Bernoulli **25**(4A), 2019, 2824–2853
https://doi.org/10.3150/18-BEJ1072

A multivariate Berry–Esseen theorem with explicit constants

MARTIN RAIČ

This paper is dedicated to the memory of Vidmantas Kastytis Bentkus (1949–2010).

University of Ljubljana, University of Primorska, and Institute of Mathematics, Physics and Mechanics, Slovenia. E-mail: martin.raic@fmf.uni-lj.si; url: valjhun.fmf.uni-lj.si/~raicm/

We provide a Lyapunov type bound in the multivariate central limit theorem for sums of independent, but not necessarily identically distributed random vectors. The error in the normal approximation is estimated for certain classes of sets, which include the class of measurable convex sets. The error bound is stated with explicit constants. The result is proved by means of Stein's method. In addition, we improve the constant in the bound of the Gaussian perimeter of convex sets.

Keywords: Berry–Esseen theorem; explicit constants; Lyapunov bound; multivariate central limit theorem; Stein's method

1. Introduction and results

Let $\mathcal{I}$ be a countable set (either finite or infinite) and let X_i, $i \in \mathcal{I}$, be independent $\mathbb{R}^d$-valued random vectors. Assume that $\mathbb{E}X_i = 0$ for all i and that $\sum_{i\in\mathcal{I}} \operatorname{Var}(X_i) = \mathbf{I}_d$. It is well known that in this case, the sum $W := \sum_{i\in\mathcal{I}} X_i$ exists almost surely and that $\mathbb{E}W = 0$ and $\operatorname{Var}(W) = \mathbf{I}_d$.

For $\mu \in \mathbb{R}^d$ and $\mathbf{\Sigma} \in \mathbb{R}^{d\times d}$, denote by $\mathcal{N}(\mu, \mathbf{\Sigma})$ the d-variate normal distribution with mean μ and covariance matrix $\mathbf{\Sigma}$. For a measurable set $A \subseteq \mathbb{R}^d$, let $\mathcal{N}(\mu, \mathbf{\Sigma})\{A\} := \mathbb{P}(Z \in A)$, and for a measurable function $f : \mathbb{R}^d \to \mathbb{R}$, denote $\mathcal{N}(\mu, \mathbf{\Sigma})\{f\} := \mathbb{E}[f(Z)]$, where $Z \sim \mathcal{N}(\mu, \mathbf{\Sigma})$.

Roughly speaking, the d-variate central limit theorem for this set-up says that if none of the summands X_i is "too large", the sum W approximately follows $\mathcal{N}(0, \mathbf{I}_d)$. The error can be measured and estimated in various ways. Here, we focus on the Lyapunov type bound

$$\sup_{A\in\mathcal{A}} \bigl|\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\}\bigr| \le K \sum_{i\in\mathcal{I}} \mathbb{E}|X_i|^3, \tag{1.1}$$

where $\mathcal{A}$ is a suitable class of subsets of $\mathbb{R}^d$ and where $|x|$ denotes the Euclidean norm of the vector x.

Fixing a class of sets for all dimensions d, an important question is the dependence of the constant K on the dimension. The latter has drawn the attention of many authors and was tackled by different techniques. The class of measurable convex sets appears as a natural extension of the classical univariate Berry–Esseen theorem. For this case and for identically distributed summands, Nagaev [18] uses Fourier transforms to derive a constant of order d. Bentkus [7] succeeds to derive a constant of order $d^{1/2}$ by the method of composition (Lindeberg–Bergström

1350-7265 © 2019 ISI/BS

method). Improving this method and taking advantage of new bounds on Gaussian perimeters of convex sets (see below), he obtains $K = 400d^{1/4}$ in [5]. In [6], the latter result is extended to not necessarily identically distributed summands, but with no explicit constant, just of order $d^{1/4}$.

In 1970, Stein [26] developed a new elegant approach to bound the error in the normal approximation. His method was subsequently extended and refined in many ways. Götze [16] derives (1.1) with $K = 157.85d + 10$ using Stein's method combined with induction. Combining with part of Bentkus's argument, Chen and Fang [10] succeed to improve this bound to $115d^{1/2}$. However, this is still of larger order than Bentkus's result.

There used to be certain doubts about the correctness of Götze's paper [16]. To present a more readable account of Götze's paper, Bhattacharya and Holmes wrote an exposition [8] of the arguments. However, they obtain a higher order dependence of the error rate on d, namely $d^{5/2}$. In Remark 2.2, we explain where they gain the extra factor of $d^{3/2}$.

Here, we combine Götze's and Bentkus's arguments to derive the following explicit variant of Bentkus's result:

Theorem 1.1. *For X_i and W as above and all measurable convex sets $A \subseteq \mathbb{R}^d$, we have*

$$\bigl|\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\}\bigr| \le \bigl(42d^{1/4} + 16\bigr) \sum_{i \in \mathcal{I}} \mathbb{E}|X_i|^3. \tag{1.2}$$

This result follows immediately from Theorems 1.2 and 1.3 below, also noticing the observations in Example 1.1.

To derive K in (1.1), it seems inevitable to include *Gaussian perimeters* of sets $A \in \mathcal{A}$ or quantities closely related to them. The Gaussian perimeter of a set $A \subseteq \mathbb{R}^d$ is defined as

$$\gamma(A) := \int_{\partial A} \phi_d(z) \mathcal{H}^{d-1}(\mathrm{d}z),$$

where ∂A denotes the topological boundary of A, $\mathcal{H}^{d-1}$ denotes the $(d-1)$-dimensional Hausdorff measure and $\phi_d(z) := (2\pi)^{-d/2} \exp(-|x|^2/2)$ denotes the standard d-variate Gaussian density.

Gaussian perimeters are closely related to Gaussian measures of neighborhoods of the boundary. Before stating it precisely, we introduce some notation:

- For a point $x \in \mathbb{R}^d$ and a non-empty set $A \subseteq \mathbb{R}^d$, denote by $\operatorname{dist}(x, A)$ the Euclidean distance from x to A.
- For a set $A \subseteq \mathbb{R}^d$, which is neither the empty set nor the whole $\mathbb{R}^d$, define the *signed distance function* of A as

$$\delta_A(x) := \begin{cases} -\operatorname{dist}\bigl(x, \mathbb{R}^d \setminus A\bigr); & x \in A, \\ \operatorname{dist}(x, A); & x \notin A. \end{cases}$$

  Moreover, for each $t \in \mathbb{R}$, define $A^t := \{x \in \mathbb{R}^d; \delta_A(x) \le t\}$. In addition, define $\varnothing^t := \varnothing$ and $(\mathbb{R}^d)^t := \mathbb{R}^d$.

- For $A \subseteq \mathbb{R}^d$, define

$$\gamma^*(A) := \sup\left\{\frac{1}{\varepsilon}\mathcal{N}(0, \mathbf{I}_d)\big\{A^\varepsilon \setminus A\big\}, \frac{1}{\varepsilon}\mathcal{N}(0, \mathbf{I}_d)\big\{A \setminus A^{-\varepsilon}\big\}; \varepsilon > 0\right\}.$$

- For a class of sets $\mathcal{A}$, define $\gamma(\mathcal{A}) := \sup_{A\in\mathcal{A}} \gamma(A)$ and $\gamma^*(\mathcal{A}) := \sup_{A\in\mathcal{A}} \gamma^*(A)$.

The following proposition is believed by some authors to be evident. However, though the proof is quite straightforward, the assertion is not immediate. As a special case of Proposition 3.1, it is proved in Section 3.

Proposition 1.1. *Let $\mathcal{A}$ be a class of certain convex sets. Suppose that $A^t \in \mathcal{A} \cup \{\varnothing\}$ for all $A \in \mathcal{A}$ and all $t \in \mathbb{R}$. Then we have $\gamma(\mathcal{A}) = \gamma^*(\mathcal{A})$.*

Let $\mathcal{C}_d$ be the class of *all* convex sets in $\mathbb{R}^d$. Denote $\gamma_d := \gamma(\mathcal{C}_d) = \gamma^*(\mathcal{C}_d)$. It is known that $\gamma_d \le 4d^{1/4}$ – see Ball [2]. Nazarov [19] shows that the order $d^{1/4}$ is correct and improved the upper bound *asymptotically*, showing that $\limsup_{d\to\infty} d^{-1/4}\gamma_d \le (2\pi)^{1/4} < 0.64$. Our next result provides an explicit bound, which is asymptotically even slightly better than Nazarov's bound.

Theorem 1.2. *For all $d \in \mathbb{N}$, we have*

$$\gamma_d \le \sqrt{\frac{2}{\pi}} + 0.59\big(d^{1/4} - 1\big) < 0.59d^{1/4} + 0.21. \tag{1.3}$$

We defer the proof to Section 3.

Remark 1.1. Though γ_d is of order $d^{1/4}$, this does not necessarily mean that this is the optimal order of the constant K in (1.1). This remains an open question.

There are interesting classes of sets $\mathcal{A}$ where there exist better bounds on $\gamma(\mathcal{A})$ than those of order $d^{1/4}$. For the class of all balls, $\gamma(\mathcal{A})$ can be bounded independently of the dimension – see Sazonov [22,23]. For the class of all rectangles, it is known that $\gamma(\mathcal{A})$ is at most of order $\sqrt{\log d}$, see Nazarov [19]. Apart from convex sets, other classes may also be interesting, e. g., the class of unions of balls which are at least Δ apart, where $\Delta > 0$ is a fixed number. Therefore, we derive a more general result; Theorem 1.1 will follow from the latter and Theorem 1.2.

