction.

In this paper, we extend the concentration inequality approach to the multivariate setting. We prove that for $W = \sum_{i=1}^{n} X_i$ being a sum of independent random vectors, standardized so that $\mathbb{E}W = 0$, $\mathbb{E}WW^T = I_{k \times k}$,

$$P(W^{(i)} \in A^{4\gamma + |A^{4\gamma}|} \setminus A^{4\gamma}) \leq 4.1k^{1/2} \epsilon + 39k^{1/2} \gamma$$

(1.1)

and with $| \cdot |$ denoting the Euclidean norm of a vector,

$$P(W \in A^{4\gamma + |X_i|} \setminus A^{4\gamma}) \leq 4.1k^{1/2} \mathbb{E}|X_i| + 39k^{1/2} \gamma$$

(1.2)

where $A$ is a convex set in $\mathbb{R}^k$, $A^\epsilon = \{x \in \mathbb{R}^k : d(x, A) \leq \epsilon\}$ for $\epsilon > 0$, $W^{(i)} = W - X_i$ and $\gamma = \sum_{i=1}^{n} \mathbb{E}|X_i|^3$. Using these concentration inequalities, we prove a normal approximation theorem for $W$ with an error bound of the order $k^{1/2} \gamma$. This dependence of $k^{1/2}$ on the dimension is better than $k^{5/2}$ and $k^{3/2}$ obtained by Bhattacharya and Holmes (2010) and $k$ as stated in Götze (1991). Our concentration inequality approach provides a new way of dealing with dependent random vectors, for example, those under local dependence, for which the induction approach is not likely to be applicable.
Comparing our result with those assuming finite third moments and using other methods in the literature, only the result of Bentkus (2003) gives a bound depending on $k^{1/4}$, which is better than $k^{1/2}$. But his result is for i.i.d. random vectors. Other results for i.i.d. random vectors, for example, by Nagaev (1976), Senatov (1980) and Sazonov (1981) depend on $k$.

This paper is organized as follows. In section 2, we develop techniques for the concentration inequality approach in the multivariate setting. In section 3, we use the concentration inequality approach to prove a multivariate normal approximation theorem for sums of independent random vectors. In section 4, we prove the technical lemmas in Section 2.

Throughout the paper, let $|\cdot|$ denote the Euclidean norm of vectors, and let $\|\cdot\|$ denote the operator norm of matrices. Let $\partial_j f$ denote the first partial derivative of $f$ along the coordinate $j$. For a positive integer $k$, $[k] = \{1, 2, \ldots, k\}$.

Finally, let $I_{k \times k}$ denote the $k$ by $k$ identity matrix.

2. Concentration inequalities

As a powerful tool of proving distributional approximations along with error bounds, the theory of Stein’s method has been extensively developed in the literature for random variables with all kinds of dependence structure. While it works well for smooth function distances, it requires much more efforts to obtain optimal bounds for non-smooth function distances such as the Kolmogorov distance. To overcome this difficulty, we consider the probability for some random variable $W$ taking values in a small interval $[a, b]$. A bound on $P(W \in [a, b])$ is called a concentration inequality. Now if $W$ is a $k$-dimensional random vector and $Z$ is a $k$-dimensional standard Gaussian random vector, the non-smooth function distance between $L(W)$ and $L(Z)$ usually means $\sup_{A \subseteq \bar{A}} |P(W \in A) - P(Z \in A)|$ where $A$ denotes the set of all convex sets in $\mathbb{R}^k$. A concentration inequality in this setting would be a bound on $P(W \in A^c \setminus \bar{A})$ where $A^c = \{x \in \mathbb{R}^k : d(x, A) \leq \epsilon\}$ where $d(x, A) = \inf_{y \in A} |x - y|$.

For a given convex set $A \subseteq \mathbb{R}^k$, $\epsilon > 0$, we define $f = f(A, \epsilon) = (f_1, f_2, \ldots, f_k)^T : \mathbb{R}^k \to \mathbb{R}^k$ as follows. For $x \in \bar{A}$ where $\bar{A}$ is the closure of $A$, $f(x) = 0$. For $x \in A^c \setminus \bar{A}$, find $x_0$ the nearest point in $\bar{A}$ from $x$, and define $f(x) = x - x_0$. For $x \in \mathbb{R}^k \setminus A^c$, find $x_0$ the nearest point in $\bar{A}$ from $x$, and $x_1$ the intersec-
tion of \( \{x_0 + t(x - x_0) : t \in [0, 1] \} \) and \( \partial A^c \), the boundary of \( A^c \), and define \( f(x) = x_1 - x_0 = f(x_1) \). We have the following four lemmas regarding to the properties of the above defined \( f \).

**Lemma 2.1.** We have

\[
|f| \leq \epsilon. \tag{2.1}
\]

**Lemma 2.2.** For all \( \xi, \eta \in \mathbb{R}^k \),

\[
\xi \cdot (f(\eta + \xi) - f(\eta)) \geq 0. \tag{2.2}
\]

**Lemma 2.3.** For every \( i \in [k] \) and any fixed \( x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n \), \( f_i \) is absolutely continuous in \( x_i \) and

\[
\partial_i f_i(x) \geq 0 \quad \text{a.e.} \tag{2.3}
\]

For \( x \in (A^c)^o \setminus \bar{A} \), where \( A^o \) is the interior of \( A \), we have a shaper lower bound for \( \partial_i f_i(x) \). Let \( \theta = (\theta_1, \theta_2, \ldots, \theta_k)^T \) be the angles between \( x - x_0 \) and the axes.

**Lemma 2.4.** For all \( i \in [k] \), \( x \in (A^c)^o \setminus \bar{A} \),

\[
\partial_i f_i(x) \geq \cos^2 \theta_i \quad \text{a.e.} \tag{2.4}
\]

We defer the proofs of the lemmas to Section 4. To obtain a concentration inequality for a random vector \( W \) of interest, we apply the above defined function \( f \) in the Stein identity for \( W \). We consider the following two cases: multivariate Gaussian vectors and sums of independent random vectors.

### 2.1. Multivariate normal distribution

**Proposition 2.5.** Let \( Z = (Z_1, Z_2, \ldots, Z_k)^T \) be a \( k \)-dimensional standard Gaussian random vector. Then for any convex set \( A \) in \( \mathbb{R}^k \) and \( \epsilon_1, \epsilon_2 \geq 0 \),

\[
P(Z \in A^{\epsilon_1} \setminus A^{-\epsilon_2}) \leq k^{1/2}(\epsilon_1 + \epsilon_2) \tag{2.5}
\]

where \( A^c = \{ x \in \mathbb{R}^k : d(x, A) \leq \epsilon \} \) and \( A^{-\epsilon} = \{ x \in \mathbb{R}^k : B(x, \epsilon) \subset A \} \) where \( B(x, \epsilon) \) is the \( k \)-dimensional ball centered in \( x \) with radius \( \epsilon \).
Proof. From the joint independence among \( \{Z_1, Z_2, \ldots, Z_k\} \) and the integration by parts formula, we have the following \( k \) functional identities for \( Z \).

\[
EZ_1f_1(Z) = E\partial_1f_1(Z),
\]

\[
\vdots
\]

\[
EZ_kf_k(Z) = E\partial_kf_k(Z).
\]

(2.6)

Using the function \( f = f(A, \epsilon) \) defined at the beginning of this section where \( A \) is a convex set in \( \mathbb{R}^k \) and \( \epsilon > 0 \) and summing up the above \( k \) equations, we have

\[
\sum_{j=1}^{k} EZ_jf_j(Z) = \sum_{j=1}^{k} E\partial_jf_j(Z).
\]

(2.7)

By Lemma 2.1, LHS of (2.7) \( \leq k^{1/2} \epsilon \). By Lemma 2.3 and Lemma 2.4, RHS of (2.7)

\[
\geq E \sum_{j=1}^{k} \cos^2 \theta_j I(Z \in (A^\epsilon)^c \setminus A) = P(Z \in (A^\epsilon)^c \setminus A).
\]

(2.8)

Therefore,

\[
P(Z \in A^\epsilon \setminus A) \leq k^{1/2} \epsilon.
\]

(2.9)

The bound (2.5) can be deduced from the above inequality by the arguments in Section 1.3 of Bhattacharya and Rao (1986) sketched as follows.

Without loss of generality, assume \( A^\epsilon \neq \emptyset \). First suppose \( A \) is bounded. Given any \( \delta > 0 \), we may choose \( x_1, x_2, \ldots, x_n \in \partial A \) such that \( \partial A \subset \{x_1, \ldots, x_n\}^\delta \). Let \( P \) be the convex hull of \( \{x_1, \ldots, x_n\} \). By taking \( \delta \) small enough, \( P^\delta \neq \emptyset \).

