STA250 Project

Zhi Zhang

February 2021

1 Introduction

Let $\{\xi_i\}_{i=1}^n$ be a sequence of independent mean-zero random vectors in $\mathbb{R}^d$. Let $W = \sum_{i=1}^n \xi_i$ and $\Sigma = \mathrm{Var}(W)$. It is well known that under finite third-moment conditions and for fixed dimension d, the distribution of W can be approximated by a Gaussian distribution with error rate $O(1/\sqrt{n})$. In this paper, we consider the approximation of probabilities of convex sets and Euclidean balls. For convex sets, Bentkus (2005) proved for the above W that if Σ is invertible and $Z \sim N(0, \Sigma)$, then

$$\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leqslant C d^{1/4} \sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^3 \tag{1}$$

where $\mathcal{A}$ is the collection of all measurable convex sets in $\mathbb{R}^d, C$ is an absolute constant and $|\cdot|$ denotes the Euclidean norm when applied to a vector.

The first main result is that up to a logarithmic factor,

$$\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leqslant \log C d^{1/4} \left(\sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4 \right)^{1/2} \tag{2}$$

The bound (2) is optimal up to the $d^{1/4}$ and the logarithmic factors.

We then consider the Gaussian approximation on the class $\mathcal{B}$ of all Euclidean balls, which is arguably most relevant for statistical applications, e.g., chi-square tests. We show that the factor $d^{1/4}$ in (2) can be removed if we replace $\mathcal{A}$ with $\mathcal{B}$. Furthermore, we obtain an error bound that typically vanishes as long as $d = o(n)$. Incidentally, the requirement $d = o(n)$ is necessary for the validity of the Gaussian approximation on balls.

For a matrix M, we use $\|M\|_{H.S.}$ to denote its Hilbert-Schmidt norm. We use C to denote positive absolute constants which may differ in different expressions. For a vector $x \in \mathbb{R}^d$, we use

$x_j, 1 \leqslant j \leqslant d$ to denote its components. For a sequence of vectors $x_i \in \mathbb{R}^d, 1 \leqslant i \leqslant n$, we use x_{ij} to denote the j th component of x_i for $1 \leqslant j \leqslant d$. Similarly, we write X_j and X_{ij} for the components of random vectors X and X_i, respectively.

1.1 Notations and Preliminary

Notation 1.1. For two vectors $x, y \in \mathbb{R}^d$, $x \cdot y$ denotes their inner product. For two $d \times d$ matrices M and N, we write $\langle M, N\rangle_{H.S.}$ for their Hilbert-Schmidt inner product.

$$\langle M, N\rangle_{\text{H.S.}} = \operatorname{tr}\left(M^T N\right) = \sum_i \langle M e_i, N e_i\rangle$$

where $\{e_i : i \in I\}$ an orthonormal basis.

Notation 1.2. For real-valued functions on $\mathbb{R}^d$ we will write $\partial_i f(x)$ for $\partial f(x)/\partial x_i, \partial_{ij} f(x)$ for $\partial^2 f(x)/\left(\partial x_i \partial x_j\right)$ and so forth.

Notation 1.3. We write ∇f and Hess$f$ for the gradient and Hessian matrix of f, respectively. In addition, we denote by $\nabla^r f(x)$ the r-th derivative of f at x regarded as an r-linear form: The value of $\nabla^r f(x)$ evaluated at $u_1, \ldots, u_r \in \mathbb{R}^d$ is given by

$$\langle \nabla^r f(x), u_1 \otimes \cdots \otimes u_r\rangle = \sum_{j_1, \ldots, j_r = 1}^{d} \partial_{j_1, \ldots, j_r} f(x) u_{1, j_1} \cdots u_{r, j_r}$$

When $u_1 = \cdots = u_r =: u$, we write $u_1 \otimes \cdots \otimes u_r = u^{\otimes r}$ for short.

Notation 1.4. For any r -linear form T, its injective norm is defined by

$$|T|_\vee := \sup_{|u_1| \vee \ldots \vee |u_r| \leqslant 1} |\langle T, u_1 \otimes \cdots \otimes u_r\rangle|$$

For an $(r-1)$ -times differentiable function $h : \mathbb{R}^d \to \mathbb{R}$, we write

$$M_r(h) := \sup_{x \neq y} \frac{\left|\nabla^{r-1} h(x) - \nabla^{r-1} h(y)\right|_\vee}{|x - y|}$$

Remark 1.1. Note that $M_r(h) = \sup_{x \in \mathbb{R}^d} \left|\nabla^r h(x)\right|_\vee$ if h is r-times differentiable.

Finally, we refer to the following bound for derivatives of the d-dimensional standard normal density ϕ, which will be used several times in the following (cf. the inequality after Eq. (4.9) of Raič (2019 b)):

$$\int_{\mathbb{R}^d} \left|\left\langle \nabla^s \phi(z), u^{\otimes s}\right\rangle\right| dz \leqslant C_s |u|^s \quad \text{for any fixed integer } s \tag{3}$$

where C_s is a constant depending only on s.

2 Main Theorem

2.1 Approximation on Convex Sets

Theorem 2.1. *Let $\xi = \{\xi_i\}_{i=1}^n$ be a sequence of centered independent random vectors in $\mathbb{R}^d$ with finite fourth moments and set $W = \sum_{i=1}^n \xi_i$. Assume* $\operatorname{Var}(W) = \Sigma$ *and Σ is invertible. Let $Z \sim N(0, \Sigma)$ be a centered Gaussian vector in $\mathbb{R}^d$ with covariance matrix Σ. Then,*

$$\sup_{A\in\mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leqslant C d^{1/4} \left(\sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4 \right)^{1/2} \left(\left| \log \left(\sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4 \right) \right| \vee 1 \right) \tag{4}$$

where $\mathcal{A}$ is the collection of all measurable convex sets in $\mathbb{R}^d$.

Remark 2.1. The bound (4) is optimal up to the $d^{1/4}$ and the logarithmic factors.

Proposition 2.1. *There is an absolute constant $C_0 > 0$ such that, for sufficiently large n, we can construct centered i.i.d. random vectors $\xi_1, \ldots, \xi_n$ in $\mathbb{R}^d$ with finite fourth moments (which may depend on n) satisfying* $\operatorname{Var}(W) = I_d$ *and*

$$\sup_{A\in\mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \geqslant C_0 \left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right)^{1/2}$$

as long as $d \leqslant \sqrt{n}/\log n$.

Remark 2.2. Consider the situation where $\xi_i = X_i/\sqrt{n}$ and $\{X_1, X_2, \ldots\}$ is a sequence of i.i.d. mean-zero random vectors in $\mathbb{R}^d$ with $\operatorname{Var}(X_i) = I_d$. In this setting, $\Sigma = I_d$, and for the d-vector X_i, we have

$$\mathbb{E} |X_i|^3 \propto d^{3/2} \qquad \mathbb{E} |X_i|^4 \propto d^2$$

In this case, the right-hand side of (4) is of the order $O\left(\frac{d^{5/2}}{n}\right)^{1/2}$ up to a logarithmic factor. In contrast, the right-hand side of (1) is of the order $O\left(\frac{d^{7/2}}{n}\right)^{1/2}$. Therefore, subject to the requirement of the existence of the fourth moment, (4) is preferable to (1) in the large-dimensional setting where $d \to \infty$.

2.2 Approximation on Euclidean Balls

In this section, we show that the factor $d^{1/4}$ appearing on the right-hand side of (4) can be removed if we restrict the approximation to the class of balls. To facilitate the application to the bootstrap, here we do not assume W and Z have the same covariance matrix.

Theorem 2.2. *Let* $\xi = \{\xi_i\}_{i=1}^n$ *be a sequence of centered independent random vectors in* $\mathbb{R}^d$ *with finite fourth moments and set* $W = \sum_{i=1}^n \xi_i$. *Let* $Z \sim N(0, \Sigma)$ *be a centered Gaussian vector in* $\mathbb{R}^d$ *with covariance matrix* Σ. *Assume* Σ *is invertible. Then*

$$\sup_{A\in\mathcal{B}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leqslant C\Psi(\delta(W, \Sigma)) \tag{5}$$

where $\Psi(x) = x(|\log x| \vee 1), \mathcal{B}$ *is the set of all Euclidean balls in* $\mathbb{R}^d$ *and*

$$\delta(W, \Sigma) := \left\| I_d - \operatorname{Var}\left(\Sigma^{-1/2}W\right)\right\|_{H.S.} + \left(\sum_{i=1}^n \mathbb{E}\left|\Sigma^{-1/2}\xi_i\right|^4\right)^{1/2}$$

Remark 2.3. We can see that if $\operatorname{Var}(W) = \Sigma$, then the typical order of the right-hand side of (5) is $O\left(\frac{d^2}{n}\right)^{1/2}$ up to a logarithmic factor. It has near-optimal dependence on n and converges to 0 if $d = o(\sqrt{n})$.

In the next result, we sacrifice the rate of n to obtain the optimal growth rate of $d = o(n)$ in terms of the dimension.