To generalize Theorem 1.1, we shall consider a class $\mathcal{A}$ of measurable sets in $\mathbb{R}^d$. For each $A \in \mathcal{A}$, take a measurable function $\rho_A : \mathbb{R}^d \to \mathbb{R}$. The latter can be considered as a generalized signed distance function: typically, one can take $\rho_A = \delta_A$, but we allow for more general functions. For each $t \in \mathbb{R}$, define

$$A^{t|\rho} := \big\{x; \rho_A(x) \le t\big\}.$$

Next, define the generalized Gaussian perimeter as

$$\gamma^*(A \mid \rho) := \sup\left\{\frac{1}{\varepsilon}\mathcal{N}(0, \mathbf{I}_d)\{A^{\varepsilon|\rho} \setminus A\}, \frac{1}{\varepsilon}\mathcal{N}(0, \mathbf{I}_d)\{A \setminus A^{-\varepsilon|\rho}\}; \varepsilon > 0\right\},$$
$$\gamma^*(\mathcal{A} \mid \rho) := \sup_{A \in \mathcal{A}} \gamma^*(A \mid \rho).$$

We shall impose the following assumptions:

(A1) $\mathcal{A}$ is closed under translations and uniform scalings by factors greater than one.
(A2) For each $A \in \mathcal{A}$ and $t \in \mathbb{R}$, $A^{t|\rho} \in \mathcal{A} \cup \{\varnothing, \mathbb{R}^d\}$.
(A3) For each $A \in \mathcal{A}$ and $\varepsilon > 0$, either $A^{-\varepsilon|\rho} = \varnothing$ or $\{x; \rho_{A^{-\varepsilon|\rho}}(x) < \varepsilon\} \subseteq A$.
(A4) For each $A \in \mathcal{A}$, $\rho_A(x) \le 0$ for all $x \in A$ and $\rho_A(x) \ge 0$ for all $x \notin A$.
(A5) For each $A \in \mathcal{A}$ and each $y \in \mathbb{R}^d$, $\rho_{A+y}(x+y) = \rho_A(x)$ for all $x \in \mathbb{R}^d$.
(A6) For each $A \in \mathcal{A}$ and each $q \ge 1$, $|\rho_{qA}(qx)| \le q|\rho_A(x)|$ for all $x \in \mathbb{R}^d$.
(A7) For each $A \in \mathcal{A}$, ρ_A is non-expansive on $\{x; \rho_A(x) \ge 0\}$, i.e., $|\rho_A(x) - \rho_A(y)| \le |x - y|$ for all x, y with $\rho_A(x) \ge 0$ and $\rho_A(y) \ge 0$.
(A8) For each $A \in \mathcal{A}$, ρ_A is differentiable on $\{x; \rho_A(x) > 0\}$. Moreover, there exists $\kappa \ge 0$, such that

$$\left|\nabla\rho_A(x) - \nabla\rho_A(y)\right| \le \frac{\kappa|x - y|}{\min\{\rho_A(x), \rho_A(y)\}}$$

for all x, y with $\rho_A(x) > 0$ and $\rho_A(y) > 0$; throughout this paper, ∇ denotes the gradient.

In addition, we state the following optional assumption:

(A1′) $\mathcal{A}$ is closed under symmetric linear transformations with the smallest eigenvalue at least one.

Remark 1.2. It is natural to define ρ_A so that (A6) is satisfied with equality. However, for our main result, Theorem 1.3, only the inequality is needed.

Remark 1.3. Assumptions (A3)–(A8) are hereditary: if the pair $(\mathcal{A}, (\rho_A)_{A\in\mathcal{A}})$ meets them and if $\mathcal{B} \subseteq \mathcal{A}$, the pair $(\mathcal{B}, (\rho_B)_{B\in\mathcal{B}})$ meets them, too.

Remark 1.4. With $\rho_A = \delta_A$, one can easily check that Assumptions (A3)–(A7) are met. Assumption (A8) is motivated by Lemma 2.2 of Bentkus [5] (see Example 1.1 below).

The following is the main result of this paper.

Theorem 1.3. *Let $W = \sum_{i\in\mathcal{I}} X_i$ be as in Theorem 1.1 and let $\mathcal{A}$ be a class of sets meeting Assumptions* (A1)–(A8) (*along with the underlying functions ρ_A*). *Then for each $A \in \mathcal{A}$, the following estimate holds true*:

$$\left|\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\}\right| \le \max\left\{27, 1 + 53\gamma^*(\mathcal{A} \mid \rho)\sqrt{1+\kappa}\right\} \sum_{i\in\mathcal{I}} \mathbb{E}|X_i|^3. \tag{1.4}$$

In addition, if $\mathcal{A}$ also satisfies (A1′), *the preceding bound can be improved to*

$$\left|\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\}\right| \le \max\left\{27, 1 + 50\gamma^*(\mathcal{A} \mid \rho)\sqrt{1+\kappa}\right\} \sum_{i\in\mathcal{I}} \mathbb{E}|X_i|^3. \tag{1.5}$$

We provide the proof in the next section.

Remark 1.5. Though explicit, the constants in Theorem 1.3 seem to be far from optimal. Consider the classical case where $\mathcal{A}$ is the class of all half-lines $(-\infty, w]$, where w runs over $\mathbb{R}$. It is straightforward to check that $\mathcal{A}$ along with $\rho_A = \delta_A$ meets Assumptions (A1)–(A8) with $\kappa = 1$. Observing that $\gamma^*(\mathcal{A} \mid \rho) = \gamma^*(\mathcal{A}) = 1/\sqrt{2\pi}$, estimate (1.5) reduces to (1.1) with $K = 29.3$. This is much worse than $K = 4.1$ obtained by Chen and Shao [11] by Stein's method, let alone than $K = 0.5583$ obtained by Shevtsova [24] by Fourier methods.

Below we give further examples of classes of sets.

Example 1.1. Consider the class $\mathcal{C}_d$ of all measurable convex sets in $\mathbb{R}^d$, along with $\rho_A = \delta_A$, which is defined in $\mathcal{C}_d \setminus \{\varnothing, \mathbb{R}^d\}$. Clearly, the latter class satisfies (A1). It is easy to verify (A2). By Lemma 2.2 of Bentkus [5], (A8) is met with $\kappa = 1$. By Remark 1.4, all other assumptions are met, too.

Example 1.2. The class of all balls in $\mathbb{R}^d$ (excluding the empty set) along with $\rho_A = \delta_A$ meets (A1) and (A2). Since the balls are convex, it meets all Assumptions (A1)–(A8).

Example 1.3. For a class of ellipsoids, $\rho_A = \delta_A$ is not suitable because an ε-neighborhood of an ellipsoid is not an ellipsoid. However, one can set $\rho_A(x) := \delta_{\mathbf{Q}A}(\mathbf{Q}x)$, where $\mathbf{Q}$ is a linear transformation mapping A into a ball (may depend on A). Notice that $\mathbf{Q}$ must be non-expansive in order to satisfy (A7).

Remark 1.6. If the random vectors X_i are identically distributed, that is, if $\mathcal{I}$ has n elements and X_i follow the same distribution as $\xi/\sqrt{n}$, the sum $\sum_{i\in\mathcal{I}} \mathbb{E}|X_i|^3$ reduces to $n^{-1/2}\mathbb{E}|\xi|^3$. However, for the class of *centered* balls, this rate of convergence is suboptimal. Using Fourier analysis, Esseen [13] succeeds to derive a convergence rate of order $n^{-d/(d+1)}$ under the existence of the fourth moment. This is possible because of symmetry: that result is in fact an asymptotic expansion of first order with vanishing first term.

Recently, Stein's method has been used by Gaunt, Pickett and Reinert [15] to derive a convergence rate of order n^{-1}, but for sufficiently smooth radially symmetric test functions rather than the indicators of centered balls. Applying Stein's method to non-smooth test functions is not straightforward: non-smoothness of test functions needs to be compensated by a kind of smoothness of the distribution of W or its modifications.

In the present paper, this is resolved by a 'bootstrapping' argument which is essentially equivalent to Götze's [16] inductive argument. The probabilities of the sets in the class $\mathcal{A}$ are a kind of invariant (see (2.22) and (2.30)). In view of characteristic functions, this is similar to the argument introduced by Tihomirov [27], which combines Stein's idea with Fourier analysis. Instead of the set probabilities, the invariant are the expectations of functions $x \mapsto e^{\mathrm{i}\langle t,x\rangle}$ for t of order $O(\sqrt{n})$. This suffices to derive a convergence rate of order $n^{-1/2}$.

Esseen [13] succeeds to go beyond this rate (in dimensions higher than one) by deriving a kind of smoothness of the distribution of W directly: see Lemma 3 ibidem. This part of the argument seems to have no relationship with Stein's method. Similarly, Barbour and Čekanavičius [4] succeed to sharply estimate the error in the asymptotic expansions for integer random variables, but although the main argument is based on Stein's method, appropriate smoothness of modifications of W is needed and derived separately: see the inequality (5.7) ibidem.

Unfortunately, smoothness of W in view of Lemma 3 of Esseen [13] is unlikely to be useful in the argument used in this paper: another kind of smoothness would be desirable. Stein's method can be successfully combined with the concentration inequality approach, as in Chen and Fang [10]. Certain modifications of that approach could be a key to improvements.

Now consider an example of a class of non-convex sets.

Example 1.4. Let $\mathscr{A}$ be the class of all unions of disjoint intervals on the real line, such that the midpoints of any two intervals are at least Δ apart, where $\Delta > 0$ is fixed. In this case, δ_A is not a suitable function because it is not sufficiently smooth. We define ρ_A as follows (see Figure 1):

- If $x \ge b = \sup A$, define $\rho_A(x) := x - b$.
- If $x \le a = \inf A$, define $\rho_A(x) := a - x$.
- If $x \notin A$ and $b \le x \le a$, where b and a are the endpoints of two successive intervals, define

$$\rho_A(x) := \frac{1}{a-b}\left[\left(\frac{a-b}{2}\right)^2 - \left(x - \frac{b+a}{2}\right)^2\right].$$

- If x is an element of an interval with endpoints a and b, which constitutes A, define

$$\rho_A(x) := -\left(\frac{b-a}{2} - \left|x - \frac{a+b}{2}\right|\right).$$

Assumptions (A1), (A2) and (A4)–(A7) are easily verified (notice that some intervals may be joined or may disappear under $A \mapsto A^{t|\rho}$, but the distances between their midpoints never decrease). To verify (A3), observe that for $\rho_A(x) \ge \delta_A(x)/2$ for all $x \notin A$. Consequently, $A^{\varepsilon|\rho} \subseteq A^{2\varepsilon}$ for all $\varepsilon > 0$. Moreover, observe that $A^{-\varepsilon|\rho} = A^{-2\varepsilon}$ for all $\varepsilon > 0$. As a result, either $A^{-\varepsilon|\rho} = \varnothing$ or $\{x; \rho_{A^{-\varepsilon|\rho}}(x) < \varepsilon\} \subseteq \{x; \delta_{A^{-2\varepsilon}}(x) < 2\varepsilon\} \subseteq A$.