For some positive integer \( m \), \( P \) can be expressed as

\[
P = \{x \in \mathbb{R}^k : u_j \cdot x \leq d_j, 1 \leq j \leq m\}
\]

where \( u_j \)'s are distinct unit vectors and \( d_j \)'s are real numbers. For each real \( a \), define

\[
P_a = \{x \in \mathbb{R}^k : u_j \cdot x \leq d_j + a, 1 \leq j \leq m\}.
\]

Then from the fact that \( P \subset A \subset P^\delta \), we have

\[
A^{\epsilon_1} \setminus A^{-\epsilon_2} \subset (P^\delta)^{\epsilon_1} \setminus P^{-\epsilon_2} \subset P_{\epsilon_1 + \delta} \setminus P_{-\epsilon_2}.
\]
Therefore,
\[ P(Z \in A') \leq P(Z \in P_{\epsilon_1+\delta} \setminus P_{\epsilon_2}) = \int_{-\epsilon_2}^{\epsilon_1+\delta} \int_{\partial P_a} \phi d\lambda_{k-1} da \tag{2.10} \]
where \( \phi \) is the density of standard \( k \)-dimensional normal distribution and \( \lambda_{k-1} \) is the Lebesgue measure in \( \mathbb{R}^{k-1} \). We used Lemma 3.9 in Bhattacharya and Rao (1986) in the last equality. From the arguments leading to (3.35) in Bhattacharya and Rao (1986),
\[ |P(Z \in (P_a)^{\epsilon_1} \setminus P_a) - \epsilon \int_{\partial P_a} \phi d\lambda_{k-1}| \leq o(\epsilon). \]
The above inequality and (2.9) result in
\[ \int_{\partial P_a} \phi d\lambda_{k-1} \leq k^{1/2}. \]
Therefore, from (2.10),
\[ P(Z \in A') \leq k^{1/2}(\epsilon_1 + \epsilon_2 + \delta). \]
The bound (2.5) is proved by letting \( \delta \to 0 \). If \( A \) is unbounded, consider \( A_r = A \cap B(0,r) \) and let \( r \to \infty \).

\[ \square \]

Remark 2.6. It is known that \( P(Z \in A^\epsilon \setminus A^{-\epsilon_2}) \leq 4k^{1/4}(\epsilon_1 + \epsilon_2) \), which is of optimal order in \( k \) (see Ball (1993) and Bentkus (2003)). It is not clear how we can obtain \( k^{1/4} \) in the bound by our approach.

2.2. Sum of independent random vectors

Proposition 2.7. Let \( k \)-dimensional random vector \( W \) be
\[ W = (W_1, \ldots, W_k)^T = \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} (X_{i1}, X_{i2}, \ldots, X_{ik})^T \]
where \( \{X_i : i \in [n]\} \) are independent random vectors such that \( \mathbb{E}X_i = 0 \) and \( \mathbb{E}W^T = I_{k \times k} \). Then, for any convex set \( A \) in \( \mathbb{R}^k \),
\[ P(W^{(i)} \in A^\gamma \setminus A^{-\gamma}) \leq 4.1k^{1/2}\epsilon + 39k^{1/2}\gamma \tag{2.11} \]
and
\[ P(W \in A^\gamma + |X_i| \setminus A^\gamma) \leq 4.1k^{1/2}\mathbb{E}|X_i| + 39k^{1/2}\gamma \tag{2.12} \]
for any \( \epsilon > 0 \) and \( i \in [n] \) where \( W^{(i)} = W - X_i \) and \( \gamma = \sum_{i=1}^{n} \gamma_i = \sum_{i=1}^{n} \mathbb{E}|X_i|^3 \).

\[ \text{imart-generic ver. 2010/09/07 file: mvn_indep.tex date: November 18, 2011} \]
Proof. With out loss of generality, assume $\gamma$ is finite. In this proof, let $\sum_{j' \neq j''}$ denote $\sum_{j'=1}^{n} \sum_{j''=n+1}^{N}$, and for a fixed $i$, let $\sum_{j \neq i}$ denote $\sum_{j \leq n, j \neq i}$. We use $f = f(A, \epsilon + 8\gamma)$ defined at the beginning of this section in the following Stein identity for $W^{(i)}$.

$$\mathbb{E}W^{(i)} \cdot f(W^{(i)}) = \sum_{j \neq i} \mathbb{E}X_j \cdot (f(W^{(i)}) - f(W^{(i)} - X_j)).$$ \hspace{1cm} (2.13)

Because $|f| \leq \epsilon + 8\gamma$, LHS of (2.13) $\leq k^{1/2}(\epsilon + 8\gamma)$. From Lemma 2.2,

RHS of (2.13)

$$\geq \sum_{j \neq i} \mathbb{E}X_j \cdot (f(W^{(i)}) - f(W^{(i)} - X_j))I(|X_j| \leq 4\gamma)I(W^{(i)} \in A^{\epsilon + 4\gamma} \setminus A^{4\gamma})$$

$$= \sum_{j \neq i} \mathbb{E}\{ \sum_{j''=1}^{k} (-X_j \cdot h_{jj''})(f(W^{(i)} - X_j) \cdot h_{jj''} - f(W^{(i)}) \cdot h_{jj''}) \} \times I(|X_j| \leq 4\gamma)I(W^{(i)} \in A^{\epsilon + 4\gamma} \setminus A^{4\gamma})$$

where we used the orthonormal basis $\{h_{j1}, \ldots, h_{jk}\}$ for each $j \neq i$ defined as follows. For each $W^{(i)} = w^{(i)} \in A^{\epsilon + 4\gamma} \setminus A^{4\gamma}$ and $X_j = x_j$, define an orthonormal basis $\{h_{j1}, \ldots, h_{jk}\}$ such that $h_{j1}$ and $w^{(i)} - w^{(i)}_0$ are parallel and $h_{j2}$ and $-x_j - (-x_j \cdot h_{j1})h_{j1}$ are parallel (0-vector is parallel to any vector). Recall that $w^{(i)}_0$ is the nearest point in $A$ from $w^{(i)}$. Then,

RHS of (2.13)

$$\geq \sum_{j \neq i} \mathbb{E}\{ (-X_j \cdot h_{j1})(f(W^{(i)} - (-X_j \cdot h_{j1})h_{j1}) \cdot h_{j1} - f(W^{(i)}) \cdot h_{j1})$$

$$+ (-X_j \cdot h_{j1})(f(W^{(i)} - X_j) \cdot h_{j1} - f(W^{(i)} + (-X_j \cdot h_{j1})h_{j1}) \cdot h_{j1})$$

$$+ (-X_j \cdot h_{j2})(f(W^{(i)} + (-X_1 \cdot h_{j1})h_{j1}) \cdot h_{j2} - f(W^{(i)}) \cdot h_{j2})$$

$$+ (-X_j \cdot h_{j2})(f(W^{(i)} - X_j) \cdot h_{j2} - f(W^{(i)} + (-X_j \cdot h_{j1})h_{j1}) \cdot h_{j2}) \} \times I(|X_j| \leq 4\gamma)I(W^{(i)} \in A^{\epsilon + 4\gamma} \setminus A^{4\gamma}).$$

If $w^{(i)} \in A^{\epsilon + 4\gamma} \setminus A^{4\gamma}$, $|x_j| \leq 4\gamma$, then we have

$$f(w^{(i)} + (-x_j \cdot h_{j1})h_{j1}) \cdot h_{j1} - f(w^{(i)}) \cdot h_{j1} = -x_j \cdot h_{j1}, \hspace{1cm} (2.14)$$

$$f(w^{(i)} + (-x_j \cdot h_{j1})h_{j1}) \cdot h_{j2} - f(w^{(i)}) \cdot h_{j2} = 0 \hspace{1cm} (2.15)$$

and

$$(-x_j \cdot h_{j2})(f(w^{(i)} - x_j) \cdot h_{j2} - f(w^{(i)} + (-x_j \cdot h_{j1})h_{j1}) \cdot h_{j2})$$

$$\geq (f(w^{(i)} - x_j) \cdot h_{j1} - f(w^{(i)} + (-x_j \cdot h_{j1})h_{j1}) \cdot h_{j1})^2. \hspace{1cm} (2.16)$$
Equations (2.14) and (2.15) follow from \( f(w(i) + (−x_j \cdot h_{j1})h_{j1}) = f(w(i)) + (−x_j \cdot h_{j1})h_{j1} \). For (2.16), consider the plane \( p \) parallel to \( h_{j1}, h_{j2} \) and containing \( w(i) \). Let \( l \) be the line parallel to \( h_{j2} \) and containing \( w(i) \). The line \( l \) divides \( p \) into two parts \( p_1, p_2 \) where \( p_1 \) is closed and \( p_2 \) is open and contains \( W(i) \). Draw a circle on \( p \) with diameter \([w(i), w(i) − x_j]\). Then \((w(i) − x_j)'\), the projection of \((w(i) − x_j)_0 \) on \( p \), must be inside the circle (or on the perimeter) and on \( p_1 \) because of the convexity of \( A \). Let \((w(i) − x_j)''\) be the projection of \((w(i) − x_j)\) on \( l \), and let \((w(i) − x_j)'''\) be the projection of \((w(i) − x_j)'\) on \( l \). Then, (2.16) follows from

\[
|((w(i) − x_j)'' − w(i))(w(i) − x_j)''' − w(i))| ≥ |(w(i) − x_j)' − (w(i) − x_j)''|^2,
\]

which is a consequence of the fact that the angle between \((W(i) − X_j)'' - (W(i) − X_j)'\) and \(W_0(i) − (W(i) − X_j)'\) is greater than or equal to \( \pi/2 \). Using \( ab ≥ a^2 − b^2/4 \),

\[
(−x_j \cdot h_{j1})(f(w(i) − x_j) \cdot h_{j1} − f(w(i) + (−x_j \cdot h_{j1})h_{j1}) \cdot h_{j1}) ≥ \frac{−(−x_j \cdot h_{j1})^2}{4} − (f(w(i) − x_j) \cdot h_{j1} − f(w(i) + (−x_j \cdot h_{j1})h_{j1}) \cdot h_{j1})^2.
\]