Theorem 2.3. *Let* ξ, W *and* Z *be as in Theorem 2.2. Assume* $\operatorname{tr}\left(\Sigma^2\right) > 0$. *Then*

$$\sup_{A\in\mathcal{B}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leqslant \frac{C}{\operatorname{tr}\left(\Sigma^2\right)^{1/4}}\sqrt{\tilde{\delta}(W, \Sigma)} \tag{6}$$

where

$$\tilde{\delta}(W, \Sigma) := \|\Sigma - \operatorname{Var}(W)\|_{H.S.} + \sum_{j=1}^d |\Sigma_{jj} - \operatorname{Var}(W_j)| + \sqrt{\sum_{i=1}^n \mathbb{E}\,|\xi_i|^4} + \sum_{j=1}^d \sqrt{\sum_{i=1}^n \mathbb{E}\left[\xi_{ij}^4\right]}$$

Remark 2.4. Since $\mathbb{E}\,|\xi_i|^4 \leqslant d\sum_{j=1}^d \mathbb{E}\xi_{ij}^4$, if $\operatorname{Var}(W) = \Sigma = I_d$, the right-hand side of (6) is bounded by

$$C \max_{1\leqslant j\leqslant d}\left(d\sum_{i=1}^n \mathbb{E}\xi_{ij}^4\right)^{1/4}$$

If $\max_{1\leqslant i\leqslant n}\max_{1\leqslant j\leqslant d}\left(\mathbb{E}\xi_{ij}^4\right)^{1/4} = O(1/\sqrt{n})$ as $n \to \infty$ as in the typical case in applications (where $\xi_{ij} = \frac{X_{ij}}{\sqrt{n}}$ for some X_{ij} not depending on n), this converges to 0 as long as $d/n \to 0$.

Remark 2.5. The inequality

$$\mathbb{E}\,|\xi_i|^4 \leqslant d\sum_{j=1}^d \mathbb{E}\xi_{ij}^4$$

can be obtained by applying Holder inequality

$$\sum_{k=1}^{d}|x_ky_k| \leq \left(\sum_{k=1}^{d}|x_k|^p\right)^{\frac{1}{p}}\left(\sum_{k=1}^{d}|y_k|^q\right)^{\frac{1}{q}}$$

with $p=q=2$ and $x_k=\xi_{ik}^2$ and $y_k=1$. In particular, we have

$$\begin{aligned}\mathbb{E}\left|\xi_i\right|^4 &= \mathbb{E}\left(\xi_{i1}^2+\cdots+\xi_{id}^2\right)^2 \\ &\leq \mathbb{E}\left[\left(\xi_{i1}^4+\cdots+\xi_{id}^4\right)(1+\cdots+1)\right] \\ &= d\sum_{j=1}^{d}\mathbb{E}\xi_{ij}^4\end{aligned}$$

Remark 2.6. Theorem 2.3 can be used to deduce Central Limit Theorems (CLTs) for $|W-a|^2$ under suitable conditions. For example, if $a=0, \Sigma=I_d, \xi_i=X_i/\sqrt{n}$ for an i.i.d. sequence of random vectors $\{X_1,\ldots,X_n\}$ with $\max_{1\leqslant j\leqslant d}\mathbb{E}\left(X_{ij}^4\right)\leqslant C$, then by Theorem 2.3, Remark 2.4 and the CLT for chi-square random variables, we have, for $d\to\infty$ and $d=o(n)$

$$\frac{|W|^2-d}{\sqrt{2d}}\to N(0,1) \quad \text{in distribution.}$$

Proposition 2.2. *Let $X_1,\ldots,X_n$ be i.i.d. standard Gaussian vectors in $\mathbb{R}^d$. Let $\{e_i\}_{i=1}^n$ be i.i.d. variables independent of $\{X_i\}_{i=1}^n$ with $\mathbb{E}e_1=0, \mathbb{E}e_1^2=1, \mathbb{E}e_1^4<\infty$ and $\operatorname{Var}\left(e_1^2\right)>0$. Assume the law of e_1 does not depend on n. Set $W:=n^{-1/2}\sum_{i=1}^n e_iX_i$ and let $Z\sim N\left(0,I_d\right)$. If*

$$\sup_{x\geqslant 0}|\mathbb{P}(|W|\leqslant x)-\mathbb{P}(|Z|\leqslant x)|\to 0$$

as $d,n\to\infty$, we must have $d/n\to 0$.

Proof. Since $\left(|Z|^2-d\right)/\sqrt{2d}$ converges in law to $N(0,1)$ as $d\to\infty$, $\left(|W|^2-d\right)/\sqrt{2d}$ also converges in law to $N(0,1)$. Since W has the same law as $\sqrt{V}Z'$ by assumption, where $V:=n^{-1}\sum_{i=1}^n e_i^2$ and $Z'\sim N\left(0,I_d\right)$ is independent of $\{e_i\}_{i=1}^{\infty}$, $\left(V\left|Z'\right|^2-d\right)/\sqrt{2d}$ should also converges in law to $N(0,1)$. Since

$$\frac{V\left|Z'\right|^2-d}{\sqrt{2d}}=V\frac{\left|Z'\right|^2-d}{\sqrt{2d}}+\sqrt{\frac{d}{2}}(V-1)=(V-1)\frac{\left|Z'\right|^2-d}{\sqrt{2d}}+\frac{\left|Z'\right|^2-d}{\sqrt{2d}}+\sqrt{\frac{d}{2}}(V-1)$$

and the first term converges to 0 in probability,

$$\frac{\left|Z'\right|^2-d}{\sqrt{2d}}+\sqrt{\frac{d}{2}}(V-1)$$

must converge in law to $N(0,1)$. In the above expression, the first term converges in law to $N(0,1)$ and the first and second terms are independent, so this implies $\sqrt{d}(V-1)=o_p(1)$ as $n\to\infty$. Since $\sqrt{n}(V-1)$ converges in law to $N\left(0,\operatorname{Var}\left(e_1^2\right)\right)$, we must have $d/n\to 0$. □

Remark 2.7. W in Proposition 2.2 can be regarded as a bootstrap approximation of Z. Remark 2.4 and Proposition 2.2 suggest that, in general, bootstrapping may not provide a more accurate approximation than the Gaussian approximation in terms of the dependence on dimension.

3 Application to Bootstrap Approximation on Balls

Notation 3.1. $X_1, \ldots, X_n$: be a sequence of centered independent vectors in $\mathbb{R}^d$ with finite fourth moments. $W := n^{-1/2} \sum_{i=1}^n X_i$ $\Sigma := Var(W)$ $Z \sim N(0, \Sigma)$ $X_1^*, \ldots, X_n^*$: be i.i.d. draws from the empirical distribution of X

3.1 Empirical bootstrap approximation for $\mathbb{P}(W \in A)$

$W^* := \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i^* - \bar{X})$., where $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$ The bootstrap analog of 2.3 is given by:

Theorem 3.1. $\operatorname{tr}(\Sigma^2) > 0$, *for any* $K > 0$, *we have*

$$\mathbb{P}\left(\sup_{A \in \mathcal{B}} |\mathbb{P}(W^* \in A|X) - \mathbb{P}(Z \in A)| > K\sqrt{\Delta_n}\right) \leqslant \frac{C}{K^2} \tag{7}$$

where

$$\Delta_n := \frac{1}{n \operatorname{tr}(\Sigma^2)^{1/2}} \left(\sqrt{\sum_{i=1}^n \mathbb{E}\,|X_i|^4} + \sum_{j=1}^d \sqrt{\sum_{i=1}^n \mathbb{E}\left[X_{ij}^4\right]} \right)$$

Remark 3.1. Compared to the non-asymptotic bound for the quantity of $\sup_{A \in \mathcal{B}} |\mathbb{P}(W^* \in A|X) - \mathbb{P}(Z \in A)|$ under additional distribution assumption on X_i. Ours Theorem 3.1 provides better dependence on the **dimension** $d (d = o(n)$ **v.s.** $d = o(n^{1/2}))$**, at least when** $\Sigma = I_d$; our result allows Σ **to be singular**; it's possible to give a non-asymptotic version of 7 but an exponential concentration if we also assume X_i are sub-Gaussian.

3.2 Wild bootstrap approximation for $\mathbb{P}(W \in A)$

Let $\{e_i\}_{i=1}^n$ be i.i.d. variables independent of $\{X_i\}_{i=1}^n$ with $\mathbb{E}e_1 = 0, \mathbb{E}e_1^2 = 1, \mathbb{E}e_1^4 < \infty$. The $W^o := \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i X_i$. is the wild bootstrap approximation of W with multiplier variables $e_1, \ldots, e_n$.

Theorem 3.2. $\operatorname{tr}(\Sigma^2) > 0$, *for any* $K > 0$, *we have*

$$\mathbb{P}\left(\sup_{A \in \mathcal{B}} |\mathbb{P}(W^o \in A|X) - \mathbb{P}(Z \in A)| > K(\mathbb{E}e_1^4)^{1/4}\sqrt{\Delta_n}\right) \leqslant \frac{C}{K^2} \tag{8}$$

where Δ_n *is defined in 3.1*

Remark 3.2. Compared to the non-asymptotic bound for the quantity of $\sup_{A\in\mathcal{B}}|\mathbb{P}(W^o \in A|X) - \mathbb{P}(Z \in A)|$ under additional distribution assumption on X_i. Ours 3.2 provides better dependence on the n **and** $d(O(d/n)^{1/4}$ **v.s.** $O(d^2/n)^{1/5}))$; ours does not require the **unit skewness assumption** $\mathbb{E}e_1^3 = 1$ on the multiplier variables; it's possible to give a non-asymptotic version of our result of 3.2.

4 Main Idea behind the Proof

4.1 Basic Decomposition

The proofs for Theorems 2.1 and 2.2-2.3 start with approximating the indicator function 1_A for $A \in \mathcal{A}$ or $A \in \mathcal{B}$ by an appropriate smooth function h. Then, the problem amounts to establishing an appropriate bound for $\mathbb{E}h(W) - \mathbb{E}h(Z)$.