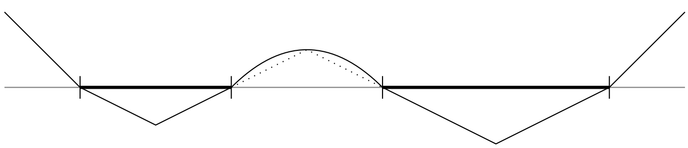

Figure 1. Construction of ρ_A for $A = [-2, 0] \cup [2, 5]$.

To verify (A8), observe that if $x \notin A$ and $b \le x \le a$, where b and a are the endpoints of two successive intervals, we have $\rho_A''(x) \le 2/(a-b) \le 1/(2\rho_A(x))$. Thus, (A8) is met with $\kappa = 1/2$.

Finally, we estimate $\gamma^*(\mathcal{A} \mid \rho)$. Let $A \in \mathcal{A}$ be a union of disjoint intervals from a_j to b_j, where j runs over $\mathcal{J}$, which is a set of successive numbers in $\mathbb{Z}$; we can assume that the intervals appear in the same order as the indices. Since $A^{\varepsilon|\rho} \subseteq A^{2\varepsilon}$ and $A^{-\varepsilon|\rho} = A^{-2\varepsilon}$, we have $\mathcal{N}(0,1)(A^\varepsilon \setminus A) \le \int_0^{2\varepsilon} \sum_{j \in \mathcal{J}} [\phi(a_j - t) + \phi(b_j + t)]\,\mathrm{d}t$ and $\mathcal{N}(0,1)(A \setminus A^{-\varepsilon}) \le \int_{-2\varepsilon}^0 \sum_{j \in \mathcal{J}} [\phi(a_j - t) + \phi(b_j + t)]\,\mathrm{d}t$, where ϕ denotes the standard univariate normal density, i. e., $\phi(x) = (2\pi)^{-1/2} e^{-x^2/2}$. Fix t, consider the terms with a_j and b_j separately, and split the sums over the indices where $a_j - t$ and $b_j + t$ are positive or negative. Estimating $a_{j+n} - a_j \ge \frac{a_{j+n-1}+b_{j+n-1}}{2} - \frac{a_j+b_j}{2} \ge (n-1)\Delta$ and $b_{j+n} - b_j \ge \frac{a_{j+n}+b_{j+n}}{2} - \frac{a_{j+1}+b_{j+1}}{2} \ge (n-1)\Delta$, and applying monotonicity of ϕ on $(-\infty, 0]$ and on $[0, \infty)$, we obtain after some calculation

$$\begin{aligned}\frac{1}{\varepsilon} \max\bigl\{\mathcal{N}(0,1)\bigl\{A^\varepsilon \setminus A\bigr\}, \mathcal{N}(0,1)\bigl\{A \setminus A^{-\varepsilon}\bigr\}\bigr\} &\le \frac{8}{\sqrt{2\pi}}\left(2 + \sum_{n=1}^\infty e^{-n^2\Delta^2/2}\right)\\ &\le \frac{8}{\sqrt{2\pi}}\left(2 + \int_0^\infty e^{-\Delta^2 x^2/2}\,\mathrm{d}x\right)\\ &= \frac{16}{\sqrt{2\pi}} + \frac{4}{\Delta}.\end{aligned}$$

The latter is the desired upper bound on $\gamma^*(\mathcal{A} \mid \rho)$.

2. Derivation of the bound in the central limit theorem

In this section, we prove Theorem 1.3. We shall use the ideas of Bentkus [5] regarding smoothing and Götze [16] regarding Stein's method. Before going to the proof, we need a few auxiliary results; we defer their proofs to the end of the section. We also introduce some further notation and conventions.

Let $x, u_1, u_2, \dots, u_r \in \mathbb{R}^d$. By $\langle \nabla^r f(x), u_1 \otimes u_2 \otimes \cdots \otimes u_r \rangle$, we denote the r-th order derivative of f at x in directions $u_1, u_2, \dots, u_r$. By components, if $u_i = (u_{i1}, u_{i2}, \dots, u_{id})$, we have

$$\bigl\langle \nabla^r f(x), u_1 \otimes u_2 \otimes \cdots \otimes u_r \bigr\rangle = \sum_{j_1, j_2, \dots, j_r} \frac{\partial^r f(x)}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_r}} u_{1j_1} u_{2j_2} \cdots u_{rj_r}.$$

Thus, $\nabla^r f(x)$ is a symmetric tensor of order r. We identify 2-tensors with linear maps or their matrices by $u \otimes v \equiv uv^T$. Observe that the Laplace operator can then be expressed as

$$\Delta f(x) = \bigl\langle \nabla^2 f(x), \mathbf{I}_d \bigr\rangle. \tag{2.1}$$

By $|T|_\vee$, we denote the *injective norm* of tensor T, that is

$$|T|_\vee := \sup_{|u_1|, |u_2|, \dots, |u_r| \le 1} \bigl|\langle T, u_1 \otimes u_2 \otimes \cdots \otimes u_r \rangle\bigr|.$$

For symmetric tensors, the supremum can be taken just over equal u_i:

Proposition 2.1 (Banach [3]; Bochnak and Siciak [9]). *If T is a symmetric tensor of order r, then $|T|_\vee = \sup_{|u|\le 1} |\langle T, u^{\otimes r}\rangle|$.*

Next, denote

$$M_0^*(f) := \frac{1}{2}\Big[\sup_{w\in\mathbb{R}^d} f(w) - \inf_{w\in\mathbb{R}^d} f(w)\Big],$$

$$M_r(f) := \sup_{\substack{w,z\in\mathbb{R}^d\\ w\neq z}} \frac{|\nabla^{r-1} f(w) - \nabla^{r-1} f(z)|_\vee}{|w-z|}; \qquad r = 1, 2, 3, \dots$$

If f is not everywhere $(r-1)$-times differentiable, we put $M_r(f) = \infty$.

Remark 2.1. This way, if $M_r(f) < \infty$, then $\nabla^{r-1} f$ exists everywhere and is Lipschitzian. In this case, by Rademacher's theorem (see Federer [14], Theorem 3.1.6), $\nabla^{r-1} f$ is almost everywhere differentiable. In addition, $M_r(f) = \sup_x |\nabla^r f(x)|_\vee$, where the supremum runs over all points where $\nabla^{r-1} f$ is differentiable.

Now we turn to auxiliary results regarding smoothing. The following one is a counterpart of Lemma 2.3 of Bentkus [5].

Lemma 2.1. *Let $\mathcal{A}$ be a class of sets which, along with the underlying functions ρ_A, meets Assumptions* (A1)–(A8). *Then for each $A \in \mathcal{A}$ and each $\varepsilon > 0$, there exist functions $f_A^{\varepsilon}, f_A^{-\varepsilon} : \mathbb{R}^d \to \mathbb{R}$, such that*:

(1) $0 \le f_A^{\varepsilon}, f_A^{-\varepsilon} \le 1$.
(2) $f_A^{\varepsilon}(x) = 1$ *for all* $x \in A$ *and* $f_A^{\varepsilon}(x) = 0$ *for all* $x \in \mathbb{R}^d \setminus A^{\varepsilon|\rho}$.
(3) $f_A^{-\varepsilon}(x) = 1$ *for all* $x \in A^{-\varepsilon|\rho}$ *and* $f_A^{-\varepsilon}(x) = 0$ *for all* $x \in \mathbb{R}^d \setminus A$.
(4) *The following bounds hold true*:

$$M_1\big(f_A^{\varepsilon}\big) \le \frac{2}{\varepsilon}, \qquad M_1\big(f_A^{-\varepsilon}\big) \le \frac{2}{\varepsilon}, \qquad M_2\big(f_A^{\varepsilon}\big) \le \frac{4(1+\kappa)}{\varepsilon^2}, \qquad M_2\big(f_A^{-\varepsilon}\big) \le \frac{4(1+\kappa)}{\varepsilon^2}.$$

(5) *For each $u \in (0,1)$, $\{x;\ f_A^{\varepsilon}(x) \ge u\} \in \mathcal{A} \cup \{\varnothing, \mathbb{R}^d\}$ and $\{x;\ f_A^{-\varepsilon}(x) \ge u\} \in \mathcal{A} \cup \{\varnothing, \mathbb{R}^d\}$.*

Proof. First, define $f_A^{\varepsilon}(x) := g(\frac{\rho_A(x)}{\varepsilon})$, where

$$g(x) := \begin{cases} 1; & x \le 0, \\ 1 - 2x^2; & 0 \le x \le 1/2, \\ 2(1-x)^2; & 1/2 \le x \le 1, \\ 0; & x \ge 1. \end{cases}$$

Requirements (1) in (2) are immediate, while (3) is irrelevant for f_A^ε. To prove (5), observe that $\{x;\, f_A^\varepsilon(x) \ge u\} = A^{\varepsilon g^{-1}(u)|\rho}$. Now we turn to (4). First, notice that $M_1(f_A^\varepsilon) \le 2/\varepsilon$ because $M_1(\rho_A) \le 1$ in $M_1(g) = 2$. Next, f_A^ε is continuously differentiable: see supplementary material [20]. Letting $B := \{x;\, \rho_A(x) \le 0\}$, take $x, y \in \mathbb{R}^d \setminus B$ with $\rho_A(x) \ge \rho_A(y)$ and estimate

$$\begin{aligned}\left|\nabla f_A^\varepsilon(x) - \nabla f_A^\varepsilon(y)\right| &\le \frac{1}{\varepsilon}\left|g'\left(\frac{\rho_A(x)}{\varepsilon}\right) - g'\left(\frac{\rho_A(y)}{\varepsilon}\right)\right|\left|\nabla \rho_A(x)\right| \\ &\quad + \frac{1}{\varepsilon}\left|g'\left(\frac{\rho_A(y)}{\varepsilon}\right)\right|\left|\nabla \rho_A(x) - \nabla \rho_A(y)\right|.\end{aligned}$$

In the first term, we apply $M_2(g) = 4$ and $M_1(\rho_A) \le 1$, while in the second, we apply $|g'(t)| \le 4t$ and (A8). Combining these estimates, we obtain $|\nabla f_A^\varepsilon(x) - \nabla f_A^\varepsilon(y)| \le 4(1+\kappa)\varepsilon^{-2}|x-y|$, noticing that we may drop the assumption that $\rho_A(x) \ge \rho_A(y)$. In other words, on $\mathbb{R}^d \setminus B$, ∇f_A^ε is Lipschitzian with constant $4(1+\kappa)\varepsilon^{-2}$. Trivially, this also holds true in the interior of B. Since ∇f_A^ε is continuous, this also holds true on the closures of both sets. Since for each $x \in \overline{\operatorname{Int} B}$ and each $y \in \overline{\mathbb{R}^d \setminus B}$, there exists z on the line segment with endpoints x and y, which is an element of both sets, ∇f_A^ε is Lipschitzian with the above-mentioned constant on the whole $\mathbb{R}^d$. Thus, f_A^ε meets all relevant requirements.