Apply (2.14)-(2.17), we obtain a lower bound of RHS of (2.13) as

\[
\text{RHS of (2.13)} ≥ \frac{3}{4} \sum_{j \neq i} E(−X_j \cdot h_{j1})^2 I(|X_j| ≤ 4\gamma) I(W(i) ∈ A^{−4\gamma} \setminus A^{4\gamma}).
\]

(2.18)

In other words, we have

\[
\text{RHS of (2.13)} ≥ \frac{3}{4} \sum_{j \neq i} E(X_j \cdot \xi(W(i)))^2 I(|X_j| ≤ 4\gamma) I(W(i) ∈ A^{4\gamma} \setminus A^{−4\gamma})
\]

\[= R\]

(2.19)

where \( \xi(W(i)) = (W_0(i) − W(i))/|W_0(i) − W(i)| \) for \( W(i) ∈ A^{4\gamma} \setminus A^{−4\gamma} \) and \( W_0(i) \) is the nearest point in \( \tilde{A} \) from \( W(i) \). We may define \( \xi(W(i)) \) to be \( e_1 \), where \( \{e_1, \ldots, e_k\} \) are the original orthonormal basis when \( W(i) \notin A^{4\gamma} \setminus A^{−4\gamma} \), since
it does not affect the value of $R$. We now obtain a lower bound of $R$.

$$R = \frac{3}{4} \sum_{j \neq i}^k \sum_{j' = 1} X_{jj'}^2 \xi(W^{(i)})_{j'}^2 I(|X_j| \leq 4\gamma) I(W^{(i)} \in A^{\gamma+4\gamma} \setminus A^{4\gamma})$$

$$+ \frac{3}{4} \sum_{j' \neq j''}^k \sum_{j' \neq i} X_{jj'} X_{j''} \xi(W^{(i)})_{j'} \xi(W^{(i)})_{j''} I(|X_j| \leq 4\gamma) I(W^{(i)} \in A^{\gamma+4\gamma} \setminus A^{4\gamma})$$

$$= R_1 + R_2.$$ 

For $R_1$,

$$R_1 = \frac{3}{4} \sum_{j' = 1}^k E \{ I(W^{(i)} \in A^{\gamma+4\gamma} \setminus A^{4\gamma}) \xi(W^{(i)})_{j'}^2 \sum_{j \neq i} X_{jj'}^2 I(|X_j| \leq 4\gamma)$$

$$+ \frac{3}{4} \sum_{j' = 1}^k \sum_{j \neq i} X_{jj'}^2 (\gamma \sum_{j' \neq i} X_{jj'} I(|X_j| \leq 4\gamma)$$

$$+ \frac{1}{4\gamma} \sum_{j' \neq i} \sum_{j \neq i} X_{jj'} I(|X_j| \leq 4\gamma)) \}$$

Using the inequality

$$ab \leq \gamma a^2 + \frac{b^2}{4\gamma},$$

(2.20)

$$|R_{1,1}| \leq \frac{3}{4} \sum_{j' = 1}^k \left\{ \gamma \sum_{j \neq i} X_{jj'}^2 I(|X_j| \leq 4\gamma)$$

$$+ \frac{1}{4\gamma} \sum_{j' \neq i} \sum_{j \neq i} X_{jj'} I(|X_j| \leq 4\gamma) \right\}^2$$

$$= \frac{3}{4} (\gamma \sum_{j' = 1}^k \sum_{j \neq i} \xi(W^{(i)})_{j'}^2 + \frac{1}{4\gamma} \sum_{j' = 1}^k \sum_{j \neq i} \text{Var}(\sum_{j \neq i} X_{jj'} I(|X_j| \leq 4\gamma)))$$

$$\leq \frac{3}{4} (\gamma \sum_{j' = 1}^k \sum_{j \neq i} \xi(W^{(i)})_{j'}^2 + \frac{1}{4\gamma} \sum_{j' = 1}^k \sum_{j \neq i} \sum_{j \neq i} [\xi(W^{(i)})_{j'}^2 I(|X_j| \leq 4\gamma)]).$$

For $R_{1,2}$,

$$R_{1,2} = \frac{3}{4} \sum_{j' = 1}^{k} \mathbb{E} I(W^{(i)} \in A^{c+\gamma} \setminus A^{4\gamma}) \xi(W^{(i)})_{j'}^2 \left[ \mathbb{E} \sum_{j \neq i} X_{j,j'}^2 I(|X_j| > 4\gamma) \right]$$

$$\geq \frac{3}{4} \sum_{j' = 1}^{k} \mathbb{E} I(W^{(i)} \in A^{c+\gamma} \setminus A^{4\gamma}) \xi(W^{(i)})_{j'}^2 (1 - \mathbb{E} X_{j,j'}^2)$$

$$- \frac{3}{4} \sum_{j' = 1}^{k} \mathbb{E} I(W^{(i)} \in A^{c+\gamma} \setminus A^{4\gamma}) \xi(W^{(i)})_{j'}^2 \mathbb{E} \sum_{j \neq i} |X_j|^3 \frac{1}{4\gamma}$$

$$\geq (1 - \gamma^{2/3}) \frac{3}{4} \mathbb{P}(W^{(i)} \in A^{c+\gamma} \setminus A^{4\gamma}) - \frac{3}{16} \mathbb{P}(W^{(i)} \in A^{c+\gamma} \setminus A^{4\gamma})$$

where we used the facts that $\mathbb{E}|X_j|^2 \leq \gamma^{2/3}$ and $|\xi(W^{(i)})| = 1$ in the last inequality.

$$R_2 = \frac{3}{4} \sum_{j' \neq j''} \mathbb{E} \sum_{j' \neq j''} X_{j,j'}X_{j,j''} \xi(W^{(i)})_{j'}^2 \xi(W^{(i)})_{j''}^2 I(|X_j| \leq 4\gamma) I(W^{(i)} \in A^{c+\gamma} \setminus A^{4\gamma})$$

$$= \frac{3}{4} \sum_{j' \neq j''} \mathbb{E} I(W^{(i)} \in A^{c+\gamma} \setminus A^{4\gamma}) \xi(W^{(i)})_{j'} \xi(W^{(i)})_{j''}$$

$$\times (\sum_{j \neq i} X_{j,j'}X_{j,j''} I(|X_j| \leq 4\gamma) - \mathbb{E} \sum_{j \neq i} X_{j,j'}X_{j,j''} I(|X_j| \leq 4\gamma))$$

$$+ \frac{3}{4} \sum_{j' \neq j''} \mathbb{E} I(W^{(i)} \in A^{c+\gamma} \setminus A^{4\gamma}) \xi(W^{(i)})_{j'} \xi(W^{(i)})_{j''} \mathbb{E} \sum_{j \neq i} X_{j,j'}X_{j,j''} I(|X_j| \leq 4\gamma)$$

$$= R_{2,1} + R_{2,2}$$

For $R_{2,1}$, using the inequality (2.20),

$$|R_{2,1}| \leq \frac{3}{4} \mathbb{E} I(W^{(i)} \in A^{c+\gamma} \setminus A^{4\gamma}) \sum_{j' \neq j''} |\xi(W^{(i)})_{j'} \xi(W^{(i)})_{j''}|$$

$$\times (\sum_{j \neq i} X_{j,j'}X_{j,j''} I(|X_j| \leq 4\gamma) - \mathbb{E} \sum_{j \neq i} X_{j,j'}X_{j,j''} I(|X_j| \leq 4\gamma))$$

$$\leq \frac{3}{4} \mathbb{E} \sum_{j' \neq j''} \left\{ \gamma |\xi(W^{(i)})_{j'} \xi(W^{(i)})_{j''}|^2 

+ \frac{(\sum_{j \neq i} X_{j,j'}X_{j,j''} I(|X_j| \leq 4\gamma) - \mathbb{E} \sum_{j \neq i} X_{j,j'}X_{j,j''} I(|X_j| \leq 4\gamma))^2}{4\gamma} \right\}$$

$$\leq \frac{3\gamma}{4} \sum_{j' \neq j''} \mathbb{E} |\xi(W^{(i)})_{j'} \xi(W^{(i)})_{j''}|^2 + \frac{3}{4} \times \frac{1}{4\gamma} \sum_{j' \neq j''} \sum_{j' \neq j''} \mathbb{E} (X_{j,j'}X_{j,j''})^2 I(|X_j| \leq 4\gamma).$$