To accomplish this, we will make use of a decomposition of $\mathbb{E}h(W) - \mathbb{E}h(Z)$ derived from the exchangeable pair approach in Stein's method for multivariate normal approximation by Chatterjee and Meckes (2008) and Reinert and Röllin (2009) along with a symmetry argument by Fang and Koike (2020a,b).

Lemma 4.1. Given a twice differentiable function $h : \mathbb{R}^d \to \mathbb{R}$ with bounded partial derivatives, we consider the Stein equation

$$\langle \operatorname{Hess} f(w), \Sigma\rangle_{H.S.} - w \cdot \nabla f(w) = h(w) - \mathbb{E}h(Z), \quad w \in \mathbb{R}^d \tag{9}$$

then

$$f(w) = \int_0^1 -\frac{1}{2(1-s)} \int_{\mathbb{R}^d} \left[h\left(\sqrt{1-s}w + \sqrt{s}\Sigma^{1/2}z\right) - \mathbb{E}h(Z)\right] \phi(z)dzds \tag{10}$$

is a solution to (9).

In the following we assume that f is thrice differentiable with bounded partial derivatives. This is true if Σ is invertible or h is thrice differentiable with bounded partial derivatives. Now we introduce the basic decomposition we will use later in proof.

Let $\{\xi_1', \ldots, \xi_n'\}$ be an independent copy of $\{\xi_1, \ldots, \xi_n\}$, and let I be a random index uniformly chosen from $\{1, \ldots, n\}$ and independent of $\{\xi_1, \ldots, \xi_n, \xi_1', \ldots, \xi_n'\}$. Define $W' = W - \xi_I + \xi_I'$. It is easy to verify that (W, W') has the same distribution as (W', W) (exchangeability) and

$$\mathbb{E}\left(W' - W \mid W\right) = -\frac{W}{n} \tag{11}$$

From exchangeability and (11) we have, with $D = W' - W$

$$\begin{aligned} 0 &= \frac{n}{2}\mathbb{E}\left[D \cdot \left(\nabla f\left(W'\right) + \nabla f(W)\right)\right] \\ &= \mathbb{E}\left[\frac{n}{2}D \cdot \left(\nabla f\left(W'\right) - \nabla f(W)\right) + nD \cdot \nabla f(W)\right] \\ &= \mathbb{E}\left[\frac{n}{2}\sum_{j,k=1}^{d} D_j D_k \partial_{jk} f(W) + R_2 + nD \cdot \nabla f(W)\right] \\ &= \mathbb{E}\left[\langle \operatorname{Hess} f(W), \Sigma\rangle_{H.S.} - R_1 + R_2 - W \cdot \nabla f(W)\right] \end{aligned} \tag{12}$$

where

$$R_1 = \sum_{j,k=1}^{d} \mathbb{E}\left\{\left(\Sigma_{jk} - \frac{n}{2}D_j D_k\right)\partial_{jk} f(W)\right\} \tag{13}$$

and

$$R_2 = \frac{n}{2}\sum_{j,k,l=1}^{d} \mathbb{E} D_j D_k D_l U \partial_{jkl} f(W + (1-U)D) \tag{14}$$

and U is a uniform random variable on $[0,1]$ independent of everything else. From (9) and (12) we have

$$\mathbb{E}h(W) - \mathbb{E}h(Z) = R_1 - R_2 \tag{15}$$

We further rewrite R_1 and R_2 respectively as follows. First, set

$$V = \left(V_{jk}\right)_{1\leqslant j,k\leqslant d} := \left(\mathbb{E}\left[\Sigma_{jk} - \frac{n}{2}D_j D_k \mid \xi\right]\right)_{1\leqslant j,k\leqslant d}$$

Then we evidently have

$$R_1 = \sum_{j,k=1}^{d} \mathbb{E} V_{jk} \partial_{jk} f(W) = \mathbb{E}\langle V,\ \operatorname{Hess}\ f(W)\rangle_{H.S.} \tag{16}$$

Also, one can easily verify that (cf. Eq.(22) of Chernozhukov, Chetverikov and Kato (2014))

$$V = \Sigma - \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\left[\xi_i\xi_i^{\top}\right] - \frac{1}{2}\sum_{i=1}^{n}\xi_i\xi_i^{\top} = (\Sigma - \operatorname{Var}(W)) - \frac{1}{2}\sum_{i=1}^{n}\left(\xi_i\xi_i^{\top} - \mathbb{E}\left[\xi_i\xi_i^{\top}\right]\right) \tag{17}$$

Next, by exchangeability we have

$$\begin{aligned} &\mathbb{E}\left[D_j D_k D_l U \partial_{jkl} f(W + (1-U)D)\right] \\ &= -\mathbb{E}\left[D_j D_k D_l U \partial_{jkl} f\left(W' - (1-U)D\right)\right] \\ &= -\mathbb{E}\left[D_j D_k D_l U \partial_{jkl} f(W + UD)\right] \end{aligned} \tag{18}$$

Hence we obtain

$$R_2 = \frac{n}{4}\sum_{j,k,l=1}^{d} \mathbb{E}\left[D_j D_k D_l U\left\{\partial_{jkl} f(W + (1-U)D) - \partial_{jkl} f(W + UD)\right\}\right] \tag{19}$$

4.2 Proof of Theorem 2.1

Since $\Sigma^{-1/2}W=\sum_{i=1}^{n}\Sigma^{-1/2}\xi_i$ and $\left\{\Sigma^{-1/2}x: x\in A\right\}\in\mathcal{A}$ for all $A\in\mathcal{A}$, it suffices to consider the case $\Sigma=I_d$. The proof is a combination of Bentkus (2003)'s smoothing, the decomposition (15), and a recursive argument by Raič (2019a). Fix $\beta_0>0$. Define

$$K\left(\beta_0\right)=\sup_{W}\frac{\sup_{A\in\mathcal{A}}|\mathbb{P}(W\in A)-\mathbb{P}(Z\in A)|}{\max\left\{\beta_0,\left(\sum_{i\in\mathcal{I}}\mathbb{E}\left|\xi_i\right|^4\right)^{1/2}\left(\left|\log\left(\sum_{i\in\mathcal{I}}\mathbb{E}\left|\xi_i\right|^4\right)\right|\vee 1\right)\right\}}\tag{20}$$

where the first supremum is taken over the family of all sums $W=\sum_{i\in\mathcal{I}}\xi_i$ of finite number of independent mean-zero random vectors with $\mathbb{E}\left|\xi_i\right|^4<\infty$ and $\operatorname{Var}(W)=I_d$. We will obtain a recursive inequality for $K\left(\beta_0\right)$ and prove that

$$K\left(\beta_0\right)\leqslant Cd^{1/4}\tag{21}$$

for an absolute constant C that does not depend on β_0. Equation (4) then follows by sending $\beta_0\to 0$.

Now we fix a $W=\sum_{i=1}^{n}\xi_i, n\geqslant 1$, in the aforementioned family

$$\bar{\beta}=\max\left\{\beta_0,\left(\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\right)^{1/2}\left(\left|\log\left(\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\right)\right|\vee 1\right)\right\}.\tag{22}$$

and for $A\in\mathcal{A},\varepsilon>0$, define

$$A^{\varepsilon}=\left\{x\in\mathbb{R}^d:\operatorname{dist}(x,A)\leqslant\varepsilon\right\}\qquad\operatorname{dist}(x,A)=\inf_{y\in A}|x-y|$$

Lemma 4.2. For any $A\in\mathcal{A}$ and $\varepsilon>0$, there exists a function $h_{A,\varepsilon}$ (which depends only on A and ε) such that

$$h_{A,\varepsilon}(x)=1\text{ for }x\in A,\quad h_{A,\varepsilon}(x)=0\text{ for }x\in\mathbb{R}^d\setminus A^{\varepsilon},\quad 0\leqslant h_{A,\varepsilon}(x)\leqslant 1$$

and

$$M_1\left(h_{A,\varepsilon}\right)\leqslant\frac{C}{\varepsilon},\quad M_2\left(h_{A,\varepsilon}\right)\leqslant\frac{C}{\varepsilon^2}\tag{23}$$

where C is an absolute constant that does not depend on A and ε.

Lemma 4.3. Let ϕ be the standard Gaussian density on $\mathbb{R}^d, d\geqslant 2$, and let A be a convex set in $\mathbb{R}^d$. Then

$$\int_{\partial A}\phi\leqslant 4d^{1/4}\tag{24}$$

Lemma 4.4. For any d -dimensional random vector W and any $\varepsilon > 0$,

$$\sup_{A\in\mathcal{A}} |P(W\in A) - P(Z\in A)| \leqslant 4d^{1/4}\varepsilon + \sup_{A\in\mathcal{A}} |\mathbb{E}h_{A,\varepsilon}(W) - \mathbb{E}h_{A,\varepsilon}(Z)| \tag{25}$$

where $h_{A,\varepsilon}$ is as in Lemma 4.2.