Now define $f_A^{-\varepsilon} \equiv 0$ if $A^{-\varepsilon} = \varnothing$ and $f_A^{-\varepsilon} := f_{A^{-\varepsilon}}^\varepsilon$ otherwise. From the above and from Assumption (A3), it follows that this function also satisfies all relevant requirements. This completes the proof. □

Throughout this section, $\mathbf{\Sigma}$ will refer to a positive-definite matrix $\mathbf{\Sigma}$ with the largest eigenvalue at most one and with the smallest eigenvalue σ^2, where $\sigma > 0$.

Lemma 2.2. *Let $\mathcal{A}$ be a class of sets, which, along with the underlying functions ρ_A, meets Assumptions (A1) and (A6). Then the following estimates hold true for all $\varepsilon > 0$:*

$$\mathcal{N}(\mu, \mathbf{\Sigma})\left\{A^{\varepsilon|\rho} \setminus A\right\} \le \frac{\gamma^*(\mathcal{A} \mid \rho)\varepsilon}{\sigma} \quad \textit{and} \quad \mathcal{N}(\mu, \mathbf{\Sigma})\left\{A \setminus A^{-\varepsilon|\rho}\right\} \le \frac{\gamma^*(\mathcal{A} \mid \rho)\varepsilon}{\sigma}.$$

Proof. Take independent random vectors $Z \sim \mathcal{N}(0, \mathbf{I}_d)$ and $R \sim \mathcal{N}(\mu, \mathbf{\Sigma} - \sigma^2\mathbf{I}_d)$. Clearly, $\sigma Z + R \sim \mathcal{N}(\mu, \mathbf{\Sigma})$. Now observe that, by (A5),

$$\begin{aligned}\mathcal{N}(\mu, \mathbf{\Sigma})\left\{A^{\varepsilon|\rho} \setminus A\right\} &= \mathbb{P}\left(\sigma Z + R \in A^{\varepsilon|\rho} \setminus A\right) \\ &= \mathbb{P}\left[Z \in \sigma^{-1}(A^{\varepsilon|\rho} - R) \setminus \sigma^{-1}(A - R)\right] \\ &= \mathbb{P}\left[Z \in \sigma^{-1}\left((A-R)^{\varepsilon|\rho}\right) \setminus \sigma^{-1}(A - R)\right] \\ &= \mathbb{E}\left[\mathcal{N}(0, \mathbf{I}_d)\left\{\sigma^{-1}\left((A-R)^{\varepsilon|\rho}\right) \setminus \sigma^{-1}(A - R)\right\}\right].\end{aligned}$$

From (A6), it follows that $\sigma^{-1}B^{\varepsilon|\rho} \subseteq (\sigma^{-1}B)^{(\varepsilon/\sigma)|\rho}$ for all $B \in \mathcal{A}$. Therefore,

$$\mathcal{N}(\mu, \mathbf{\Sigma})\left\{A^{\varepsilon|\rho} \setminus A\right\} \le \mathbb{E}\left[\mathcal{N}(0, \mathbf{I}_d)\left\{\left(\sigma^{-1}(A-R)\right)^{(\varepsilon/\sigma)|\rho} \setminus \left(\sigma^{-1}(A-R)\right)\right\}\right] \le \frac{\gamma^*(\mathcal{A} \mid \rho)\varepsilon}{\sigma}$$

(notice that $\sigma^{-1}(A - R) \in \mathscr{A}$ by (A1)). Analogously, we obtain $\mathcal{N}(\mu, \mathbf{\Sigma})\{A \setminus A^{-\varepsilon}\} \le \gamma^*(\mathscr{A} \mid \rho)\varepsilon/\sigma$. This completes the proof. □

Lemma 2.3. *Let a class $\mathscr{A}$ along with the underlying functions ρ_A meet Assumptions* (A1)–(A8). *Take a linear map $\mathbf{L}\colon \mathbb{R}^d \to \mathbb{R}^d$ with the smallest singular value at least one. Then the class $\tilde{\mathscr{A}} := \{\mathbf{L}A;\, A \in \mathscr{A}\}$ along with the underlying functions $\tilde{\rho}_{\tilde{A}}(x) := \rho_{\mathbf{L}^{-1}\tilde{A}}(\mathbf{L}^{-1}x)$ meets these assumptions with the same κ in* (A8). *Moreover,*

$$\gamma^*(\tilde{\mathscr{A}} \mid \tilde{\rho}) \le \|\mathbf{L}\|\gamma^*(\mathscr{A} \mid \rho). \tag{2.2}$$

Proof. Assumptions (A1), (A3), (A4), (A5) and (A6) are straightforward to check. To verify (A2), observe that

$$\tilde{A}^{t|\tilde{\rho}} = \{x;\, \tilde{\rho}_{\tilde{A}}(x) \le t\} = \{x;\, \rho_{\mathbf{L}^{-1}\tilde{A}}(\mathbf{L}^{-1}x) \le t\} = \mathbf{L}(\mathbf{L}^{-1}\tilde{A})^{t|\rho}.$$

Assumption (A7) follows from the fact that $\mathbf{L}^{-1}$ is non-expansive. To verify (A8), observe that, by the chain rule, $\nabla\tilde{\rho}_{\tilde{A}}(x) = \mathbf{L}^{-T}\nabla\rho_{\mathbf{L}^{-1}\tilde{A}}(\mathbf{L}^{-1}x)$, and use again that $\mathbf{L}^{-1}$ is non-expansive. Finally, observe that

$$\begin{aligned}\mathcal{N}(0, \mathbf{I}_d)\{\tilde{A}^{\varepsilon|\tilde{\rho}} \setminus \tilde{A}\} &= \mathcal{N}(0, \mathbf{I}_d)\{\mathbf{L}(\mathbf{L}^{-1}\tilde{A})^{\varepsilon|\rho} \setminus \tilde{A}\} \\ &= \mathcal{N}(0, \mathbf{L}^{-1}\mathbf{L}^{-T})\{(\mathbf{L}^{-1}\tilde{A})^{\varepsilon|\rho} \setminus \mathbf{L}^{-1}\tilde{A}\} \\ &\le \|\mathbf{L}\|\gamma^*(\mathscr{A} \mid \rho)\varepsilon\end{aligned}$$

by Lemma 2.2. An analogous inequality holds true for $\tilde{A} \setminus \tilde{A}^{-\varepsilon|\tilde{\rho}}$. Taking the supremum over $\tilde{A} \in \tilde{\mathscr{A}}$, we obtain (2.2). □

Now we turn to Stein's method, which will be implemented in view of the proof of Lemma 1 of Slepian [25]. We recall the procedure briefly; for an exposition, see Röllin [21] and Appendix H of Chernozhukov, Chetverikov and Kato [12]. Let f be a bounded measurable function. For $0 \le \alpha \le \pi/2$, define

$$\mathcal{U}_\alpha f(w) := \int_{\mathbb{R}^d} f(w\cos\alpha + z\sin\alpha)\phi_d(z)\,\mathrm{d}z. \tag{2.3}$$

For a random variable W, $\mathbb{E}[\mathcal{U}_\alpha f(W)]$ can be regarded as an interpolant between $\mathbb{E}[f(W)]$ and $\mathcal{N}(0, \mathbf{I}_d)\{f\}$. A straightforward calculation shows that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\mathcal{U}_\alpha f(w) = \mathcal{S}\mathcal{U}_\alpha f(w)\tan\alpha,$$

where $\mathcal{S}$ denotes the *Stein operator*:

$$\mathcal{S}g(w) := \Delta g(w) - \langle\nabla g(w), w\rangle \tag{2.4}$$

and where Δ denotes the Laplacian. Integrating over α and taking expectation, we find that

$$\mathbb{E} f(W) - \mathcal{N}(0, \mathbf{I}_d)\{f\} = -\int_0^{\pi/2} \mathbb{E}\big[\mathcal{S}\,\mathcal{U}_\alpha f(W)\big] \tan\alpha \,\mathrm{d}\alpha. \tag{2.5}$$

Notice that for $0 < \alpha \le \pi/2$, $\mathcal{U}_\alpha f$ is infinitely differentiable, so that $\mathcal{S}\,\mathcal{U}_\alpha f$ is well-defined. Differentiability can be shown by integration by parts. In particular, we shall need

$$\nabla^3 \mathcal{U}_\alpha f(w) = -\cot^3\alpha \int_{\mathbb{R}^d} f(w\cos\alpha + z\sin\alpha) \nabla^3 \phi_d(z)\,\mathrm{d}z \tag{2.6}$$

$$= -\frac{\cos^3\alpha}{\sin\alpha} \int_{\mathbb{R}^d} \nabla^2 f(w\cos\alpha + z\sin\alpha) \otimes \nabla\phi_d(z)\,\mathrm{d}z. \tag{2.7}$$

The proof is straightforward and is therefore left to the reader (cf. Section 2 of Bhattacharya and Holmes [8]). Observe that (2.7) remains true for all w if ∇f is Lipschitzian, that is, $M_2(f) < \infty$ (see Remark 2.1).