From the bounds on $|R_{1,1}|$ and $|R_{2,1}|$,

$$|R_{1,1}| + |R_{2,1}| \leq \frac{3\gamma}{4} + \frac{3}{16\gamma} \mathbb{E} \sum_{j \neq i} |X_j|^4 I(|X_j| \leq 4\gamma) \leq \frac{3\gamma}{2}.$$
A lower bound of $R_{2,2}$ can be obtained as follows. Let $\widetilde{W}^{(i)}$ be an independent copy of $W^{(i)}$.

\[ R_{2,2} \]

\[
\begin{align*}
R_{2,2} & = \frac{3}{4} \sum_{j' \neq j''} \mathbb{E}I(W^{(i)} \in A^{e+4\gamma} \setminus A^{4\gamma}) \xi(W^{(i)})_{j'} \xi(W^{(i)})_{j''} \\
& \times \left[ \mathbb{E}\sum_{j \neq i} X_{j'j}X_{j''j} - \mathbb{E}\sum_{j \neq i} X_{j'j}X_{j''j}I(|X_j| > 4\gamma) \right] \\
& \geq \frac{3}{4} \mathbb{E}|X|^2 \mathbb{P}(W^{(i)} \in A^{e+4\gamma} \setminus A^{4\gamma}) \\
& - \frac{3}{4} \sum_{j' \neq j''} \sum_{j \neq i} \mathbb{E}I(\widetilde{W}^{(i)} \in A^{e+4\gamma} \setminus A^{4\gamma}) |X_{j'j}X_{j''j}| \mathbb{E}\tilde{\xi}(W^{(i)})_{j'} \mathbb{E}\tilde{\xi}(W^{(i)})_{j''}I(|X_j| > 4\gamma) \\
& \geq \frac{3}{4} 2^{2/3} \mathbb{P}(W^{(i)} \in A^{e+4\gamma} \setminus A^{4\gamma}) - \frac{3}{4} \mathbb{E}I(\widetilde{W}^{(i)} \in A^{e+4\gamma} \setminus A^{4\gamma}) \sum_{j \neq i} |X_j|^2 I(|X_j| > 4\gamma) \\
& \geq \frac{3}{4} 2^{2/3} \mathbb{P}(W^{(i)} \in A^{e+4\gamma} \setminus A^{4\gamma}) - \frac{3}{16} \mathbb{P}(W^{(i)} \in A^{e+4\gamma} \setminus A^{4\gamma})
\end{align*}
\]

where we used the facts that $\mathbb{E}\sum_{j \neq i} X_{j'j}X_{j''j} = -\mathbb{E}X_{j'j}X_{j''j}$ for $j' \neq j''$ and $\sum_{j' = 1}^k |X_{j'j}X_{j''j}|(\mathbb{E}f(W^{(i)})_{j'} \mathbb{E}f(W^{(i)})_{j''}) \leq |X_j|$. Therefore,

RHS of (2.13)

\[
\geq (1 - \gamma^{2/3}) \frac{3}{4} \mathbb{P}(W^{(i)} \in A^{e+4\gamma} \setminus A^{4\gamma}) - \frac{3}{16} \mathbb{P}(W^{(i)} \in A^{e+4\gamma} \setminus A^{4\gamma}) - \frac{3\gamma}{2}
\]

Recall that LHS of (2.13) $\leq k^{1/2}(\epsilon + 8\gamma)$, we have

\[
\left( \frac{3}{8} - \frac{3}{2} \gamma^{2/3} \right) \mathbb{P}(W^{(i)} \in A^{e} \setminus A) \\
\leq k^{1/2} \epsilon + 8k^{1/2} \gamma + \frac{3}{2} \gamma.
\]

(2.21)

When $\gamma > 1/39$, (2.11) is true. When $\gamma \leq 1/39$, (2.11) is obtained by solving (2.21).

To prove (2.12), let $f^{X_i} = f(A, |X_i| + 8\gamma)$ be defined at the beginning of this section. Consider the following Stein identity,

\[
\mathbb{E}W^{(i)} \cdot f^{X_i}(W) = \sum_{j \neq i} \mathbb{E}X_j \cdot (f^{X_i}(W) - f^{X_i}(W - X_j)).
\]
We have
\[
\mathbb{E}|W^{(i)}(|X_i| + 8\gamma) \\
\geq \sum_{j \neq i} \mathbb{E}X_j \cdot (f^{X_i}(W) - f^{X_i}(W - X_j))I(W \in A^{4\gamma + |X_i| \setminus A^{4\gamma})I(|X_j| \leq 4\gamma).
\]

The bound (2.12) can be proved by applying the same argument leading to (2.11).

\[\square\]

3. Multivariate normal approximation

In this section, we prove a multivariate normal approximation result (Theorem 3.5) by applying the concentration inequality approach in Stein’s method. A multivariate version of the Stein equation was given in Götze (1991) as well as in Barbour (1990) as follows.

\[\Delta f(w) - w \cdot \nabla f(w) = h(w) - \mathbb{E}h(Z) \quad (3.1)\]

where \(h\) is a test function and \(Z\) is a standard \(k\)-dimensional Gaussian random vector.

If the test function \(h\) is smooth enough, the above equation can be solved and one of its solution can be expressed as

\[f(w) = -\frac{1}{2} \int_0^1 \frac{1}{1-s} \int_{\mathbb{R}^k} [h(\sqrt{1-s}w + \sqrt{s}z) - \mathbb{E}h(Z)]\phi(z)dz ds \quad (3.2)\]

where \(\phi(z)\) is the density function of the \(k\)-dimensional standard normal distribution at \(z \in \mathbb{R}^k\). When \(\nabla h\) is Lipschitz, the second derivatives of \(f\) can be calculated as

\[\partial_{jj'} f(w) = -\frac{1}{2} \int_0^1 \frac{1}{1-s} \int_{\mathbb{R}^k} h(\sqrt{1-s}w + \sqrt{s}z)\partial_{jj'} \phi(z)dz ds + \frac{1}{2} \int_0^1 \frac{1}{\sqrt{s}} \int_{\mathbb{R}^k} \partial_j h(\sqrt{1-s}w + \sqrt{s}z)\partial_{j'} \phi(z)dz ds \quad (3.3)\]

For each test function \(h = I_A\) where \(A\) is a convex set in \(\mathbb{R}^k\), a smoothed version of it was introduced by Bentkus (2003)

\[h_\epsilon(w) = \psi\left(\frac{d(w, A)}{\epsilon}\right) \quad (3.4)\]
where \( \epsilon > 0 \) and function \( \psi \) is defined as

\[
\psi(x) = \begin{cases} 
1, & x < 0 \\
1 - 2x^2, & 0 \leq x < \frac{1}{2} \\
2(1 - x)^2, & \frac{1}{2} \leq x < 1 \\
0, & 1 \leq x.
\end{cases}
\] (3.5)

The next lemma was proved in Bentkus (2003).

**Lemma 3.1.** The above defined function \( h_\epsilon \) satisfies:

\[
h_\epsilon(w) = 1 \text{ for } w \in A, \quad h_\epsilon(w) = 0 \text{ for } w \in \mathbb{R}^k \setminus A^\epsilon, \quad 0 \leq h_\epsilon \leq 1,
\] (3.6)

and

\[
|\nabla h_\epsilon(w)| \leq \frac{2}{\epsilon}, \quad |\nabla h_\epsilon(w_1) - \nabla h_\epsilon(w_2)| \leq \frac{8|w_1 - w_2|}{\epsilon^2}.
\] (3.7)

For a convex set \( A \) and \( \gamma \geq 0 \), defining \( g_{1,\epsilon} = h_\epsilon \) for \( h = I_{A^\epsilon} \), we have

\[
P(W \in A) - P(Z \in A) \leq P(W \in A^{4\gamma}) - P(Z \in A)
\]

\[
\leq E_{g_{1,\epsilon}}(W) - E_{g_{1,\epsilon}}(Z) + E_{g_{1,\epsilon}}(Z) - P(Z \in A)
\]

\[
\leq E_{g_{1,\epsilon}}(W) - E_{g_{1,\epsilon}}(Z) + P(Z \in A^{4\gamma} \setminus A)
\]

\[
\leq E_{g_{1,\epsilon}}(W) - E_{g_{1,\epsilon}}(Z) + k^{1/2}(4\gamma + \epsilon)
\]

where we used (2.5). If \( A^{-4\epsilon-4\gamma} = \emptyset \),

\[
P(W \in A) - P(Z \in A) \geq -P(Z \in A \setminus A^{-4\epsilon-4\gamma}) \geq -k^{1/2}(4\gamma + \epsilon).
\]

If not, defining \( g_{2,\epsilon} = h_\epsilon \) for \( h = I_{(A^{-4\epsilon-4\gamma})^c} \), we have

\[
P(W \in A) - P(Z \in A) \geq E_{g_{2,\epsilon}}(W) - E_{g_{2,\epsilon}}(Z) + E_{g_{2,\epsilon}}(Z) - P(Z \in A)
\]

\[
\geq E_{g_{2,\epsilon}}(W) - E_{g_{2,\epsilon}}(Z) + P(Z \in A \setminus A^{-4\epsilon-4\gamma})
\]

\[
\geq E_{g_{2,\epsilon}}(W) - E_{g_{2,\epsilon}}(Z) - k^{1/2}(4\gamma + \epsilon).
\]

Therefore, we have the following smoothing lemma.