We now fix $A \in \mathcal{A}$ (will take sup later), $0 < \varepsilon \leqslant 1$, write $h := h_{A,\varepsilon}$ and proceed to bound $|\mathbb{E}h(W) - \mathbb{E}h(Z)|$ by the decomposition (15). Consider the solution f to the Stein equation (9) with $\Sigma = I_d$

$$f(w) = \int_0^1 -\frac{1}{2(1-s)} \int_{\mathbb{R}^d} \left[h\left(\sqrt{1-s}w + \sqrt{s}z\right) - \mathbb{E}h(Z)\right] \phi(z)dzds$$

Since h has bounded partial derivatives up to the second order and $\Sigma = I_d$ is invertible, f is thrice differentiable with bounded partial derivatives. Using the integration by parts formula, we have for $1 \leqslant j,k,l \leqslant d$ and any constant $0 \leqslant c_0 \leqslant 1$ that

$$\begin{aligned}\partial_{jk}f(w) = &\int_0^{c_0} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} \partial_j h(\sqrt{1-s}w + \sqrt{s}z)\partial_k\phi(z)dzds \\ &+ \int_{c_0}^1 -\frac{1}{2s} \int_{\mathbb{R}^d} h(\sqrt{1-s}w + \sqrt{s}z)\partial_{jk}\phi(z)dzds\end{aligned} \tag{26}$$

and

$$\begin{aligned}\partial_{jkl}f(w) = &\int_0^{c_0} \frac{\sqrt{1-s}}{2\sqrt{s}} \int_{\mathbb{R}^d} \partial_{jk} h(\sqrt{1-s}w + \sqrt{s}z)\partial_l\phi(z)dzds \\ &+ \int_{c_0}^1 -\frac{\sqrt{1-s}}{2s} \int_{\mathbb{R}^d} \partial_j h(\sqrt{1-s}w + \sqrt{s}z)\partial_{kl}\phi(z)dzds\end{aligned} \tag{27}$$

We first bound R_1 in (16). We will utilize the following lemma.

Lemma 4.5. For $k \geqslant 1$ and each map $a : \{1,\ldots,d\}^k \to \mathbb{R}$, we have

$$\int_{\mathbb{R}^d} \left(\sum_{i_1,\ldots,i_k=1}^d a(i_1,\ldots,i_k) \frac{\partial_{i_1\ldots i_k}\phi(z)}{\phi(z)}\right)^2 \phi(z)dz \leqslant k! \sum_{i_1,\ldots,i_k=1}^d (a(i_1,\ldots,i_k))^2 \tag{28}$$

Now, using the expression of $\partial_{jk}f$ in (26) with $c_0 = \varepsilon^2$, we have

$$R_1 = R_{11} + R_{12}$$

where

$$R_{11} = \sum_{j,k=1}^d \mathbb{E}\left[V_{jk} \int_0^{\varepsilon^2} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} \partial_j h(\sqrt{1-s}W + \sqrt{s}z)\partial_k\phi(z)dzds\right]$$

and

$$R_{12}=\sum_{j,k=1}^{d}\mathbb{E}\left[V_{jk}\int_{\varepsilon^2}^{1}-\frac{1}{2s}\int_{\mathbb{R}^d}h(\sqrt{1-s}W+\sqrt{s}z)\partial_{jk}\phi(z)dzds\right]$$

For R_{11}, we use the Cauchy-Schwarz inequality and the bounds (23) and (28) and obtain

$$\begin{aligned}|R_{11}|&=\left|\int_0^{\varepsilon^2}\frac{1}{2\sqrt{s}}\int_{\mathbb{R}^d}\mathbb{E}\sum_{j=1}^{d}\partial_j h(\sqrt{1-s}W+\sqrt{s}z)\sum_{k=1}^{d}V_{jk}\frac{\partial_k\phi(z)}{\phi(z)}\phi(z)dzds\right|\\&\leqslant\frac{C}{\varepsilon}\int_0^{\varepsilon^2}\frac{1}{2\sqrt{s}}\int_{\mathbb{R}^d}\mathbb{E}\left\{\sum_{j=1}^{d}\left(\sum_{k=1}^{d}V_{jk}\frac{\partial_k\phi(z)}{\phi(z)}\right)^2\right\}^{1/2}\phi(z)dzds\\&\leqslant\frac{C}{\varepsilon}\int_0^{\varepsilon^2}\frac{1}{2\sqrt{s}}\left\{\int_{\mathbb{R}^d}\mathbb{E}\sum_{j=1}^{d}\left(\sum_{k=1}^{d}V_{jk}\frac{\partial_k\phi(z)}{\phi(z)}\right)^2\phi(z)dz\right\}^{1/2}ds\\&\leqslant\frac{C}{\varepsilon}\int_0^{\varepsilon^2}\frac{1}{2\sqrt{s}}\left\{\mathbb{E}\sum_{j=1}^{d}\sum_{k=1}^{d}V_{jk}^2\right\}^{1/2}ds\leqslant C\left\{\sum_{j,k=1}^{d}\mathbb{E}V_{jk}^2\right\}^{1/2}\end{aligned}\tag{29}$$

Recall that $\mathrm{Var}(W)=\Sigma$

$$V=\Sigma-\frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\left[\xi_i\xi_i^\top\right]-\frac{1}{2}\sum_{i=1}^{n}\xi_i\xi_i^\top=(\Sigma-\mathrm{Var}(W))-\frac{1}{2}\sum_{i=1}^{n}\left(\xi_i\xi_i^\top-\mathbb{E}\left[\xi_i\xi_i^\top\right]\right)$$

we have

$$\mathbb{E}V_{jk}^2=\frac{1}{4}\mathrm{Var}\left[\sum_{i=1}^{n}\xi_{ij}\xi_{ik}\right]=\frac{1}{4}\sum_{i=1}^{n}\mathrm{Var}\left[\xi_{ij}\xi_{ik}\right]\leqslant\frac{1}{4}\sum_{i=1}^{n}\mathbb{E}\left[\xi_{ij}^2\xi_{ik}^2\right]$$

and therefore,

$$|R_{11}|\leqslant C\left\{\sum_{j,k=1}^{d}\sum_{i=1}^{n}\mathbb{E}\left[\xi_{ij}^2\xi_{ik}^2\right]\right\}^{1/2}=C\left\{\sum_{i=1}^{n}\mathbb{E}\left[\sum_{j=1}^{d}\xi_{ij}^2\right]^2\right\}^{1/2}=C\left(\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\right)^{1/2}$$

Applying similar arguments, we have, for R_{12},

$$\begin{aligned}|R_{12}|&=\left|\int_{\varepsilon^2}^{1}\left(-\frac{1}{2s}\right)\int_{\mathbb{R}^d}\mathbb{E}h(\sqrt{1-s}W+\sqrt{s}z)\sum_{j,k=1}^{d}V_{jk}\frac{\partial_{jk}\phi(z)}{\phi(z)}\phi(z)dzds\right|\\&\leqslant\int_{\varepsilon^2}^{1}\frac{1}{2s}\int_{\mathbb{R}^d}\mathbb{E}\left|\sum_{j,k=1}^{d}V_{jk}\frac{\partial_{jk}\phi(z)}{\phi(z)}\right|\phi(z)dzds\\&\leqslant\int_{\varepsilon^2}^{1}\frac{1}{2s}\left\{\int_{\mathbb{R}^d}\mathbb{E}\left[\sum_{j,k=1}^{d}V_{jk}\frac{\partial_{jk}\phi(z)}{\phi(z)}\right]^2\phi(z)dz\right\}^{1/2}ds\\&\leqslant C\int_{\varepsilon^2}^{1}\frac{1}{2s}\left\{\mathbb{E}\sum_{j,k=1}^{d}V_{jk}^2\right\}^{1/2}ds\leqslant C|\log\varepsilon|\left(\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\right)^{1/2}\end{aligned}\tag{30}$$

therefore,

$$|R_1| \leqslant C(|\log \varepsilon| \vee 1) \left(\sum_{i=1}^{n} \mathbb{E} |\xi_i|^4 \right)^{1/2} \tag{31}$$

Next, we bound R_2. Take $0 < \eta \leqslant 1$ arbitrarily. Using the expression of $\partial_{jkl} f$ in (27) with $c_0 = \eta^2$ and the two equivalent expressions (14) and (19) for R_2, we have

$$R_2 = R_{21} + R_{22}$$

where

$$\begin{aligned} R_{21} =& \frac{1}{2} \sum_{i=1}^{n} \sum_{j,k,l=1}^{d} \mathbb{E} U \left(\xi'_{ij} - \xi_{ij}\right) \left(\xi'_{ik} - \xi_{ik}\right) \left(\xi'_{il} - \xi_{il}\right) \int_0^{\eta^2} \frac{\sqrt{1-s}}{2\sqrt{s}} \\ & \times \int_{\mathbb{R}^d} \partial_{jk} h \left(\sqrt{1-s} \left(W + (1-U)\left(\xi'_i - \xi_i\right)\right) + \sqrt{s} z \right) \partial_l \phi(z) dz ds \end{aligned}$$