Now we turn to the *Stein expectation* $\mathbb{E}[\mathcal{S} g(W)]$. The following result, which is essentially a counterpart of Lemma 2.9 of Götze [16], expresses it in a way which is useful for its estimation.

Lemma 2.4 (Stein Expectation). *Let X_i, $i \in \mathcal{I}$, be independent $\mathbb{R}^d$-valued random vectors with sum W, which satisfies $\mathbb{E}W = 0$ and $\mathrm{Var}(W) = \mathbf{I}_d$. Then for any bounded three times continuously differentiable function g with bounded derivatives,*

$$\mathbb{E}\big[\mathcal{S} g(W)\big] = \sum_{i\in\mathcal{I}} \mathbb{E}\big[\big\langle \nabla^3 g(W_i + \theta X_i),\, X_i \otimes \tilde{X}_i^{\otimes 2} - (1-\theta) X_i^{\otimes 3}\big\rangle\big],$$

where $W_i = W - X_i$, $\tilde{X}_i$ is an independent copy of X_i, θ is uniformly distributed over $[0, 1]$, *and $\tilde{X}_i$ and θ are independent of each other and all other variates.*

Proof. Recalling (2.1), write

$$\Delta g(W) = \big\langle \nabla^2 g(W), \mathbf{I}_d\big\rangle = \big\langle \nabla^2 g(W), \mathrm{Var}(W)\big\rangle = \sum_{i\in\mathcal{I}} \big\langle \nabla^2 g(W), \mathrm{Var}(X_i)\big\rangle$$

$$= \sum_{i\in\mathcal{I}} \big\langle \nabla^2 g(W), \mathbb{E}\big(X_i^{\otimes 2}\big)\big\rangle.$$

Plugging into (2.4), we obtain

$$\mathbb{E}\big[\mathcal{S} g(W)\big] = \sum_{i\in\mathcal{I}} \mathbb{E}\big[\big\langle \nabla^2 g(W_i + X_i), \mathbb{E}X_i^{\otimes 2}\big\rangle - \big\langle \nabla g(W_i + X_i), X_i\big\rangle\big]$$

$$= \sum_{i\in\mathcal{I}} \mathbb{E}\big[\big\langle \nabla^2 g(W_i + X_i), \tilde{X}_i^{\otimes 2}\big\rangle - \big\langle \nabla g(W_i + X_i), X_i\big\rangle\big].$$

Taylor expansion centered at W_i yields

$$\mathbb{E}\big[\mathcal{S}g(W)\big] = \sum_{i\in\mathcal{I}} \mathbb{E}\big[\big\langle\nabla^2 g(W_i), \tilde{X}_i^{\otimes 2}\big\rangle + \big\langle\nabla^3 g(W_i+\theta X_i), X_i\otimes\tilde{X}_i^{\otimes 2}\big\rangle - \big\langle\nabla g(W_i), X_i\big\rangle - \big\langle\nabla^2 g(W_i), X_i^{\otimes 2}\big\rangle - (1-\theta)\big\langle\nabla^3 g(W_i+\theta X_i), X_i^{\otimes 3}\big\rangle\big].$$

By independence, the first and the fourth term cancel and the third term vanishes because $\mathbb{E}X_i = 0$. This completes the proof. □

Now we turn to the estimation of several integrals related to the multivariate normal distribution. Define constants $c_0, c_1, c_2, \dots$ by

$$c_r := \int_{-\infty}^{\infty} \big|\phi_1^{(r)}(z)\big|\,\mathrm{d}z.$$

Lemma 2.5. *For each bounded measurable function f, each $r\in\mathbb{N}$ and each $u\in\mathbb{R}^d$, we have*

$$\left|\int_{\mathbb{R}^d} f(z)\big\langle\nabla^r\phi_d(z), u^{\otimes r}\big\rangle\,\mathrm{d}z\right| \le c_r M_0^*(f)|u|^r.$$

Proof. First, observe that since the function $F(x) = \int_{\mathbb{R}^d}\phi_d(z+x)\,\mathrm{d}z$ is constant, we have $\int_{\mathbb{R}^d}\langle\nabla^r\phi_d(z), u^{\otimes r}\rangle\,\mathrm{d}z = \langle\nabla^r F(0), u^{\otimes r}\rangle = 0$. Therefore, f can be replaced by $f-b$, where b is arbitrary constant. As a result,

$$\left|\int_{\mathbb{R}^d} f(z)\big\langle\nabla^r\phi_d(z), u^{\otimes r}\big\rangle\,\mathrm{d}z\right| \le \sup|f-b|\int_{\mathbb{R}^d}\big|\big\langle\nabla^r\phi_d(z), u^{\otimes r}\big\rangle\big|\,\mathrm{d}z.$$

Choosing $b = (\inf f + \sup f)/2$, we have $\sup|f-b| = M_0^*(f)$. Next, since ϕ_d is spherically symmetric, we can replace u by $|u|e_1$, where $e_1 = (1,0,\dots,0)$. Writing $z = (z_1, z')$, we have $\langle\nabla^r\phi_d(z), e_1^{\otimes r}\rangle = \phi_1^{(r)}(z_1)\phi_{d-1}(z')$, so that

$$\int_{\mathbb{R}^d}\big|\big\langle\nabla^r\phi_d(z), e_1^{\otimes r}\big\rangle\big|\,\mathrm{d}z = \int_{\mathbb{R}}\big|\phi_1^{(r)}(z_1)\big|\,\mathrm{d}z_1\int_{\mathbb{R}^{d-1}}\phi_{d-1}\big(z'\big)\,\mathrm{d}z' = c_r.$$

Combining this with previous observations, the result follows. □

Remark 2.2. At this step, Bhattacharya and Holmes [8] gain the extra factor of $d^{3/2}$ in their bound. Instead of taking advantage of spherical symmetry, they estimate by components – see the estimates (3.12)–(3.15) ibidem. Götze's paper [16] comes to this step in the estimate (2.7) ibidem, where the result of Lemma 2.5 is actually used, but no argument is provided.

Lemma 2.6. *Let $f:\mathbb{R}^d\to\mathbb{R}$ be bounded and measurable. Take $0<\alpha\le\pi/2$. Then for all $r\in\mathbb{N}$ and all $\mu, u\in\mathbb{R}^d$,*

$$\big|\big\langle\mathcal{N}(\mu,\mathbf{\Sigma})\big\{\nabla^r\mathcal{U}_\alpha f\big\}, u^{\otimes r}\big\rangle\big| \le c_r M_0^*(f)\frac{\cos^r\alpha}{\sigma^r}|u|^r.$$

Remark 2.3. The expression $\mathcal{N}(\mu, \mathbf{\Sigma})\{\nabla^r \mathcal{U}_\alpha f\}$ is an expectation of a random tensor of order r and is therefore a deterministic tensor. This allows us to define $\langle \mathcal{N}(\mu, \mathbf{\Sigma})\{\nabla^r \mathcal{U}_\alpha f\}, u^{\otimes r}\rangle$.

Proof of Lemma 2.6. Write

$$\big\langle \mathcal{N}(\mu, \mathbf{\Sigma})\big\{\nabla^r \mathcal{U}_\alpha f\big\}, u^{\otimes r}\big\rangle = \mathbb{E}\big[\big\langle \nabla^r \mathcal{U}_\alpha f\big(\mathbf{\Sigma}^{1/2} Z + \mu\big), u^{\otimes r}\big\rangle\big] = \big\langle \nabla^r F(\mu), u^{\otimes r}\big\rangle, \tag{2.8}$$

where $F(\mu) := \mathbb{E}[\mathcal{U}_\alpha f(\mathbf{\Sigma}^{1/2} Z + \mu)]$ and where Z is a standard d-variate normal random vector. If Z' is another such vector independent of Z, we can write

$$F(\mu) = \mathbb{E}\big[f\big(\big(\mathbf{\Sigma}^{1/2} Z + \mu\big)\cos\alpha + Z' \sin\alpha\big)\big] = \int_{\mathbb{R}^d} f(\mu\cos\alpha + \mathbf{Q}_\alpha z)\phi_d(z)\,\mathrm{d}z,$$

where $\mathbf{Q}_\alpha := (\mathbf{\Sigma}\cos^2\alpha + \mathbf{I}_d \sin^2\alpha)^{1/2}$. Substituting $y = \mathbf{Q}_\alpha^{-1}\mu\cos\alpha + z$, we obtain

$$F(\mu) = \int_{\mathbb{R}^d} f(\mathbf{Q}_\alpha y)\phi_d\big(y - \mathbf{Q}_\alpha^{-1}\mu\cos\alpha\big)\,\mathrm{d}y.$$

Differentiation yields

$$\begin{aligned}\big\langle \nabla^r F(\mu), u^{\otimes r}\big\rangle &= (-1)^r \cos^r\alpha \int_{\mathbb{R}^d} f(\mathbf{Q}_\alpha y)\big\langle \nabla^r \phi_d\big(y - \mathbf{Q}_\alpha^{-1}\mu\cos\alpha\big), v^{\otimes r}\big\rangle\,\mathrm{d}y \\ &= (-1)^r \cos^r\alpha \int_{\mathbb{R}^d} f(\mu\cos\alpha + \mathbf{Q}_\alpha z)\big\langle \nabla^r \phi_d(z), v^{\otimes r}\big\rangle\,\mathrm{d}z,\end{aligned}$$

where $v = \mathbf{Q}_\alpha^{-1} u$. By Lemma 2.5, we can estimate

$$\big|\big\langle \nabla^r F(\mu), u^{\otimes r}\big\rangle\big| \le c_r \cos^r\alpha\, M_0^*(f)|v|^r. \tag{2.9}$$

Noting that $\|\mathbf{Q}_\alpha^{-1}\| = (\sigma^2\cos^2\alpha + \sin^2\alpha)^{-1/2} \le 1/\sigma$ and plugging into (2.9) and (2.8) in turn, the result follows. ☐