**Lemma 3.2.** For any \( k \)-dimensional random vector \( W \),

\[
\sup_{A \in \mathcal{A}} |P(W \in A) - P(Z \in A)| \leq \sup_{h = I_{A^{4\gamma}}, A \in \mathcal{A}} |E_{h_\epsilon}(W) - E_{h_\epsilon}(Z)| + k^{1/2}(\epsilon + 4\gamma)
\] (3.8)

where \( Z \) is a standard \( k \)-dimensional Gaussian random vector, \( \mathcal{A} \) is the set of all the convex sets in \( \mathbb{R}^k \), \( \epsilon > 0 \), \( \gamma \geq 0 \) and \( h_\epsilon \) is defined as in (3.4).
The following lemma from Bentkus (2003) will be used in this section.

**Lemma 3.3.** For a $k$-dimensional vector $x$,

$$
\int_{\mathbb{R}^k} |x_j \partial_j \phi(z)| dz \leq \sqrt{\frac{2}{\pi}} |x|, \quad (3.9)
$$

$$
\int_{\mathbb{R}^k} |x_j x_j' x_j'' \partial_{jj'j''} \phi(z)| dz \leq 2 \frac{1 + 4e^{-3/2}}{\sqrt{2\pi}} |x|^3. \quad (3.10)
$$

Using the same argument as in Bentkus (2003) when proving Lemma 3.3, we obtain the following lemma.

**Lemma 3.4.** For $k$-dimensional vectors $u, v$, we have

$$
\int_{\mathbb{R}^k} |u_j v_j' v_j'' \partial_{jj'j''} \phi(z)| dz \leq 2(1 + \sqrt{\frac{2}{\pi}}) ||u|| ||v||^2. \quad (3.11)
$$

**Proof.** It is straightforward to verify that

$$
\sum_{j,j',j''=1}^k u_j v_j' v_j'' \partial_{jj'j''} \phi(z) = (||v||^2 (u \cdot z) + 2(u \cdot v)(v \cdot z) - (u \cdot z)(v \cdot z)^2) \phi(z). \quad (3.12)
$$

From (3.12), we only need to consider the projection of $z$ in the two-dimensional space spanned by vectors $u, v$. Therefore, the constant obtained is dimension free and the rough upper bound (3.11) is calculated as follows. Let $Z_1, Z_2$ be two independent 1-dimensional standard Gaussian variables, then

$$
\int_{\mathbb{R}^k} |u_j v_j' v_j'' \partial_{jj'j''} \phi(z)| dz \\
\leq ||u|| ||v||^2 (\mathbb{E}|3Z_1 - Z_1^3| + \mathbb{E}|Z_2(1 - Z_2^2)|) \leq 2(1 + \sqrt{\frac{2}{\pi}}) ||u|| ||v||^2.
$$

\[\square\]

**Theorem 3.5.** Let $k$-dimensional random vector $W$ be

$$
W = (W_1, \ldots, W_k)^T = \sum_{i=1}^n X_i = \sum_{i=1}^n (X_{i1}, X_{i2}, \ldots, X_{ik})^T
$$

where $\{X_i : i \in [n]\}$ are independent such that $\mathbb{E}X_i = 0$ for each $i$ and $\mathbb{E}WW^T = I_{k \times k}$. Then,

$$
\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leq 115k^{1/2} \gamma \quad (3.13)
$$
where \( A \) is the set of all the convex sets in \( \mathbb{R}^k \), \( Z \) is a standard \( k \)-dimensional Gaussian vector and \( \gamma = \sum_{i=1}^{n} \gamma_i = \sum_{i=1}^{n} \mathbb{E}|X_i|^3 \).

**Proof.** Without loss of generality, assume \( \gamma \) is finite. Let \( f_\varepsilon \) be the solution to the Stein equation \((3.1)\) with test function \( h_\varepsilon \) defined in \((3.4)\) where \( h = I_{A^\varepsilon} \), for some \( A \in \mathcal{A} \). With \( W^{(i)} = W - X_i \), we have

\[
\mathbb{E}[\Delta f_\varepsilon(W)] - \mathbb{E}W \cdot \nabla f_\varepsilon(W) = \mathbb{E}[\Delta f_\varepsilon(W)] - \sum_{i=1}^{n} \mathbb{E}X_i \cdot (\nabla f_\varepsilon(W) - \nabla f_\varepsilon(W^{(i)}))
\]

\[
= \mathbb{E}[\Delta f_\varepsilon(W)] - \sum_{i=1}^{n} \mathbb{E}X_i \cdot (\text{Hess} f_\varepsilon(W^{(i)}) X_i)
\]

\[
- \sum_{i=1}^{n} \mathbb{E}X_i \cdot (\nabla f_\varepsilon(W) - \nabla f_\varepsilon(W^{(i)}) - \text{Hess} f_\varepsilon(W^{(i)}) X_i)
\]

\[
= R_1 - R_2
\]

where

\[
R_1 = \sum_{i=1}^{n} \sum_{j,j'=1}^{k} \mathbb{E}X_{ij}X_{ij'} \mathbb{E}[\partial_{jj'} f_\varepsilon(W) - \partial_{jj'} f_\varepsilon(W^{(i)})]
\]

and

\[
R_2 = \sum_{i=1}^{n} \sum_{j,j'=1}^{k} \mathbb{E}X_{ij}X_{ij'} [\partial_{jj'} f_\varepsilon(W^{(i)}) + UX_i] - \partial_{jj'} f_\varepsilon(W^{(i)})]
\]

where \( U \) is an independent uniform random variable in \([0,1]\). From \((3.3)\), \( R_2 \) can be expressed as

\[
R_2 = \sum_{i=1}^{n} \sum_{j,j'=1}^{k} \mathbb{E}X_{ij}X_{ij'} \int_{1}^{\infty} \left( -\frac{1}{2s} \right) \int_{\mathbb{R}^k} \left[ h_\varepsilon(\sqrt{1-s}W^{(i)}) + \sqrt{1-s}U X_i + \sqrt{s}z \right]
\]

\[
- h_\varepsilon(\sqrt{1-s}W^{(i)}) + \sqrt{s}z)] \partial_{jj'} \phi(z) dz ds
\]

\[
+ \sum_{i=1}^{n} \sum_{j,j'=1}^{k} \mathbb{E}X_{ij}X_{ij'} \int_{0}^{1} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^k} \left[ \partial_{jj'} h_\varepsilon(\sqrt{1-s}W^{(i)}) + \sqrt{1-s}U X_i + \sqrt{s}z \right]
\]

\[
- \partial_{jj'} h_\varepsilon(\sqrt{1-s}W^{(i)}) + \sqrt{s}z)] \partial_{jj'} \phi(z) dz ds
\]

\[
= R_{2,1} + R_{2,2}.
\]

Introducing another independent uniform random variable \( U' \) in \([0,1]\) and using
the integration by parts formula,

\[
R_{2,1} = \sum_{i=1}^{n} \sum_{j,j'=1}^{k} \mathbb{E} U_{X_{ij}} X_{ij'} X_{ij''} \int_{r^2}^{1} \frac{\sqrt{1-s}}{2\sqrt{s}} \times \int_{R^k} h_e(\sqrt{1-s}U(i) + \sqrt{1-s}U' X_i + \sqrt{s}z) \partial_j j' j'' \phi(z) dz ds
\]

and

\[
R_{2,2} = \sum_{i=1}^{n} \sum_{j,j'=1}^{k} \mathbb{E} U_{X_{ij}} X_{ij'} X_{ij''} \int_{r^2}^{1} \frac{\sqrt{1-s}}{2\sqrt{s}} \times \int_{R^k} \partial_j j' j'' h_e(\sqrt{1-s}U(i) + \sqrt{1-s}U' X_i + \sqrt{s}z) \partial_j j'' \phi(z) dz ds
\]

We first use the concentration inequality in Proposition 2.7 to bound \(R_{2,2}\). Define any linear transform of a set to be the image of the linear transform of all the elements in the set. Notice that by (3.7) and Proposition 2.7,

\[
|\mathbb{E} U X_i (\sum_{j=1}^{k} X_{ij'} \partial_j j' \nabla h_e(\sqrt{1-s}W^{(i)} + \sqrt{s}z + \sqrt{1-s}U' X_i \cdot X_i))| \\
\leq \frac{8}{\epsilon^2} |X_i|^2 \mathbb{E} U X_i I(\sqrt{1-s}W^{(i)} \in A^c) - (\sqrt{s}z + \sqrt{1-s}U' X_i)
\leq |X_i|^2 (32.8k^{1/2} \frac{1}{\epsilon \sqrt{1-s}} + 312k^{1/2} \frac{\gamma}{\epsilon^2})
\]