and

$$\begin{aligned} R_{22} =& \frac{1}{4} \sum_{i=1}^{n} \sum_{j,k,l=1}^{d} \mathbb{E} U \left(\xi'_{ij} - \xi_{ij}\right) \left(\xi'_{ik} - \xi_{ik}\right) \left(\xi'_{il} - \xi_{il}\right) \int_{\eta^2}^{1} - \frac{\sqrt{1-s}}{2s} \\ & \times \int_{\mathbb{R}^d} \left[\partial_j h \left(\sqrt{1-s} \left(W + (1-U)\left(\xi'_i - \xi_i\right)\right) + \sqrt{s} z \right) - \partial_j h \left(\sqrt{1-s} \left(W + U \left(\xi'_i - \xi_i\right)\right) + \sqrt{s} z \right) \right] \partial_{kl} \phi(z) dz ds \\ =& \frac{1}{4} \sum_{i=1}^{n} \sum_{j,k,l,m=1}^{d} \mathbb{E} U(1-2U) \left(\xi'_{ij} - \xi_{ij}\right) \left(\xi'_{ik} - \xi_{ik}\right) \left(\xi'_{il} - \xi_{il}\right) \left(\xi'_{im} - \xi_{im}\right) \int_{\eta^2}^{1} - \frac{1-s}{2s} \\ & \times \int_{\mathbb{R}^d} \partial_{jm} h \left(\sqrt{1-s} \left(W + (U + (1-2U)U')\left(\xi'_i - \xi_i\right)\right) + \sqrt{s} z \right) \partial_{kl} \phi(z) dz ds \end{aligned} \tag{32}$$

where U' is a uniform random variable on $[0,1]$ independent of everything else and we used the mean value theorem in the last equality. Let $W^{(i)} = W - \xi_i$ for $i \in \{1, \ldots, n\}$. We will use the fact that ∇h is non-zero only in $A^{\varepsilon} \backslash A$ and bound

$$\mathbb{P}\left(\sqrt{1-s} W^{(i)} \in A_i^{\varepsilon} \backslash A_i \mid U, U', \xi_i, \xi'_i \right)$$

where $0 < s < 1$ and A_i is a convex set which may depend on U, U', ξ_i, ξ'_i, s and z. Let Σ_i be the covariance matrix of $W^{(i)}$ and let σ_i be its smallest eigenvalue, which will be assumed to be positive in Case 1 below. We have

$$\begin{aligned} & \mathbb{P}\left(\sqrt{1-s} W^{(i)} \in A_i^{\varepsilon} \setminus A_i \mid U, U', \xi_i, \xi'_i \right) \\ =& \mathbb{P}\left(\Sigma_i^{-1/2} W^{(i)} \in \frac{1}{\sqrt{1-s}} \Sigma_i^{-1/2} \left(A_i^{\varepsilon} \setminus A_i\right) \mid U, U', \xi_i, \xi'_i \right) \\ \leqslant & 4 d^{1/4} \frac{\varepsilon}{\sigma_i \sqrt{1-s}} + 2 \sup_{A \in \mathcal{A}} \left| \mathbb{P}\left(\Sigma_i^{-1/2} W^{(i)} \in A \right) - P(Z \in A) \right| \end{aligned} \tag{33}$$

where we used the $4d^{1/4}$ upper bound for the Gaussian surface area of any convex set in Lemma 4.3. From the definition in (20), we have

$$\begin{aligned}
&\sup_{A\in\mathcal{A}}\left|\mathbb{P}\left(\Sigma^{-1/2}W^{(i)}\in A\right)-P(Z\in A)\right| \\
\leq &K\left(\beta_0\right)\max\left\{\beta_0,\left(\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{E}\left|\Sigma_i^{-1/2}\xi_j\right|^4\right)^{1/2}\left(\left|\log\left(\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{E}\left|\Sigma_i^{-1/2}\xi_j\right|^4\right)\right|\vee 1\right)\right\}
\end{aligned}\tag{34}$$

Set $\beta_*=0.19$ and $\sigma_*=(1-\beta_*)^{1/2}=0.9$. Recall that

$$\bar{\beta}=\max\left\{\beta_0,\left(\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\right)^{1/2}\left(\left|\log\left(\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\right)\right|\vee 1\right)\right\}$$

- Case-1: $\bar{\beta}\leqslant\beta_*/d^{1/4}$.

  In this case, we have

  $$\mathbb{E}\left|\xi_i\right|^2\leqslant\sqrt{\mathbb{E}\left|\xi_i\right|^4}\leqslant\bar{\beta}\leqslant\beta_*/d^{1/4}\leqslant\beta_*$$

  for each unit vector $u\in\mathbb{R}^d$

  $$\langle\Sigma_i u,u\rangle=u^\top\Sigma_i u=u^\top\left(I_d-\mathbb{E}\xi_i\xi_i^\top\right)u=1-\mathbb{E}\left(\xi_i\cdot u\right)^2\geqslant 1-\mathbb{E}\left|\xi_i\right|^2\geqslant 1-\beta_*$$

  this implies $\sigma_i\geqslant\sigma_*$. Note that $x^{\frac{1}{2}}(|\log x|\vee 1)$ is an increasing function. Therefore, from (34), we have, by increasing $\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{E}\left|\Sigma_i^{-1/2}\xi_j\right|^4$ to $\frac{1}{\sigma_*^4}\sum_{j=1}^{n}\mathbb{E}\left|\xi_j\right|^4$,

  $$\sup_{A\in\mathcal{A}}\left|\mathbb{P}\left(\Sigma^{-1/2}W^{(i)}\in A\right)-P(Z\in A)\right|\leqslant K\left(\beta_0\right)\max\left\{\beta_0,\frac{2\bar{\beta}}{\sigma_*^2}\right\}\leqslant CK\left(\beta_0\right)\bar{\beta}\tag{35}$$

  Applying (23), (33), (35) and (3) we have

  $$\left|R_{21}\right|\leqslant\frac{C}{\varepsilon^2}\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^3\left(d^{1/4}\varepsilon+K\left(\beta_0\right)\bar{\beta}\right)\eta\tag{36}$$

  and

  $$\left|R_{22}\right|\leqslant\frac{C}{\varepsilon^2}\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\left(d^{1/4}\varepsilon+K\left(\beta_0\right)\bar{\beta}\right)|\log\eta|\tag{37}$$

  Now, if $\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4<\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^3$, choose $\eta=\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4/\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^3<1$. Note that we have by the Cauchy-Schwarz inequality

  $$\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^3\leqslant\sqrt{\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^2\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4}=\sqrt{d\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4}$$

Thus we obtain

$$|\log \eta| \leqslant \frac{1}{2}\log d - \frac{1}{2}\log\left(\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\right)$$

Since $\left(\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\right)^{1/2} \leqslant \bar{\beta} \leqslant \beta_*/d^{1/4}$, we have $\frac{1}{2}\log d \leq 2\log\beta_* - \log\left(\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\right)$ and

$$|\log \eta| \leqslant C\left|\log\left(\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\right)\right|$$

Hence, we have

$$|R_{21}| + |R_{22}| \leqslant \frac{C}{\varepsilon^2}\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\left(d^{1/4}\varepsilon + K\left(\beta_0\right)\bar{\beta}\right)\left(\left|\log\left(\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\right)\right| \vee 1\right) \tag{38}$$

This inequality also holds true if $\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4 \geqslant \sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^3$ by taking $\eta = 1$.

$$\begin{aligned}
&\sup_{A\in\mathcal{A}}|P(W\in A) - P(Z\in A)| \\
\leqslant& 4d^{1/4}\varepsilon + C(|\log\varepsilon| \vee 1)\left(\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\right)^{1/2} \\
&+ \frac{C}{\varepsilon^2}\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\left(d^{1/4}\varepsilon + K\left(\beta_0\right)\bar{\beta}\right)\left(\left|\log\left(\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\right)\right| \vee 1\right)
\end{aligned} \tag{39}$$

Choose $\varepsilon = \min\left\{\left[2C\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\left(\left|\log\left(\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4\right)\right| \vee 1\right)\right]^{1/2}, 1\right\}$ with the same absolute constant C as in the third term on the right-hand side of (39) If $\varepsilon < 1$, then (39) can be simplified to

$$\sup_{A\in\mathcal{A}}|P(W\in A) - P(Z\in A)| \leqslant \left(Cd^{1/4} + \frac{K\left(\beta_0\right)}{2}\right)\bar{\beta}$$

hence

$$\frac{\sup_{A\in\mathcal{A}}|P(W\in A) - P(Z\in A)|}{\bar{\beta}} \leqslant Cd^{1/4} + \frac{K\left(\beta_0\right)}{2} \tag{40}$$

If $\varepsilon = 1$, then $\sum_{i=1}^{n}\mathbb{E}\left|\xi_i\right|^4$ and $\bar{\beta}$ are bounded away from 0 by an absolute constant; hence

$$\frac{\sup_{A\in\mathcal{A}}|P(W\in A) - P(Z\in A)|}{\bar{\beta}} \leqslant \frac{1}{\bar{\beta}} \leqslant C \tag{41}$$

- Case-2: $\bar{\beta} > \beta_*/d^{1/4}$, we trivially estimate

$$\frac{\sup_{A\in\mathcal{A}}|P(W\in A) - P(Z\in A)|}{\bar{\beta}} \leqslant \frac{1}{\bar{\beta}} \leqslant \frac{d^{1/4}}{\beta_*} \leqslant Cd^{1/4} \tag{42}$$

Combining both cases together, we have

$$\frac{\sup_{A\in\mathcal{A}}|P(W\in A)-P(Z\in A)|}{\bar{\beta}}\leqslant Cd^{1/4}+\frac{K\left(\beta_0\right)}{2}$$

Note that the right-hand side of the above bound does not depend on W. Taking supremum over W, we obtain

$$K\left(\beta_0\right)\leqslant Cd^{1/4}+\frac{K\left(\beta_0\right)}{2} \tag{43}$$

4.3 The Outline of the Proof for Theorem 2.2

We first note that, for any $d\times d$ orthogonal matrix U, we have $UW=\sum_{i=1}^n U\xi_i, UZ\sim N\left(0,U\Sigma U^\top\right), \delta\left(UW,U\Sigma U^\top\right)=\delta(W,\Sigma)$ and $UB\in\mathcal{B}$ for all $B\in\mathcal{B}$. Therefore, it is enough to prove (5) when Σ is diagonal with positive entries. The proof is a combination of Zhilova (2020)'s smoothing, a Gaussian anti-concentration inequality for ellipsoids by Giessing and Fan (2020), the decomposition (15), and a recursive argument by Raič (2019a).