Lemma 2.7. *Let $\mathcal{A}$ be a family of measurable sets in $\mathbb{R}^d$, which, along with the underlying functions ρ_A, meets Assumptions* (A1)–(A8). *Take an $\mathbb{R}^d$-valued random vector W, such that there exist a vector $\mu \in \mathbb{R}^d$, a positive-definite matrix $\mathbf{\Sigma}$ and a constant $D \ge 0$, such that for each $A \in \mathcal{A}$,*

$$\big|\mathbb{P}(W \in A) - \mathcal{N}(\mu, \mathbf{\Sigma})(A)\big| \le D. \tag{2.10}$$

Then for each $\varepsilon > 0$ and each $f \in \{f_A^\varepsilon, f_A^{-\varepsilon}\}$, where f_A^ε and $f_A^{-\varepsilon}$ are as in Lemma 2.1, *we have*

$$\int_0^{\pi/2} \big|\mathbb{E}\big(\nabla^3 \mathcal{U}_\alpha f(W)\big)\big|_\vee \tan\alpha\,\mathrm{d}\alpha \le \frac{c_3}{6\sigma^3} + \sqrt{2(1+\kappa)c_1 c_3}\bigg(\frac{\gamma^*(\mathcal{A} \mid \rho)}{\sigma} + \frac{4D}{\varepsilon}\bigg). \tag{2.11}$$

Proof. Fix $A \in \mathcal{A}$ and $\varepsilon > 0$, and let $f = f_A^\varepsilon$ or $f = f_A^{-\varepsilon}$. In the first case, define $A_1 := A$ and $A_2 := A^{\varepsilon|\rho}$, while in the second case, define $A_1 := A^{-\varepsilon|\rho}$ and $A_2 := A$.

Similarly as observed in Remark 2.3, $\mathbb{E}(\nabla^3 \mathcal{U}_\alpha f(W))$ is a tensor because it is an expectation of a random tensor. Since the latter is symmetric, so is its expectation. By Proposition 2.1, its injective norm can be expressed as

$$\big|\mathbb{E}\big(\nabla^3 \mathcal{U}_\alpha f(W)\big)\big|_\vee = \sup_{|u|\le 1} \big|H_\alpha(u)\big|, \tag{2.12}$$

where

$$H_\alpha(u) := \big\langle \mathbb{E}\big(\nabla^3 \mathcal{U}_\alpha f(W)\big), u^{\otimes 3}\big\rangle = \mathbb{E}\big[\big\langle \nabla^3 \mathcal{U}_\alpha f(W), u^{\otimes 3}\big\rangle\big]. \tag{2.13}$$

Fix $0 < \beta < \pi/2$ and $u \in \mathbb{R}^d$ with $|u| \le 1$. We distinguish the cases $0 < \alpha \le \beta$ and $\beta < \alpha \le \pi/2$. In the first case, write, applying (2.7),

$$H_\alpha(u) = -\frac{\cos^3\alpha}{\sin\alpha} \int_{\mathbb{R}^d} F_\alpha(z) \big\langle \nabla\phi_d(z), u\big\rangle \, \mathrm{d}z,$$

where

$$F_\alpha(z) := \mathbb{E}\big[\big\langle \nabla^2 f(W\cos\alpha + z\sin\alpha), u^{\otimes 2}\big\rangle\big].$$

Notice that by Part (4) of Lemma 2.1 and Rademacher's theorem (see Remark 2.1), $\nabla^2 f$ is defined almost everywhere. By Fubini's theorem, the latter also holds for F_α. Moreover, where it is defined, we have, by Parts (2) and (4) of Lemma 2.1,

$$\big|F_\alpha(z)\big| \le \frac{4(1+\kappa)}{\varepsilon^2} \mathbb{P}(W\cos\alpha + z\sin\alpha \in A_2 \setminus A_1). \tag{2.14}$$

First, we estimate the right-hand side with W replaced by a d-variate normal random vector with the same mean and covariance matrix. Lemma 2.2 yields

$$\mathcal{N}\big(\mu\cos\alpha + z\sin\alpha, \boldsymbol{\Sigma}\cos^2\alpha\big)\{A_2 \setminus A_1\} \le \frac{\gamma^*(\mathcal{A} \mid \rho)\varepsilon}{\sigma\cos\alpha}. \tag{2.15}$$

To estimate the remainder, combine (2.10), (A1), (A2) and the fact that $A_1 \subseteq A_2$, resulting in

$$\big|\mathbb{P}(W\cos\alpha + z\sin\alpha \in A_2 \setminus A_1) - \mathcal{N}\big(\mu\cos\alpha + z\sin\alpha, \boldsymbol{\Sigma}\cos^2\alpha\big)\{A_2 \setminus A_1\}\big| \le 2D. \tag{2.16}$$

Combining (2.14), (2.15) and (2.16), we obtain

$$\big|F_\alpha(z)\big| \le \frac{4(1+\kappa)}{\varepsilon^2}\left(\frac{\gamma^*(\mathcal{A} \mid \rho)\varepsilon}{\sigma\cos\alpha} + 2D\right) \le \frac{4(1+\kappa)}{\varepsilon^2\cos\alpha}\left(\frac{\gamma^*(\mathcal{A} \mid \rho)\varepsilon}{\sigma} + 2D\right).$$

From Lemma 2.5, it follows that

$$\big|H_\alpha(u)\big| \le \frac{4(1+\kappa)c_1\cos^2\alpha}{\varepsilon\sin\alpha}\left(\frac{\gamma^*(\mathcal{A} \mid \rho)}{\sigma} + \frac{2D}{\varepsilon}\right). \tag{2.17}$$

Now we turn to the case $\alpha \geq \beta$, where we estimate $|H_\alpha(u)|$ in a different way. First, we estimate the right-hand side of (2.13) with W replaced by a d-variate normal random vector with the same mean and covariance matrix. Lemma 2.6 yields

$$\left|\left\langle \mathcal{N}(\mu, \mathbf{\Sigma})\left\{\nabla^3 \mathcal{U}_\alpha f\right\}, u^{\otimes 3}\right\rangle\right| \leq \frac{c_3 \cos^3 \alpha}{2\sigma^3}. \tag{2.18}$$

To estimate the remainder, write, applying (2.6),

$$H_\alpha(u) - \left\langle \mathcal{N}(\mu, \mathbf{\Sigma})\left\{\nabla^3 \mathcal{U}_\alpha f\right\}, u^{\otimes 3}\right\rangle = -\cot^3 \alpha \int_{\mathbb{R}^d} G_\alpha(z)\left\langle \nabla^3 \phi_d(z), u^{\otimes 3}\right\rangle \mathrm{d}z, \tag{2.19}$$

where

$$G_\alpha(z) := \mathbb{E}\left[f(W\cos\alpha + z\sin\alpha)\right] - \mathcal{N}\left(\mu\cos\alpha + z\sin\alpha, \mathbf{\Sigma}\cos^2\alpha\right)\{f\}.$$

Noting that $0 \leq f \leq 1$, write $f(x) = \int_0^1 \mathbf{1}(x \in \tilde{A}_t)\,\mathrm{d}t$, where $\tilde{A}_t := \{x;\, f(x) \geq t\}$. Consequently,

$$\begin{aligned} G_\alpha(z) &= \int_0^1 \left[\mathbb{P}(W\cos\alpha + z\sin\alpha \in \tilde{A}_t) - \mathcal{N}\left(\mu\cos\alpha + z\sin\alpha, \mathbf{\Sigma}\cos^2\alpha\right)\{\tilde{A}_t\}\right]\mathrm{d}t \\ &= \int_0^1 \left[\mathbb{P}(W \in \tilde{A}_{t,\alpha,z}) - \mathcal{N}(\mu, \mathbf{\Sigma})\{\tilde{A}_{t,\alpha,z}\}\right]\mathrm{d}t, \end{aligned}$$

where $\tilde{A}_{t,\alpha,z} := (\tilde{A}_t - \sin\alpha z)\cos^{-1}\alpha$. By Part (5) of Lemma 2.1, $\tilde{A}_t \in \mathscr{A} \cup \{\varnothing, \mathbb{R}^d\}$ for all $t \in (0,1)$. By Assumption (A1), the same is true for $\tilde{A}_{t,\alpha,z}$. Therefore, $|G_\alpha(z)| \leq D$ (observe that (2.10) is trivially true for $A \in \{\varnothing, \mathbb{R}^d\}$). Applying (2.18), (2.19) and Lemma 2.5, we obtain

$$\left|H_\alpha(u)\right| \leq \frac{c_3 \cos^3 \alpha}{2\sigma^3} + c_3 D \cot^3 \alpha. \tag{2.20}$$

Taking the supremum over u in (2.17) and (2.20), applying (2.12) and integrating, we obtain

$$\begin{aligned} \int_0^{\pi/2} \left|\mathbb{E}\left[\nabla^3 \mathcal{U}_\alpha f(W)\right]\right|_\vee \tan\alpha\,\mathrm{d}\alpha &\leq \int_0^\beta \frac{4(1+\kappa)c_1\cos\alpha}{\varepsilon}\left(\frac{\gamma^*(\mathscr{A} \mid \rho)}{\sigma} + \frac{2D}{\varepsilon}\right)\mathrm{d}\alpha \\ &\quad + c_3 \int_\beta^{\pi/2} \left(\frac{\cos^2\alpha\sin\alpha}{2\sigma^3} + D\cot^2\alpha\right)\mathrm{d}\alpha \\ &\leq \frac{4(1+\kappa)c_1\tan\beta}{\varepsilon}\left(\frac{\gamma^*(\mathscr{A} \mid \rho)}{\sigma} + \frac{2D}{\varepsilon}\right) \\ &\quad + \frac{c_3}{6\sigma^3} + c_3 D\cot\beta. \end{aligned} \tag{2.21}$$

Now choose β so that the sum of the terms with D is optimal. This occurs at $\beta = \arctan(\varepsilon \times \sqrt{\frac{c_3}{8(1+\kappa)c_1}})$. Plugging into (2.21), we obtain (2.11), completing the proof. □

Now we are ready to prove the main result.