Therefore,

\[
|R_{2,2}| \leq \frac{1}{2} \sum_{i=1}^{n} |X_i|^2 \int_{r^2}^{1} \frac{\sqrt{1-s}}{2\sqrt{s}} (32.8k^{1/2} \frac{1}{\epsilon \sqrt{1-s}} + 312k^{1/2} \frac{\gamma}{\epsilon^2}) \times \int_{R^k} |X_{ij'} \partial_j j' \phi(z)| dz ds
\leq \sqrt{\frac{2}{\pi}} \gamma (16.4k^{1/2} + 156k^{1/2} \frac{\gamma}{\epsilon})
\]

where we used Lemma 3.3. Next, we make use of the concentration inequality in Proposition 2.7 to bound \(R_{2,1}\) by a quantity involving \(\gamma\), \(\epsilon\) and \(\sup_{A \in A} |P(W \in A) - P(Z \in A)|\). Write \(R_{2,1} = R'_{2,1} + R''_{2,1}\) by separating the sum over \(i\) into two
parts according to $\gamma_i \leq 8 \gamma^3$ or else. Write $R'_{2,1} = R'_{2,1.1} + R'_{2,1.2}$ by subtracting a term with $W^{(i)}$ replaced by an independent $k$-dimensional standard Gaussian vector $Z$ and adding the same term, i.e.,

$$R'_{2,1.1} = \sum_{i: \gamma_i \leq 8 \gamma^3, j, j', j'' = 1}^k \mathbb{E} U X_{ij} X_{ij'} X_{ij''} \int_{\mathbb{R}^k} \frac{1}{2^{3/2}} \sqrt{1-s} \, ds$$

and

$$R'_{2,1.2} = \sum_{i: \gamma_i \leq 8 \gamma^3, j, j', j'' = 1}^k \mathbb{E} U X_{ij} X_{ij'} X_{ij''} \int_{\mathbb{R}^k} \frac{1}{2^{3/2}} \sqrt{1-s} \, ds$$

By introducing an independent copy $\tilde{X}_1$ of $X_1$, $\tilde{W} = W^{(i)} + \tilde{X}_1$ has the same distribution as $W$ and is independent of $X_1$. We have

$$\mathbb{E} U^{(i)} \{ h_\varepsilon(\sqrt{1-s} W^{(i)} + \sqrt{s} Z + \sqrt{1-s} U U' X_i)$$

$$- h_\varepsilon(\sqrt{1-s} Z + \sqrt{s} Z + \sqrt{1-s} U U' X_i) \}$$

$$\leq \mathbb{E} U^{(i)} \{ I(W^{(i)}) \leq \frac{1}{\sqrt{1-s}} (A^{4 \gamma} - \sqrt{s} Z - \sqrt{1-s} U U' X_i)$$

$$- I(Z \in \frac{1}{\sqrt{1-s}} (A^{4 \gamma} - \sqrt{s} Z - \sqrt{1-s} U U' X_i) \}$$

$$\leq \mathbb{E} U^{(i)} \{ I(W^{(i)}) + \tilde{X}_1 \leq \frac{1}{\sqrt{1-s}} (A^{4 \gamma} - \sqrt{s} Z - \sqrt{1-s} U U' X_i)$$

$$+ \frac{1}{\sqrt{1-s}} (A^{4 \gamma} - \sqrt{s} Z - \sqrt{1-s} U U' X_i)$$

$$+ I(Z \in \frac{1}{\sqrt{1-s}} (A^{4 \gamma} - \sqrt{s} Z - \sqrt{1-s} U U' X_i)$$

$$- I(Z \in \frac{1}{\sqrt{1-s}} (A^{4 \gamma} - \sqrt{s} Z - \sqrt{1-s} U U' X_i) \}.$$
such that \( \text{Cov}(W,W) = I_{k \times k} \) and the sum of absolute third moments of the summands is bounded by \( \gamma \). Using the concentration inequalities in Proposition 2.5 and Proposition 2.7, we have

\[
E[U,U',X_i] \left[ h_\epsilon(\sqrt{1-s}W^{(i)} + \sqrt{1-s}z + \sqrt{1-s}U'X_i) - h_\epsilon(\sqrt{1-s}Z + \sqrt{1-s}z + \sqrt{1-s}U'X_i) \right] \leq 4.1k^{1/2}E|X_i| + 39k^{1/2} + k^{1/2}\frac{\epsilon}{\sqrt{1-s}} + \delta_\gamma.
\]

(3.18)

After proving a lower bound in same way as proving the upper bound, we can use Lemma 3.3 to bound \( R'_{2,1,1} \) by

\[
|R'_{2,1,1}| \leq \frac{1 + 4e^{-3/2}}{\sqrt{2\pi}}(47.2k^{1/2} + k^{1/2} + \frac{\epsilon}{\sqrt{1-s}}) \sum_{i,\gamma_i \leq 8\gamma^3} E|X_i|^3.
\]

For \( R'_{2,1,2} \), using the integration by parts formula and noticing that \( \sqrt{1-s}Z + \sqrt{s}\tilde{Z} \) has the same distribution as \( Z \) where \( \tilde{Z} \) is an independent copy of standard normal \( Z \),

\[
E[X_i] \int_{s_2}^{1} \frac{\sqrt{1-s}}{2s^{3/2}} \int_{\mathbb{R}^k} h_\epsilon(\sqrt{1-s}Z + \sqrt{1-s}U'X_i)\partial_{jj',\gamma''} \phi(z)dzds = -E[X_i] \int_{s_2}^{1} \frac{\sqrt{1-s}}{2s} \int_{\mathbb{R}^k} \partial_{jj',\gamma''} h_\epsilon(\sqrt{1-s}Z + \sqrt{1-s}U'X_i)\phi(z)dzds.
\]

Therefore, by Lemma 3.3,

\[
|R'_{2,1,2}| \leq \frac{1 + 4e^{-3/2}}{3\sqrt{2\pi}} \sum_{i,\gamma_i \leq 8\gamma^3} E|X_i|^3.
\]

(3.19)

We remark that in the above calculation we used the third derivatives of \( h_\epsilon \) which does not exist. However, we can smooth \( h_\epsilon \) first then use limiting arguments to show that the final equality holds even if \( h_\epsilon \) does not have third derivatives.

Now we turn to bounding \( |R''_{2,1}| \) where

\[
R''_{2,1} = \sum_{i,\gamma_i > 8\gamma^3} \sum_{j,j',\gamma''=1} E[X_i X_{ij'} X_{ij''}] \int_{s_2}^{1} \frac{\sqrt{1-s}}{2s^{3/2}} \times \int_{\mathbb{R}^k} h_\epsilon(\sqrt{1-s}W^{(i)} + \sqrt{1-s}U'X_i + \sqrt{s}z)\partial_{jj',\gamma''} \phi(z)dzds.
\]

For each \( X_i \) such that \( \gamma_i > 8\gamma^3 \), define \( N_i \) to be the positive square root of the inverse of the matrix \( I_{k \times k} - \text{Cov}(X_i, X_i) \). Then we have the following bound on
the operator norm of \( N_i \).

\[
||N_i|| = \sqrt{||(I_{k\times k} - \text{Cov}(X_i, X_i))^{-1}||} \leq \left( \frac{1}{1 - ||\text{Cov}(X_i, X_i)||} \right)^{1/2}
\]

\[
= \left( \frac{1}{1 - \sup_{u \neq 0} u'\text{Cov}(X_i, X_i)u} \right)^{1/2} = \left( \frac{1}{1 - \sup_{u \neq 0} E(u'X_i^2)} \right)^{1/2}
\]

\[
\leq \left( \frac{1}{1 - E|X_i|^2} \right)^{1/2} \leq \left( \frac{1}{1 - \gamma_i^{2/3}} \right)^{1/2}.
\]

(3.20)

Note that

\[
N_i W^{(i)} = \sum_{i', i' \neq i} N_i X_{i'}
\]

(3.21)

is a sum of \( n \) independent random vectors (with one 0-vector) with

\[
\mathbb{E}N_i X_{i'} = 0, \quad \text{Cov}(N_i W^{(i)}, N_i W^{(i)}) = I_{k\times k}
\]

(3.22)

and

\[
\sum_{i', i' \neq i} \mathbb{E}|N_i X_{i'}|^3 \leq \frac{\gamma_i - \gamma_i}{(1 - \gamma_i^{2/3})^{3/2}} \leq \frac{\gamma_i - \gamma_i}{(1 - \gamma_i^{2/3})^{2/3}} \leq \gamma
\]