Fix $\beta_0>0$. Define

$$K'\left(\beta_0\right)=\sup_{W,\Sigma}\frac{\sup_{A\in\mathcal{B}}\left|\mathbb{P}(W\in A)-\mathbb{P}\left(\Sigma^{1/2}Z_0\in A\right)\right|}{\max\left\{\beta_0,\Psi(\delta(W,\Sigma))\right\}} \tag{44}$$

where $Z_0\sim N\left(0,I_d\right)$ and the first supremum is taken over the family of all sums $W=\sum_{i\in\mathcal{I}}\xi_i$ of finite number of independent centered random vectors with $\mathbb{E}\left|\xi_i\right|^4<\infty$, and diagonal matrices Σ with positive entries. We will obtain a recursive inequality for $K'\left(\beta_0\right)$ and prove that

$$K'\left(\beta_0\right)\leqslant C \tag{45}$$

for an absolute constant C that does not depend on β_0. Equation (5) then follows by sending $\beta_0\to 0$.

Now we fix a $W=\sum_{i=1}^n\xi_i, n\geqslant 1$, and Σ in the aforementioned family (will take sup in (5.53) . Let

$$\bar{\beta}=\max\left\{\beta_0,\Psi(\delta(W,\Sigma))\right\} \tag{46}$$

We write σ_j for the j -th diagonal entry of $\Sigma^{1/2}$. To prove theorem 2.2, we need some technical lemmas that is applicable in the case of Euclidean ball.

Lemma 4.6. For any $A\in\mathcal{B}$ and $\varepsilon>0$, there exists $a\mathrm{C}^\infty$ function $\tilde{h}_{A,\varepsilon}$ (which depends only on A and ε) such that

$$\tilde{h}_{A,\varepsilon}(x)=1 \text{ for } x\in A,\quad \tilde{h}_{A,\varepsilon}(x)=0 \text{ for } x\in\mathbb{R}^d\backslash A^{\varepsilon},\quad 0\leqslant\tilde{h}_{A,\varepsilon}(x)\leqslant 1 \tag{47}$$

and

$$M_r\left(\tilde{h}_{A,\varepsilon}\right) \leqslant \frac{C}{\varepsilon^r} \quad \text{for } r=1,2,3,4 \tag{48}$$

and

$$\sup_{x\in\mathbb{R}^d}\left|\left\langle M, \operatorname{Hess}\tilde{h}_{A,\varepsilon}(x)\right\rangle_{H.S.}\right| \leqslant \frac{C}{\varepsilon^2}\left(\|M\|_{H.S.}+\sum_{j=1}^d |M_{jj}|\right) \tag{49}$$

for any $d\times d$ matrix $M=(M_{jk})_{1\leqslant j,k\leqslant d}$, where C is an absolute constant that does not depend on A,ε or M.

Lemma 4.7. For any d -dimensional random vector W and any $\varepsilon>0$,

$$\sup_{A\in\mathcal{B}}|P(W\in A)-P(Z\in A)| \leqslant \sup_{A\in\mathcal{B}}\mathbb{P}\left(Z\in A^{\varepsilon}\backslash A\right)+\sup_{A\in\mathcal{B}}\left|\mathbb{E}\tilde{h}_{A,\varepsilon}(W)-\mathbb{E}\tilde{h}_{A,\varepsilon}(Z)\right| \tag{50}$$

where $\tilde{h}_{A,\varepsilon}$ is as in Lemma 4.6.

Set $\tilde{\sigma}:=\operatorname{tr}\left(\Sigma^2\right)^{1/4}$. The following anti-concentration inequality is an immediate consequence of Giessing and Fan (2020, Corollary 5):

Lemma 4.8. Assume $\tilde{\sigma}>0$. For any $\varepsilon>0$,

$$\sup_{A\in\mathcal{B}}\mathbb{P}\left(Z\in A^{\varepsilon}\backslash A\right)\leqslant C\tilde{\sigma}^{-1}\varepsilon$$

Lemma 4.9. Ψ is an increasing function on $(0,\infty)$. Moreover, $\Psi(cx)\leqslant(c+\Psi(c))\Psi(x)$ for all $x>0$ and $c\geqslant 1$.

We now fix $A\in\mathcal{B}$ (will take sup in (5.51)), $0<\varepsilon\leqslant\tilde{\sigma}$, write $h:=\tilde{h}_{A,\varepsilon}$ and proceed to bound $|\mathbb{E}h(W)-\mathbb{E}h(Z)|$ by the decomposition (15) Consider the solution f to the Stein equation (9), which is given by (10)

$$f(w)=\int_0^1 -\frac{1}{2(1-s)}\int_{\mathbb{R}^d}\left[h\left(\sqrt{1-s}w+\sqrt{s}\Sigma^{1/2}z\right)-\mathbb{E}h(Z)\right]\phi(z)dzds$$

Since h has bounded partial derivatives up to the third order, f is thrice differentiable with bounded partial derivatives. Using the integration by parts formula, we have for $1\leqslant j,k,l\leqslant d$ and any $0\leqslant c_0\leqslant 1$ that

$$\begin{aligned}\partial_{jk}f(w)=&\int_0^{c_0}\frac{1}{2\sqrt{s}}\int_{\mathbb{R}^d}\partial_j h\left(\sqrt{1-s}w+\sqrt{s}\Sigma^{1/2}z\right)\sigma_k^{-1}\partial_k\phi(z)dzds\\&+\int_{c_0}^1 -\frac{1}{2s}\int_{\mathbb{R}^d}h\left(\sqrt{1-s}w+\sqrt{s}\Sigma^{1/2}z\right)\sigma_j^{-1}\sigma_k^{-1}\partial_{jk}\phi(z)dzds\end{aligned} \tag{51}$$

and

$$\partial_{jkl} f(w) = \int_0^{c_0} \frac{\sqrt{1-s}}{2\sqrt{s}} \int_{\mathbb{R}^d} \partial_{jk} h\left(\sqrt{1-s}w + \sqrt{s}\Sigma^{1/2} z\right) \sigma_l^{-1} \partial_l \phi(z) dz ds$$
$$+ \int_{c_0}^{1} \frac{\sqrt{1-s}}{2s^{3/2}} \int_{\mathbb{R}^d} h\left(\sqrt{1-s}w + \sqrt{s}\Sigma^{1/2} z\right) \sigma_j^{-1}\sigma_k^{-1}\sigma_l^{-1} \partial_{jkl}\phi(z) dz ds. \tag{52}$$

We first bound R_1 in decomposition. Using the expression of $\partial_{jk} f$ in (51) with $c_0 = (\varepsilon/\tilde{\sigma})^2$, we have

$$R_1 = R_{11} + R_{12}$$

where

$$R_{11} = \sum_{j,k=1}^{d} \mathbb{E}\left[V_{jk} \int_0^{(\varepsilon/\tilde{\sigma})^2} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} \partial_j h\left(\sqrt{1-s}W + \sqrt{s}\Sigma^{1/2} z\right) \sigma_k^{-1} \partial_k \phi(z) dz ds\right]$$

and

$$R_{12} = \sum_{j,k=1}^{d} \mathbb{E}\left[V_{jk} \int_{(\varepsilon/\tilde{\sigma})^2}^{1} -\frac{1}{2s} \int_{\mathbb{R}^d} h\left(\sqrt{1-s}W + \sqrt{s}\Sigma^{1/2} z\right) \sigma_j^{-1}\sigma_k^{-1} \partial_{jk}\phi(z) dz ds\right]$$

For R_{11}, applying analogous arguments to (29), we obtain

$$|R_{11}| \leqslant C\tilde{\sigma}^{-1} \left\{\sum_{j,k=1}^{d} \sigma_k^{-2} \mathbb{E} V_{jk}^2\right\}^{1/2} \leqslant C \left\{\sum_{j,k=1}^{d} (\sigma_j\sigma_k)^{-2} \mathbb{E} V_{jk}^2\right\}^{1/2}$$

where we used the inequality $\tilde{\sigma} \geqslant \sigma_j$ to derive the last inequality. The triangle inequality yields, for V in (17),

$$\left\{\sum_{j,k=1}^{d} (\sigma_j\sigma_k)^{-2} \mathbb{E} V_{jk}^2\right\}^{1/2} \leqslant \left\| I_d - \mathrm{Var}\left(\Sigma^{-1/2} W\right)\right\|_{H.S.} + \frac{1}{2}\left\{\sum_{j,k=1}^{d} (\sigma_j\sigma_k)^{-2}\, \mathrm{Var}\left[\sum_{i=1}^{n} \xi_{ij}\xi_{ik}\right]\right\}^{1/2}$$