Proof of Theorem 1.3. First, we prove the case where $\mathscr{A}$ also meets (A1′). Throughout the argument, fix $\mathscr{A}$ along with the underlying functions ρ_A. For each $\beta_0 > 0$, define

$$K(\beta_0) := \sup \frac{|\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\}|}{\max\{\sum_{i\in\mathcal{I}} \mathbb{E}|X_i|^3, \beta_0\}}, \tag{2.22}$$

where the supremum runs over the family of all sums $W = \sum_{i\in\mathcal{I}} X_i$ of independent random vectors with $\mathbb{E}X_i = 0$ and $\mathrm{Var}(W) = \mathbf{I}_d$, and over all $A \in \mathscr{A}$. Now fix $\beta_0 > 0$, a sum $W = \sum_{i\in\mathcal{I}} X_i$ in the aforementioned family and a set $A \in \mathscr{A}$. From Lemma 2.1, it follows that

$$0 \le \mathcal{N}(0, \mathbf{I}_d)\{f_A^{\varepsilon|\rho}\} - \mathcal{N}(0, \mathbf{I}_d)\{A\} \le \mathcal{N}(0, \mathbf{I}_d)\{A^{\varepsilon|\rho}\} - \mathcal{N}(0, \mathbf{I}_d)\{A\} \le \gamma^*(\mathscr{A} \mid \rho)\varepsilon,$$
$$0 \le \mathcal{N}(0, \mathbf{I}_d)\{A\} - \mathcal{N}(0, \mathbf{I}_d)\{f_A^{-\varepsilon|\rho}\} \le \mathcal{N}(0, \mathbf{I}_d)\{A\} - \mathcal{N}(0, \mathbf{I}_d)\{A^{-\varepsilon|\rho}\} \le \gamma^*(\mathscr{A} \mid \rho)\varepsilon.$$

Consequently,

$$\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\} \le \mathbb{E}f_A^{\varepsilon}(W) - \mathcal{N}(0, \mathbf{I}_d)\{f_A^{\varepsilon}\} + \gamma^*(\mathscr{A} \mid \rho)\varepsilon,$$
$$\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\} \ge \mathbb{E}f_A^{-\varepsilon}(W) - \mathcal{N}(0, \mathbf{I}_d)\{f_A^{-\varepsilon}\} - \gamma^*(\mathscr{A} \mid \rho)\varepsilon.$$

Therefore,

$$\begin{aligned}\bigl|\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\}\bigr| \le \max\bigl\{\bigl|\mathbb{E}f(W) - \mathcal{N}(0, \mathbf{I}_d)\{f\}\bigr|; f \in \bigl\{f_A^{\varepsilon}, f_A^{-\varepsilon}\bigr\}\bigr\} \\ + \gamma^*(\mathscr{A} \mid \rho)\varepsilon\end{aligned} \tag{2.23}$$

Let $f \in \{f_A^{\varepsilon}, f_A^{-\varepsilon}\}$, and let $\tilde{X}_i$ and θ be as in Lemma 2.4. Applying (2.5) and Lemma 2.4 in turn, and conditioning on X_i, $\tilde{X}_i$ and θ, we obtain

$$\mathbb{E}f(W) - \mathcal{N}(0, \mathbf{I}_d)\{f\} = -\int_0^{\pi/2} \sum_{i\in\mathcal{I}} \mathbb{E}\bigl[\bigl\langle T_i(\alpha), X_i \otimes \tilde{X}_i^{\otimes 2} - (1-\theta)X_i^{\otimes 3}\bigr\rangle\bigr] \tan\alpha \, \mathrm{d}\alpha,$$

where

$$T_i(\alpha) := \mathbb{E}\bigl[\nabla^3 \mathcal{U}_\alpha f(W_i + \theta X_i) | X_i, \tilde{X}_i, \theta\bigr]$$

is a random tensor of order three. Now estimate

$$\bigl|\mathbb{E}f(W) - \mathcal{N}(0, \mathbf{I}_d)\{f\}\bigr| \le \sum_{i\in\mathcal{I}} \mathbb{E}\biggl[\int_0^{\pi/2} \bigl|T_i(\alpha)\bigr|_\vee \tan\alpha \, \mathrm{d}\alpha \bigl(|X_i||\tilde{X}_i|^2 + (1-\theta)|X_i|^3\bigr)\biggr]. \tag{2.24}$$

To estimate $\int_0^{\pi/2} |T_i(\alpha)|_\vee \tan\alpha \, \mathrm{d}\alpha$, we shall use the conditional counterpart of Lemma 2.7 given X_i, $\tilde{X}_i$ and θ. To apply it, we need to estimate

$$D_{i,A} := \bigl|\mathbb{P}(W_i + \theta X_i \in A \mid X_i, \tilde{X}_i, \theta) - \mathcal{N}(\theta X_i, \mathbf{\Sigma}_i)\{A\}\bigr|,$$

where $\mathbf{\Sigma}_i = \mathrm{Var}(W_i)$. Assume that $\mathbf{\Sigma}_i$ is non-singular. In this case, we may write

$$D_{i,A} = \left|\mathbb{P}\big(\mathbf{\Sigma}_i^{-1/2} W_i \in \mathbf{\Sigma}_i^{-1/2}(A - \theta X_i) \mid X_i, \tilde{X}_i, \theta\big) - \mathcal{N}(0, \mathbf{I}_d)\big\{\mathbf{\Sigma}_i^{-1/2}(A - \theta X_i)\big\}\right|.$$

To estimate $D_{i,A}$, we apply the 'bootstrapping' argument: we refer to (2.22) with $\mathbf{\Sigma}_i^{-1/2} W_i$ in place of W, noting independence of W_i and $(X_i, \tilde{X}_i, \theta)$, and observing that $\mathbf{\Sigma}_i^{-1/2} W_i$ is a sum of independent random vectors with vanishing expectations and with $\mathrm{Var}(\mathbf{\Sigma}_i^{-1/2} W_i) = \mathbf{I}_d$. Furthermore, observe that, given θ and X_i, we have $\mathbf{\Sigma}_i^{-1/2}(A - \theta X_i) \in \mathscr{A}$ by (A1′). Denoting by σ_i^2 the smallest eigenvalue of $\mathbf{\Sigma}_i$ (with $\sigma_i > 0$), observe that $\mathbb{E}|\mathbf{\Sigma}_i^{-1/2} X_j|^3 \le \sigma_i^{-3} \mathbb{E}|X_j|^3$ (notice that $\sigma_i \le 1$). By (2.22), we have

$$D_{i,A} \le K(\beta_0) \max\left\{\frac{1}{\sigma_i^3} \sum_{j \in \mathcal{I} \setminus \{i\}} \mathbb{E}|X_j|^3, \beta_0\right\} \le \frac{K(\beta_0)\bar{\beta}}{\sigma_i^3},$$

where $\bar{\beta} := \max\{\sum_{j \in \mathcal{I}} \mathbb{E}|X_j|^3, \beta_0\}$. Applying Lemma 2.7 to the conditional distribution of W given X_i, $\tilde{X}_i$ and θ, we find that

$$\int_0^{\pi/2} \big|T_i(\alpha)\big|_\vee \tan\alpha \, \mathrm{d}\alpha \le B_i := \frac{c_3}{6\sigma_i^3} + \sqrt{2(1+\kappa)c_1 c_3}\left(\frac{\gamma^*(\mathscr{A} \mid \rho)}{\sigma_i} + \frac{4K(\beta_0)}{\sigma_i^3}\frac{\bar{\beta}}{\varepsilon}\right).$$

Now (2.24) reduces to

$$\big|\mathbb{E} f(W) - \mathcal{N}(0, \mathbf{I}_d)\{f\}\big| \le \sum_{i \in \mathcal{I}} B_i \mathbb{E}\big(|X_i||\tilde{X}_i|^2 + (1-\theta)|X_i|^3\big) \le \frac{3}{2} \sum_{i \in \mathcal{I}} B_i \mathbb{E}|X_i|^3, \qquad (2.25)$$

with the last inequality being due to Hölder's inequality.

Now fix $0 < \beta_* < 1$ (an explicit value will be chosen later) and assume first that $\bar{\beta} \le \beta_*$. By Jensen's inequality, $\mathbb{E}|X_i|^2 \le (\mathbb{E}|X_i|^3)^{2/3} \le \bar{\beta}^{2/3} \le \beta_*^{2/3}$ for all $i \in \mathcal{I}$. Next, for each unit vector $u \in \mathbb{R}^d$,

$$\langle \mathbf{\Sigma}_i u, u \rangle = u^T \mathbf{\Sigma}_i u = u^T \big(\mathbf{I}_d - \mathbb{E} X_i X_i^T\big) u = 1 - \mathbb{E}\langle X_i, u\rangle^2 \ge 1 - \mathbb{E}|X_i|^2 \ge 1 - \beta_*^{2/3}.$$

Therefore, $\sigma_i^2 \ge 1 - \beta_*^{2/3}$ for all $i \in \mathcal{I}$. In particular, the matrices $\mathbf{\Sigma}_i$ are non-singular and the quantities B_i can be uniformly bounded. Letting $\sigma_* := (1 - \beta_*^{2/3})^{1/2}$, (2.25) reduces to

$$\big|\mathbb{E} f(W) - \mathcal{N}(0, \mathbf{I}_d)\{f\}\big| \le \left[\frac{c_3}{4\sigma_*^3} + \sqrt{2(1+\kappa)c_1 c_3}\left(\frac{3\gamma^*(\mathscr{A} \mid \rho)}{2\sigma_*} + \frac{6K(\beta_0)}{\sigma_*^3}\frac{\bar{\beta}}{\varepsilon}\right)\right]\bar{\beta}. \qquad (2.26)$$

Recalling (2.23), we obtain

$$\big|\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\}\big| \le \left[\frac{c_3}{4\sigma_*^3} + \sqrt{2(1+\kappa)c_1 c_3}\left(\frac{3\gamma^*(\mathscr{A} \mid \rho)}{2\sigma_*} + \frac{6K(\beta_0)}{\sigma_*^3}\frac{\bar{\beta}}{\varepsilon}\right)\right]\bar{\beta} + \gamma^*(\mathscr{A} \mid \rho)\varepsilon.$$