(3.23)

where we used the fact that \( \gamma_i > 8\gamma^3 \) in the last inequality. Therefore, \( N_i W^{(i)} \) can be regarded as a standardized sum of \( n \) independent random vectors with sum of absolute third moments of the summands less than \( \gamma \). We write \( R''_{2,1} \) into two parts as

\[
R''_{2,1,1} = \sum_{i' > 8\gamma^3} \sum_{j,j',j'' = 1}^{k} \mathbb{E}U_{ij}X_{ij'}X_{ij''} \int_{\mathbb{R}^k} \frac{1}{2s^{3/2}} \int \left[ h_c(\sqrt{1 - s}N_i^{-1}(N_i W^{(i)}) + \sqrt{s}Z + \sqrt{1 - s}UU'X_i)ight.
\]

\[
- h_c(\sqrt{1 - s}N_i^{-1}Z + \sqrt{s}Z + \sqrt{1 - s}UU'X_i)] \partial_{jj'} \phi(z)dzds
\]

and

\[
R''_{2,1,2} = \sum_{i' > 8\gamma^3} \sum_{j,j',j'' = 1}^{k} \mathbb{E}U_{ij}X_{ij'}X_{ij''} \int_{\mathbb{R}^k} \frac{1}{2s^{3/2}} \int \left[ h_c(\sqrt{1 - s}N_i^{-1}Z + \sqrt{s}Z + \sqrt{1 - s}UU'X_i)] \partial_{jj'} \phi(z)dzds.
\]
From
\[ \mathbb{E}^{U', U} [h_{\epsilon}(\sqrt{1 - s} N_i^{-1} (N_i W^{(i)}) + \sqrt{s} z + \sqrt{1 - s} U' U X_i)] \]
\[ - h_{\epsilon}(\sqrt{1 - s} N_i^{-1} Z + \sqrt{s} z + \sqrt{1 - s} U' U X_i)] \]
\[ \leq \mathbb{E}^{U', U} [I(N_i W^{(i)}) \in \frac{N_i}{\sqrt{1 - s}} (A^{4\gamma + \epsilon} - \sqrt{s} z - \sqrt{1 - s} U' U X_i)] \]
\[ - I(Z \in \frac{N_i}{\sqrt{1 - s}} (A^{4\gamma} - \sqrt{s} z - \sqrt{1 - s} U' U X_i))] \]
\[ \leq \delta_{\gamma} + k^{1/2} \frac{\epsilon}{\sqrt{1 - s}} ||N_i|| \]
and a similar lower bound, we have
\[ |R_{2,1,1}'| \leq \sum_{i: \gamma > S^{(y)}} \frac{1}{\sqrt{2\pi}} \mathbb{E} |X_i|^3 \left( \frac{\delta_{\gamma}}{\epsilon} + k^{1/2} \frac{1}{\sqrt{1 - \gamma^{2/3}}} \right). \]
Therefore,
\[ |R_{2,1,1}'| \leq \frac{1}{\sqrt{2\pi}} \frac{1 + 4e^{-3/2}}{\epsilon} \frac{\delta_{\gamma}}{\epsilon} + k^{1/2} \frac{1}{\sqrt{1 - \gamma^{2/3}}} + 47.2k^{1/2} \frac{\gamma_{\epsilon}}{\epsilon}. \] (3.24)
Using a similar argument leading to (3.19), \( R_{2,1,2}'' \) can be written as
\[ R_{2,1,2}'' = \sum_{i: \gamma > S^{(y)}} \sum_{j,j' = 1}^k \mathbb{E} U X_{ij} X_{ij'} X_{ij''} \int_{\mathbb{R}^2} \frac{1}{2} \sqrt{1 - s} \]
\[ \times \int_{\mathbb{R}^k} h_{\epsilon}(Z + \sqrt{1 - s} U U' X_i) \partial_{j'j''} \phi_{\Sigma_i}(z) dz \] (3.25)
where \( \Sigma_i = I_{k \times k} - (1 - s) \text{Cov}(X_i, X_i) \) and \( \phi_{\Sigma_i} \) is the density function of \( N(0, \Sigma_i) \).
From
\[ \int_{\mathbb{R}^k} \left| \sum_{j,j' = 1}^k X_{ij} X_{ij'} X_{ij''} \partial_{j'j''} \phi_{\Sigma_i}(z) \right| dz \]
\[ = \int_{\mathbb{R}^k} \left| \sum_{j,j' = 1}^k (N_i^2 X_i)_{ij} (N_i^2 X_i)_{ij'} (N_i^2 X_i)_{ij''} \partial_{j'j''} \phi(z) \right| dz \]
where \( N_i^2 \) is the positive square root of the inverse of \( \Sigma_i \),
\[ |R_{2,1,2}'| \leq \sum_{i: \gamma > S^{(y)}} \mathbb{E} |X_i|^3 \frac{1 + 4e^{-3/2}}{3\sqrt{2\pi}} \left( \frac{1}{1 - \gamma^{2/3}} \right)^{3/2}. \] (3.26)
We used the fact that \( ||N_i^2|| \leq \left( \frac{1}{1 - \gamma^{2/3}} \right)^{1/2}, \) which can be proved as in (3.20), in the above inequality. Therefore,
\[ |R_{2,1,2}| \leq \frac{1 + 4e^{-3/2}}{3\sqrt{2\pi}} \gamma \left( \frac{1}{1 - \gamma^{2/3}} \right)^{3/2}. \] (3.27)
Observing that $R_1$ can be written as

$$R_1 = \sum_{i=1}^{n} \sum_{j,j'=1}^{k} \mathbb{E} \tilde{X}_{ij} \tilde{X}_{ij}' [\partial_{jj'} f_{\epsilon}(W) - \partial_{jj'} f_{\epsilon}(W^{(i)})]$$

where $\tilde{X}_i$ is an independent copy of $X_i$, we can bound it similarly as for $R_2$ as follows.

$$|R_{1,2}| \leq 2 \sqrt{2} \frac{\gamma}{\pi} (16.4k^{1/2} + 156k^{1/2} \frac{\gamma}{\epsilon}),$$  
(3.28)

$$|R_{1,1,1}| \leq 2(1 + \sqrt{2} \frac{\gamma}{\pi}) \gamma \left( \frac{\delta_\gamma}{\epsilon} + k^{1/2} \frac{1}{\sqrt{1 - \gamma^2/3}} + 47.2k^{1/2} \frac{\gamma}{\epsilon} \right),$$  
(3.29)

$$|R_{1,1,2}| \leq \frac{2}{3} (1 + \sqrt{2} \frac{\gamma}{\pi}) \gamma \left( \frac{1}{1 - \gamma^2/3} \right)^{3/2}.$$  
(3.30)

Note that the constants are different from those of $R_2$ because we use (3.11) instead of (3.10) and an extra 2 comes from the fact that there is no $U$ in $R_1$.

From the bounds (3.29), (3.30), (3.28), (3.24), (3.27), (3.17) and the smoothing inequality (3.8), with $c_0 = 2(1 + \sqrt{2} \frac{\gamma}{\pi}) + \frac{1 + 4e^{-3/2}}{\sqrt{2\pi}}$

$$(1 - \frac{\gamma c_0}{\epsilon})^2 \leq (49.2 \sqrt{2} \frac{\gamma}{\pi} + \frac{c_0}{\sqrt{1 - \gamma^2/3}} + \frac{c_0}{3(1 - \gamma^2/3)^{3/2}})k^{1/2}$$

$$+ (468 \sqrt{2} \frac{\gamma}{\pi} + 47.2c_0)k^{1/2} \frac{\gamma^2}{\epsilon} + k^{1/2}(4\gamma + \epsilon).$$

Let $\epsilon = 33\gamma$, and without loss of generality let $\gamma \leq 1/115$. The bound (3.13) is proved by solving the above inequality.

4. Proofs of lemmas

We prove Lemma 2.1 to 2.4 in this section.

**Proof of Lemma 2.1.** The lemma is true by observing that for $x \in \mathbb{R}^k \backslash A'$, $x_0$ must be the nearest point of $x_1$ in $\bar{A}$ where $x_0, x_1$ as defined above Lemma 2.1.

**Proof of Lemma 2.2.** Because $x_0$, the nearest point in $\bar{A}$ from $x$, depends on $x$, the validity of (2.2) is not obvious. We consider the following three cases. All the other cases can be reduced to these cases.

Case 1: $\eta \in \bar{A}$, $\eta + \xi \in \bar{A}$.

Case 2: $\eta \in \bar{A} \backslash \bar{A}$, $\eta + \xi \in \bar{A} \backslash \bar{A}$. 
Case 3: $\eta \in \mathbb{R}^k \setminus A'$, $\eta + \xi \in \mathbb{R}^k \setminus A'$.

In case 1, since $f(\eta) = f(\eta + \xi) = 0$, (2.2) is satisfied.

From the facts that (2.2) is equivalent to

$$(-\xi) \cdot (f(\eta + \xi + (-\xi)) - f(\eta + \xi)) \geq 0$$

and

$$\xi \cdot (\eta - \eta_0) > 0 \implies (-\xi) \cdot ((\eta + \xi) - (\eta + \xi)_0) < 0,$$

which can be proved using a similar argument as in the next paragraph, we only need to consider the following situation in case 2.