Besides, we have

$$\begin{aligned}\left\{\sum_{j,k=1}^{d} (\sigma_j\sigma_k)^{-2}\, \mathrm{Var}\left[\sum_{i=1}^{n} \xi_{ij}\xi_{ik}\right]\right\}^{1/2} &= \left\{\sum_{i=1}^{n}\sum_{j,k=1}^{d} (\sigma_j\sigma_k)^{-2}\, \mathrm{Var}\left[\xi_{ij}\xi_{ik}\right]\right\}^{1/2} \\ &\leqslant \left\{\sum_{i=1}^{n}\sum_{j,k=1}^{d} (\sigma_j\sigma_k)^{-2}\, \mathbb{E}\xi_{ij}^2\xi_{ik}^2\right\}^{1/2} \\ &= \left\{\sum_{i=1}^{n} \mathbb{E}\left[\left(\sum_{j=1}^{d} \sigma_j^{-2}\xi_{ij}^2\right)^2\right]\right\}^{1/2} \\ &= \left(\sum_{i=1}^{n} \mathbb{E}\left|\Sigma^{-1/2}\xi_i\right|^4\right)^{1/2}\end{aligned}$$

Consequently, we obtain

$$|R_{11}| \leqslant C\delta(W, \Sigma)$$

For R_{12}, we apply analogous arguments to (30) and obtain

$$|R_{12}| \leqslant C|\log(\varepsilon/\tilde{\sigma})| \left\{ \sum_{j,k=1}^{d} (\sigma_j\sigma_k)^{-2}\, \mathbb{E}V_{jk}^2 \right\}^{1/2} \leqslant C|\log(\varepsilon/\tilde{\sigma})|\delta(W, \Sigma)$$

Therefore,

$$|R_1| \leqslant C(|\log(\varepsilon/\tilde{\sigma})| \vee 1)\delta(W, \Sigma) \tag{53}$$

Next, we bound R_2 in 19. Using the expression of $\partial_{jkl} f$ in 52 with $c_0 = (\varepsilon/\tilde{\sigma})^2$, we have

$$R_{21} = \frac{1}{4}\sum_{i=1}^{n}\sum_{j,k,l=1}^{d} \mathbb{E}U\left(\xi'_{ij} - \xi_{ij}\right)\left(\xi'_{ik} - \xi_{ik}\right)\left(\xi'_{il} - \xi_{il}\right)\int_0^{(\varepsilon/\tilde{\sigma})^2} \frac{\sqrt{1-s}}{2\sqrt{s}} \times \int_{\mathbb{R}^d} \phi_{21}(z, s)dzds$$

$$R_{22} = \frac{1}{4}\sum_{i=1}^{n}\sum_{j,k,l=1}^{d} \mathbb{E}U\left(\xi'_{ij} - \xi_{ij}\right)\left(\xi'_{ik} - \xi_{ik}\right)\left(\xi'_{il} - \xi_{il}\right)\int_{(\varepsilon/\tilde{\sigma})^2}^{1} \frac{\sqrt{1-s}}{2s^{3/2}} \times \int_{\mathbb{R}^d} \phi_{22}(z, s)dzds$$

and

$$\begin{aligned}
\phi_{21}(z, s) = \Big[&\partial_{jk}h\left(\sqrt{1-s}\left(W + (1-U)\left(\xi'_i - \xi_i\right)\right) + \sqrt{s}\Sigma^{1/2}z\right) \\
&-\partial_{jk}h\left(\sqrt{1-s}\left(W + U\left(\xi'_i - \xi_i\right)\right) + \sqrt{s}\Sigma^{1/2}z\right)\Big]\,\sigma_l^{-1}\partial_l\phi(z) \\
\phi_{22}(z, s) = \Big[&h\left(\sqrt{1-s}\left(W + (1-U)\left(\xi'_i - \xi_i\right)\right) + \sqrt{s}\Sigma^{1/2}z\right) \\
&-h\left(\sqrt{1-s}\left(W + U\left(\xi'_i - \xi_i\right)\right) + \sqrt{s}\Sigma^{1/2}z\right)\Big]\,\sigma_j^{-1}\sigma_k^{-1}\sigma_l^{-1}\partial_{jkl}\phi(z)
\end{aligned}$$

Let $W^{(i)} = W - \xi_i$ for $i \in \{1, \ldots, n\}$. We will use the mean value theorem for the differences involving h in the above two expressions as in (5.25), the fact that ∇h is non-zero only in $A^{\varepsilon}\backslash A$ and bound

$$\mathbb{P}\left(\sqrt{1-s}W^{(i)} \in A_i^{\varepsilon}\backslash A_i \mid U, U', \xi_i, \xi'_i\right)$$

where $0 < s < 1, U'$ is a uniform random variable on [0,1] independent of everything else, and A_i is a Euclidean ball which may depend on U, U', ξ_i, ξ'_i, s and $\Sigma^{1/2}z$. We have by Lemma 4.8

$$\mathbb{P}\left(\sqrt{1-s}W^{(i)} \in A_i^{\varepsilon} \setminus A_i \mid U, U', \xi_i, \xi'_i\right) \leqslant C\frac{\varepsilon}{\tilde{\sigma}\sqrt{1-s}} + 2\sup_{A\in\mathcal{B}}\left|\mathbb{P}\left(W^{(i)} \in A\right) - \mathbb{P}(Z \in A)\right| \tag{54}$$

From (44), we have

$$\sup_{A\in\mathcal{B}}\left|\mathbb{P}\left(W^{(i)} \in A\right) - \mathbb{P}(Z \in A)\right| \leqslant K'\left(\beta_0\right)\max\left\{\beta_0, \Psi\left(\delta\left(W^{(i)}, \Sigma\right)\right)\right\} \tag{55}$$

Since

$$\left\|\operatorname{Var}\left(\Sigma^{-1/2}W\right)-\operatorname{Var}\left(\Sigma^{-1/2}W^{(i)}\right)\right\|_{H.S.}=\sqrt{\sum_{j,k=1}^{d}\left(\mathbb{E}\left(\Sigma^{-1/2}\xi_i\right)_j\left(\Sigma^{-1/2}\xi_i\right)_k\right)^2}$$
$$\leqslant\mathbb{E}\left|\Sigma^{-1/2}\xi_i\right|^2\leqslant\sqrt{\mathbb{E}\left|\Sigma^{-1/2}\xi_i\right|^4}$$

and $\sqrt{x}+\sqrt{y}\leqslant\sqrt{2(x+y)}$ for any $x,y\geqslant 0$, we have

$$\delta\left(W^{(i)},\Sigma\right)\leqslant\left\|I_d-\operatorname{Var}\left(\Sigma^{-1/2}W\right)\right\|_{H.S.}+\sqrt{\mathbb{E}\left|\Sigma^{-1/2}\xi_i\right|^4}+\sqrt{\sum_{\substack{j=1\\j\neq i}}^{n}\mathbb{E}\left|\Sigma^{-1/2}\xi_j\right|^4}$$
$$\leqslant\left\|I_d-\operatorname{Var}\left(\Sigma^{-1/2}W\right)\right\|_{H.S.}+\sqrt{2\sum_{j=1}^{n}\mathbb{E}\left|\Sigma^{-1/2}\xi_j\right|^4}\leqslant\sqrt{2}\delta(W,\Sigma)$$

Hence, we obtain by Lemma 4.9

$$\Psi\left(\delta\left(W^{(i)},\Sigma\right)\right)\leqslant 2\sqrt{2}\Psi(\delta(W,\Sigma))\leqslant 2\sqrt{2}\bar{\beta}$$

Thus we conclude

$$\sup_{A\in\mathcal{B}}\left|\mathbb{P}\left(W^{(i)}\in A\right)-P(Z\in A)\right|\leqslant K'\left(\beta_0\right)\max\left\{\beta_0,2\sqrt{2}\bar{\beta}\right\}=2\sqrt{2}K'\left(\beta_0\right)\bar{\beta}\tag{56}$$

Using the mean value theorem for R_{21},R_{22} and applying (48), (54), (56) and (3) we have

$$|R_{21}|+|R_{22}|\leqslant\frac{C\tilde{\sigma}^2}{\varepsilon^2}\sum_{i=1}^{n}\mathbb{E}\left|\Sigma^{-1/2}\xi_i\right|^4\left(\frac{\varepsilon}{\tilde{\sigma}}+2\sqrt{2}K'\left(\beta_0\right)\bar{\beta}\right)\tag{57}$$

where we also used the inequality $\max_{1\leqslant j\leqslant d}\sigma_j\leqslant\tilde{\sigma}$. From Lemmas 4.7-4.8, Equations (15), (53) and (57) we have

$$\begin{aligned}&\sup_{A\in\mathcal{B}}|\mathbb{P}(W\in A)-\mathbb{P}(Z\in A)|\\&\leqslant C\tilde{\sigma}^{-1}\varepsilon+C(|\log(\varepsilon/\tilde{\sigma})|\vee 1)\delta(W,\Sigma)+\frac{C\tilde{\sigma}^2}{\varepsilon^2}\sum_{i=1}^{n}\mathbb{E}\left|\Sigma^{-1/2}\xi_i\right|^4\left(\frac{\varepsilon}{\tilde{\sigma}}+K'\left(\beta_0\right)\bar{\beta}\right)\end{aligned}\tag{58}$$

Choose $\varepsilon=\min\left\{\tilde{\sigma}\left[2C\sum_{i=1}^{n}\mathbb{E}\left|\Sigma^{-1/2}\xi_i\right|^4\right]^{1/2},\tilde{\sigma}\right\}$ for the same absolute constant C as in the third term on the right-hand side of (58)

- If $\varepsilon<\tilde{\sigma}$, then from (58)

$$\sup_{A\in\mathcal{B}}|\mathbb{P}(W\in A)-\mathbb{P}(Z\in A)|\leqslant\left(C+\frac{K'\left(\beta_0\right)}{2}\right)\bar{\beta}$$

hence

$$\frac{\sup_{A\in\mathcal{B}}|\mathbb{P}(W\in A)-\mathbb{P}(Z\in A)|}{\bar{\beta}}\leqslant C+\frac{K'\left(\beta_0\right)}{2}\tag{59}$$

- If $\varepsilon = \tilde{\sigma}$, then $\sum_{i=1}^{n} \mathbb{E} \left|\Sigma^{-1/2}\xi_i\right|^4$ and $\bar{\beta}$ are bounded away from 0 by an absolute constant; hence,
$$\frac{\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)|}{\bar{\beta}} \leqslant \frac{1}{\bar{\beta}} \leqslant C$$

Note that the right-hand sides of the above two bounds do not depend on W or Σ. Taking supremum over W and Σ, we obtain

$$K'(\beta_0) \leqslant C + \frac{K'(\beta_0)}{2} \tag{60}$$

This implies (45) hence (5).

4.4 Proof of Theorem 2.3

Fix $A \in \mathcal{B}$ (will take sup in the end of the proof), $\varepsilon > 0$, write $h := \tilde{h}_{A,\varepsilon}$ as in Lemma 4.6 and proceed to bound $|\mathbb{E}h(W) - \mathbb{E}h(Z)|$ by the decomposition (15) . Consider the solution f to the Stein equation (9), which is given by (10). Note that we can rewrite f as

$$f(w) = \int_0^1 -\frac{1}{2(1-s)} \mathbb{E}[h(\sqrt{1-s}w + Z) - \mathbb{E}h(Z)]ds$$

Since h has bounded partial derivatives up to the fourth order, f is four times differentiable and

$$\nabla^r f(w) = \int_0^1 -\frac{(1-s)^{r/2-1}}{2} \mathbb{E}\left[\nabla^r h(\sqrt{1-s}w + \sqrt{s}Z)\right] ds \tag{61}$$

We first bound R_1 in (16). Using (61), we obtain

$$R_1 = -\frac{1}{2} \int_0^1 \mathbb{E}\left[\langle V, \operatorname{Hess} h(W^s)\rangle_{H.S.}\right] ds$$

where $W^s := \sqrt{1-s}W + \sqrt{s}Z$. Since ∇h is non-zero only in $A^\varepsilon \setminus A$, we have

$$R_1 = -\frac{1}{2} \int_0^1 \mathbb{E}\left[\langle V, \operatorname{Hess} h(W^s)\rangle_{H.S.} 1_{\{W^s \in A^\varepsilon \setminus A\}}\right] ds$$

Therefore, using (49) we obtain

$$\begin{aligned}
|R_1| &\leqslant \frac{1}{2} \int_0^1 \mathbb{E} \left|\langle V, \operatorname{Hess} h(W^s)\rangle_{H.S.}\right| 1_{\{W^s \in A^\varepsilon \setminus A\}} ds \\
&\leqslant \frac{C}{\varepsilon^2} \int_0^1 \mathbb{E} \left(\|V\|_{H.S.} + \sum_{j=1}^{d} |V_{jj}| \right) 1_{\{W^s \in A^\varepsilon \setminus A\}} ds \\
&= \frac{C}{\varepsilon^2} \int_0^1 \mathbb{E} \left(\|V\|_{H.S.} + \sum_{j=1}^{d} |V_{jj}| \right) \mathbb{P}\left(W^s \in A^\varepsilon \setminus A \mid \xi\right) ds
\end{aligned}$$

Since Z is independent of ξ, Lemma 4.8 yields

$$\mathbb{P}\left(W^s \in A^\varepsilon \setminus A \mid \xi\right) \leqslant \frac{C\varepsilon}{\tilde{\sigma}\sqrt{s}}$$

Thus we deduce

$$|R_1| \leqslant \frac{C}{\tilde{\sigma}\varepsilon}\mathbb{E}\left(\|V\|_{H.S.} + \sum_{j=1}^{d}|V_{jj}|\right)\int_0^1 \frac{1}{\sqrt{s}}ds \leqslant \frac{C}{\tilde{\sigma}\varepsilon}\mathbb{E}\left(\|V\|_{H.S.} + \sum_{j=1}^{d}|V_{jj}|\right)$$

Using (17) and the triangle inequality, we obtain

$$\begin{aligned}
&\mathbb{E}\left(\|V\|_{H.S.} + \sum_{j=1}^{d}|V_{jj}|\right)\\
\leqslant&\|\Sigma - \mathrm{Var}(W)\|_{H.S.} + \sum_{j=1}^{d}|\Sigma_{jj} - \mathrm{Var}\left(W_j\right)| + \frac{1}{2}\sqrt{\sum_{j,k=1}^{d}\sum_{i=1}^{n}\mathrm{Var}\left(\xi_{ij}\xi_{ik}\right)} + \frac{1}{2}\sum_{j=1}^{d}\sqrt{\sum_{i=1}^{n}\mathrm{Var}\left(\xi_{ij}^2\right)}\\
\leqslant&\|\Sigma - \mathrm{Var}(W)\|_{H.S.} + \sum_{j=1}^{d}|\Sigma_{jj} - \mathrm{Var}\left(W_j\right)| + \frac{1}{2}\sqrt{\sum_{j,k=1}^{d}\sum_{i=1}^{n}\mathbb{E}\left[\xi_{ij}^2\xi_{ik}^2\right]} + \frac{1}{2}\sum_{j=1}^{d}\sqrt{\sum_{i=1}^{n}\mathbb{E}\left[\xi_{ij}^4\right]}\\
\leqslant&\tilde{\delta}(W,\Sigma)
\end{aligned}$$

Therefore, we conclude

$$|R_1| \leqslant \frac{C}{\tilde{\sigma}\varepsilon}\tilde{\delta}(W,\Sigma) \tag{62}$$

Next we bound R_2 in (19). We rewrite it as

$$R_2 = \frac{n}{4}\sum_{j,k,l,m=1}^{d}\mathbb{E}\left[D_jD_kD_lD_mU(1-2U)\partial_{jklm}f(W+\tilde{D})\right] \tag{63}$$

where $\tilde{D} := UD + U'(1-2U)D$ and U' is a uniform random variable on [0,1] independent of everything else. Now we set $\tilde{W}^s := \sqrt{1-s}(W+\tilde{D}) + \sqrt{s}Z$. Then, using (61), we can rewrite R_2 as

$$R_2 = n\int_0^1 -\frac{1-s}{8}\mathbb{E}U(1-2U)\left\langle \nabla^4 h\left(\tilde{W}^s\right), D^{\otimes 4}\right\rangle ds$$

Since ∇h is non-zero only in $A^\varepsilon \setminus A$, we can further rewrite it as

$$R_2 = n\int_0^1 -\frac{1-s}{8}\mathbb{E}U(1-2U)\left\langle \nabla^4 h\left(\tilde{W}^s\right), D^{\otimes 4}\right\rangle 1_{\left\{\tilde{W}^s\in A^\varepsilon\setminus A\right\}}ds$$

Therefore, using (48), we obtain

$$
\begin{aligned}
|R_2| &\leqslant \frac{n}{8}\int_0^1 \mathbb{E}\left|\left\langle \nabla^4 h\left(\tilde{W}^s\right), D^{\otimes 4}\right\rangle\right| 1_{\left\{\tilde{W}^s\in A^\varepsilon\setminus A\right\}} ds \\
&\leqslant \frac{Cn}{\varepsilon^4}\int_0^1 \mathbb{E}|D|^4 1_{\left\{\tilde{W}^s\in A^\varepsilon\setminus A\right\}} ds \\
&= \frac{Cn}{\varepsilon^4}\int_0^1 \mathbb{E}|D|^4 \mathbb{P}\left(\tilde{W}^s \in A^\varepsilon \setminus A \mid D, U, U'\right) ds
\end{aligned}
$$

Since Z is independent of D, U and U', Lemma 4.8 yields

$$
\mathbb{P}\left(\tilde{W}^s \in A^\varepsilon \setminus A \mid D, U, U'\right) \leqslant \frac{C\varepsilon}{\tilde{\sigma}\sqrt{s}}
$$

Thus we conclude

$$
|R_2| \leqslant \frac{Cn}{\tilde{\sigma}\varepsilon^3}\mathbb{E}|D|^4 \int_0^1 \frac{1}{\sqrt{s}} ds \leqslant \frac{C}{\tilde{\sigma}\varepsilon^3}\sum_{i=1}^n \mathbb{E}\left|\xi_i\right|^4 \tag{64}
$$

From Lemmas 4.7-4.8 and Equations (15), (62), (64), we have

$$
\sup_{A\in\mathcal{B}} |\mathbb{P}(W\in A) - \mathbb{P}(Z\in A)| \leqslant C\tilde{\sigma}^{-1}\varepsilon + \frac{C}{\tilde{\sigma}\varepsilon}\tilde{\delta}(W,\Sigma) + \frac{C}{\tilde{\sigma}\varepsilon^3}\sum_{i=1}^n \mathbb{E}\left|\xi_i\right|^4 \tag{65}
$$

and by choosing $\varepsilon = \sqrt{\tilde{\delta}(W,\Sigma)}$, we obtain

$$
\sup_{A\in\mathcal{B}} |\mathbb{P}(W\in A) - \mathbb{P}(Z\in A)| \leqslant \frac{C}{\tilde{\sigma}}\sqrt{\tilde{\delta}(W,\Sigma)} \tag{66}
$$

This completes the proof.