Assume $\xi \cdot (\eta - \eta_0) \leq 0$. Let $p_1$ be the plane containing points $\eta_0, \eta, \eta + \xi$. Let the point $(\eta + \xi)'$ be on $p_1$ such that $(\eta + \xi)' - (\eta + \xi)$ is parallel to $\eta_0 - \eta$ and $\eta + \xi' - \eta_0$ is parallel to $\xi$. Let $p_2$ be the $(k-1)$-dimensional hyperplane orthogonal to $\xi$ and containing $(\eta + \xi)'$. The hyperplane $p_2$ divides $\mathbb{R}^k$ into two parts $s_1, s_2$ where $s_1$ is closed and contains $\eta$. If $(\eta + \xi)_0$, the nearest point in $A$ from $\eta + \xi$, is in $s_1$, (2.2) is satisfied. If not, let $(\eta + \xi)''$ be the projection of $(\eta + \xi)_0$ on $p_1$. Then the angle between $\eta_0 - (\eta + \xi)''$ and $\eta + \xi - (\eta + \xi)''$ is less than $\pi/2$. This means that the angle between $\eta_0 - (\eta + \xi)_0$ and $\eta + \xi - (\eta + \xi)_0$ is less than $\pi/2$, which contradicts with the fact that $(\eta + \xi)_0$ is the nearest point in $A$ from $\eta + \xi$.

The validity of (2.2) in case 3 can be proved similarly.

**Proof of Lemma 2.3.** We first prove $f_i$ is $1$-Lipschitz in direction $i$. From (4.2), we only need to prove

$$|f_i(x + he_i) - f_i(x)| \leq h, \quad h > 0$$

(4.3)

in the following two cases.

Case 1: $x, x + he_i \in A' \setminus \bar{A}$ and $e_i \cdot (x - x_0) \leq 0$.

Case 2: $x, x + he_i \notin A'$ and $e_i \cdot (x - x_0) \leq 0$.

For case 1, let $p_1$ be the plane parallel to $x-x_0$, $e_i$ and containing $x$. Let $(x + he_i)'$ be on $p_1$ such that $(x + he_i)' - (x + he_i)$ is parallel to $x - x_0$ and $(x + he_i)' - x_0$ is parallel to $e_i$. Let $p_2$ be the $(k-1)$-dimensional hyperplane orthogonal to $e_i$ and containing $(x + he_i)'$, and let $p_3$ be the $(k-1)$-dimensional hyperplane orthogonal to $x - x_0$ and containing $x_0$. Let $(x + he_i)''$ be the projection of $x + he_i$ on $p_3$ and, let $x'$ be the intersection of the line $\{x_0 + t(x - x_0) : t \in \mathbb{R}\}$
with $p_2$. Then, $(x + he_i)_0$, the projection of $(x + he_i)_0$ on $p_1$, must be within the trapezoid $\{x_0, x', (x + he_i)', (x + he_i)''\}$ (including the boundary), which implies $h \geq f_i(x + he_i) - f_i(x) \geq 0$. Therefore, \((4.3)\) is satisfied. Case 2 is similar.

Since $f_i$ is 1-Lipschitz in direction $i$, $\partial_i f_i$ exist a.e.. From Lemma 2.2,

$$\frac{f_i(x + he_i) - f_i(x)}{h} = \frac{(he_i) \cdot (f(x + he_i) - f(x))}{h^2} \geq 0, \forall h \in \mathbb{R}, h \neq 0.$$ 

Therefore,

$$\partial_i f_i(x) = \lim_{h \to 0} \frac{f_i(x + he_i) - f_i(x)}{h} \geq 0 \text{ a.e.}$$

**Proof of Lemma 2.4.** If $\theta_i = 0$, $f_i(x) = x - x_0 = x_i - x_0$. Note that $x_0$ does not change by moving $x$ a little in the direction of $e_i$. So $\partial_i f_i(x) = 1 = \cos^2 \theta_i$.

If $\theta_i = \pi/2$, Lemma 2.4 follows from Lemma 2.3.

If $0 < \theta_i < \pi/2$ and $h > 0$ small enough such that $x + he_i \in (A')^o \ \bar{\cup} \ \bar{A}$. Let $p_1$ be the $(k - 1)$-dimensional hyperplane orthogonal to $x - x_0$ which contains $x_0$. Let $(x + he_i)'$ be the projection of $x + he_i$ on $p_1$. Let $p_2$ be the $(k - 1)$-dimensional hyperplane orthogonal to $x_0 - (x + he_i)'$ which contains $(x + he_i)'$.

The hyperplane $p_1$ divides $\mathbb{R}^k$ into two parts $s_1, s_2$ where $s_2$ is open and contains $x$; the hyperplane $p_2$ divides $\mathbb{R}^k$ into two parts $s_3, s_4$ where $s_3$ is closed and contains $x$. By observing

$$(x + he_i - (x + he_i)') \cdot e_i = f_i(x) + \cos^2 \theta_i h$$

and $(x + he_i)_0$ must be in $s_1 \cap s_3$, we have,

$$f_i(x + he_i) \geq (x + he_i - (x + he_i)') \cdot e_i = f_i(x) + \cos^2 \theta_i h.$$ 

This implies

$$\frac{f_i(x + he_i) - f_i(x)}{h} \geq \cos^2 \theta_i.$$ 

Therefore,

$$\lim_{h \to 0^+} \frac{f_i(x + he_i) - f_i(x)}{h} \geq \cos^2 \theta_i \text{ a.e.}$$

So $\partial_i f_i(x) \geq \cos^2 \theta_i$ a.e.. For the other possible choices of $\theta_i$, the arguments are similar. This completes the proof of Lemma 2.4.

**Acknowledgement**

This work is based on part of the Ph.D. thesis of the second author. The second author is thankful to the first author for his guidance and helpful discussions.
References

BALL, K. (1993) The reverse isoperimetric problem of Gaussian measure, Discrete Comput. Geom. 10 411-420.

BARBOUR, A.D. (1990) Stein’s method for diffusion approximations. Probab. Theory Related Fields 84 297-322.

BENTKUS, V. (2003) On the dependence of the Berry-Esseen bound on dimension. J. Statist. Plann. Inference 113 385-402.

BHATTACHARIYA, R.N. AND HOLMES, S. (2010) An exposition of Götze’s Estimation of the Rate of Convergence in the Multivariate Central Limit Theorem. Technical Report, Stanford University.

BHATTACHARYA, R.N. AND RAO, R.R. (1986) Normal approximation and asymptotic expansions. Wiley.

BOLTHAUSEN, E. (1984) An estimate of the remainder in a combinatorial central limit theorem. Z. Wahrscheinlichkeitstheorie verw. Gebiete 66 379-386.

CHATTERJEE, S. AND MECKES, E. (2008) Multivariate normal approximation using exchangeable pairs. ALEA Lat. Am. J. Probab. Math. Stat. 4 257-283.

CHEN, L.H.Y. (1986) The rate of convergence in a central limit theorem for dependent random variables with arbitrary index set. IMA Preprint Series 243 Univ. Minnesota.

CHEN, L.H.Y. (1998) Stein’s method: some perspectives with applications. Probability Towards 2000. L. Accardi and C.C. Heyde, eds., Lecture Notes in Statistics 128 Springer Verlag, 515-528.

CHEN, L.H.Y., GOLDSTEIN, L. AND SHAO, Q.M. (2010) Normal approximation by Stein’s method. Springer.

CHEN, L.H.Y. AND SHAO, Q.M. (2001) A non-uniform Berry-Esseen bound via Stein’s method. Probab. Theory Related Fields 120, 236-254.

CHEN, L.H.Y. AND SHAO, Q.M. (2004) Normal approximation under local dependence. Ann. Probab. 32 1985-2028.

GOLDSTEIN, L. AND RINOTT, Y. (1996) Multivariate normal approximation by Stein’s method and size bias couplings. Appl. Probab. Index 33, 1-17.

GÖTZE, F. (1991) On the rate of convergence in the multivariate CLT. Ann. Probab. 19 724-739.

NAGAEV, S.V. (1976) An estimate of the remainder term in the multidimen-
sional central limit theorem. Proc. Third Japan-USSR Symp. Probab. Theory. Lecture Notes in Math. 550 419-438. Springer, Berlin.

RAIČ, M. (2003) Normal approximation with Stein’ method. Proceedings of the Seventh Young Statisticians Meeting.

REINERT, G. AND RÖLLIN, A. (2009) Multivariate normal approximation with Stein’s method of exchangeable pairs under a general linearity condition. Ann. Probab. 37, 2150-2173.

RINOTT, Y. AND ROTAR, V. (1996) A multivariate CLT for local dependence with $n^{-1/2} \log n$ rate and applications to multivariate graph related statistics. J. Multivariate Anal. 56 333-350.

SAZONOV, V.V. (1981) Normal approximation - some recent advances. Springer.

SENATOV, V.V. (1980) Uniform estimates of the rate of convergence in the multidimensional central limit theorem. Teor. Veroyatn. Primen. 25 757-770.

STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. Proc. Sixth Berkeley Symp. Math. Stat. Prob. 2 Univ. California Press. Berkeley, Calif., 583-602.