Choosing $\varepsilon := 12\bar{\beta}\sqrt{2(1+\kappa)c_1c_3}/\sigma_*^3$, this reduces to

$$\begin{aligned}&\big|\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\}\big| \\ &\quad \le \left[\frac{K(\beta_0)}{2} + \frac{c_3}{4\sigma_*^3} + \gamma^*(\mathscr{A} \mid \rho)\sqrt{2(1+\kappa)c_1c_3}\left(\frac{3}{2\sigma_*} + \frac{12}{\sigma_*^3}\right)\right]\bar{\beta}. \end{aligned} \tag{2.27}$$

Now we are left with the case $\bar{\beta} \ge \beta_*$. We trivially estimate

$$\big|\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\}\big| \le 1 \le \frac{\bar{\beta}}{\beta_*}. \tag{2.28}$$

Dividing estimates (2.27) and (2.28) by $\bar{\beta}$, taking the supremum over all $A \in \mathscr{A}$ and all sums W, and plugging into (2.22), we obtain

$$K(\beta_0) \le \max\left\{\frac{1}{\beta_*}, \frac{K(\beta_0)}{2} + \frac{c_3}{4\sigma_*^3} + \gamma^*(\mathscr{A} \mid \rho)\sqrt{2(1+\kappa)c_1c_3}\left(\frac{3}{2\sigma_*} + \frac{12}{\sigma_*^3}\right)\right\}.$$

Since $K(\beta_0) \le 1/\beta_0 < \infty$, it follows that

$$K(\beta_0) \le \max\left\{\frac{1}{\beta_*}, \frac{c_3}{2\sigma_*^3} + \gamma^*(\mathscr{A} \mid \rho)\sqrt{2(1+\kappa)c_1c_3}\left(\frac{3}{\sigma_*} + \frac{24}{\sigma_*^3}\right)\right\}. \tag{2.29}$$

Choose $\beta_* := 1/27$, which is approximately optimal for the class of all half-lines on the real line. Straightforward numerical estimation yields $K(\beta_0) \le \max\{27, 1 + 50\gamma^*(\mathscr{A} \mid \rho)\sqrt{1+\kappa}\}$; this holds true for all $\beta_0 > 0$. Thus, for a fixed sum $W = \sum_{i \in \mathcal{I}} X_i$, one can plug the preceding estimate into (2.22), choosing $\beta_0 := \sum_{i \in \mathcal{I}} \mathbb{E}|X_i|^3$; (1.5) follows.

Now we turn to the case where $\mathscr{A}$ does not necessarily meet Assumption (A1′). This time, fix $\kappa \ge 0$ and for each $\beta_0, \gamma_0 > 0$, define

$$K(\beta_0, \gamma_0) := \sup \frac{|\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\}|}{\max\{\sum_{i \in \mathcal{I}} \mathbb{E}|X_i|^3, \beta_0\} \max\{\gamma^*(\mathscr{A} \mid \rho), \gamma_0\}}, \tag{2.30}$$

where the supremum runs over the family of all sums $W = \sum_{i \in \mathcal{I}} X_i$ of independent random vectors with $\mathbb{E}X_i = 0$ and $\mathrm{Var}(W) = \mathbf{I}_d$, all classes $\mathscr{A}$ which, along with the underlying functions ρ_A, satisfy Assumptions (A1)–(A8) (with the chosen κ), and all $A \in \mathscr{A}$.

Now fix $\beta_0, \gamma_0 > 0$, a sum $W = \sum_{i \in \mathcal{I}} X_i$ in the aforementioned family, a class $\mathscr{A}$ along with functions ρ_A satisfying Assumptions (A1)–(A8), and a set $A \in \mathscr{A}$. We proceed as in the previous case up to the estimation of $D_{i,A}$. For the latter, we now refer to (2.30), again with $\mathbf{\Sigma}_i^{-1/2} W_i$ in place of W. However, the set $\mathbf{\Sigma}_i^{-1/2}(A - \theta X_i)$ might not be in $\mathscr{A}$. Instead, it is in the class $\tilde{\mathscr{A}} := \{\mathbf{\Sigma}_i^{-1/2} A';\, A' \in \mathscr{A}\}$. Thus, we may take $\mathbf{\Sigma}_i^{-1/2}(A - \theta X_i)$ in place of A provided that we take $\tilde{\mathscr{A}}$ in place of $\mathscr{A}$. By Lemma 2.3, we may take the latter provided that we take the underlying family of functions $\tilde{\rho}_{\tilde{A}'}(x) := \rho_{\mathbf{\Sigma}_i^{1/2} \tilde{A}'}(\mathbf{\Sigma}_i^{1/2} x)$, $\tilde{A}' \in \tilde{\mathscr{A}}$, in place of the family $\rho_{A'}$, $A' \in \mathscr{A}$: in this case, κ stays the same. Denoting by σ_i^2 the smallest eigenvalue of $\mathbf{\Sigma}_i$ (with $\sigma_i > 0$), recall that

$\mathbb{E}|\mathbf{\Sigma}_i^{-1/2}X_j|^3 \le \sigma_i^{-3}\mathbb{E}|X_j|^3$ and observe that, again by Lemma 2.3, $\gamma^*(\tilde{\mathcal{A}} \mid \tilde{\rho}) \le \gamma^*(\mathcal{A} \mid \rho)/\sigma_i$ (notice that $\sigma_i \le 1$). By (2.30), we have

$$D_{i,A} \le K(\beta_0, \gamma_0) \max\left\{\frac{1}{\sigma_i^3} \sum_{j \in \mathcal{I}\setminus\{i\}} \mathbb{E}|X_j|^3, \beta_0\right\} \max\left\{\frac{\gamma^*(\mathcal{A} \mid \rho)}{\sigma_i}, \gamma_0\right\} \le \frac{K(\beta_0, \gamma_0)\bar{\beta}\bar{\gamma}}{\sigma_i^4},$$

where $\bar{\beta} := \max\{\sum_{j\in\mathcal{I}} \mathbb{E}|X_j|^3, \beta_0\}$ and $\bar{\gamma} := \max\{\gamma^*(\mathcal{A} \mid \rho), \gamma_0\}$. Applying Lemma 2.7 to the conditional distribution of W given X_i, $\tilde{X}_i$ and θ, we find that

$$\int_0^{\pi/2} \big|T_i(\alpha)\big|_\vee \tan\alpha \, \mathrm{d}\alpha \le B_i := \frac{c_3}{6\sigma_i^3} + \sqrt{2(1+\kappa)c_1c_3}\left(\frac{\bar{\gamma}}{\sigma_i^2} + \frac{4K(\beta_0, \gamma_0)}{\sigma_i^4}\frac{\bar{\beta}\bar{\gamma}}{\varepsilon}\right).$$

Again, fix $0 < \beta_* < 1$, let $\sigma_* := (1-\beta_*^{2/3})^{1/2}$ and assume first that $\bar{\beta} \le \beta_*$. By the same argument as in the first part, we derive

$$\big|\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\}\big| \le \left[\frac{c_3}{4\sigma_*^3} + \sqrt{2(1+\kappa)c_1c_3}\left(\frac{3\bar{\gamma}}{2\sigma_*^2} + \frac{6K(\beta_0, \gamma_0)}{\sigma_*^4}\frac{\bar{\beta}\bar{\gamma}}{\varepsilon}\right)\right]\bar{\beta} + \bar{\gamma}\varepsilon.$$

Choosing $\varepsilon := 12\bar{\beta}\sqrt{2(1+\kappa)c_1c_3}/\sigma_*^4$, this reduces to

$$\begin{aligned}
&\big|\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\}\big| \\
&\quad \le \frac{c_3\bar{\beta}}{4\sigma_*^3} + \left[\frac{K(\beta_0, \gamma_0)}{2} + \sqrt{2(1+\kappa)c_1c_3}\left(\frac{3}{2\sigma_*^2} + \frac{12}{\sigma_*^4}\right)\right]\bar{\beta}\bar{\gamma} \\
&\quad \le \left[\frac{K(\beta_0, \gamma_0)}{2} + \frac{c_3}{4\gamma_0\sigma_*^3} + \sqrt{2(1+\kappa)c_1c_3}\left(\frac{3}{2\sigma_*^2} + \frac{12}{\sigma_*^4}\right)\right]\bar{\beta}\bar{\gamma}. \qquad (2.31)
\end{aligned}$$

In the case $\bar{\beta} \ge \beta_*$, we trivially estimate

$$\big|\mathbb{P}(W \in A) - \mathcal{N}(0, \mathbf{I}_d)\{A\}\big| \le 1 \le \frac{\bar{\beta}\bar{\gamma}}{\beta_*\gamma_0}. \qquad (2.32)$$

Divide the estimates (2.27) and (2.28) by $\bar{\beta}\bar{\gamma}$ and take the supremum over all $A \in \mathcal{A}$, all sums W, and all families $\mathcal{A}$ (along with functions ρ_A). Plugging into (2.30), we obtain

$$K(\beta_0, \gamma_0) \le \max\left\{\frac{1}{\beta_*\gamma_0}, \frac{K(\beta_0, \gamma_0)}{2} + \frac{c_3}{4\sigma_*^3\gamma_0} + \sqrt{2(1+\kappa)c_1c_3}\left(\frac{3}{2\sigma_*^2} + \frac{12}{\sigma_*^4}\right)\right\}.$$

Since $K(\beta_0, \gamma_0) \le 1/(\beta_0\gamma_0) < \infty$, it follows that

$$K(\beta_0, \gamma_0) \le \max\left\{\frac{1}{\beta_*\gamma_0}, \frac{c_3}{2\sigma_*^3\gamma_0} + \sqrt{2(1+\kappa)c_1c_3}\left(\frac{3}{\sigma_*^2} + \frac{24}{\sigma_*^4}\right)\right\}. \qquad (2.33)$$

As in the first case, choose $\beta_* := 1/27$. Straightforward numerical estimation yields $K(\beta_0, \gamma_0) \le \max\{27/\gamma_0, 1/\gamma_0 + 53\sqrt{1+\kappa}\}$; this holds true for all $\beta_0, \gamma_0 > 0$. Thus, for a fixed sum $W =$

$\sum_{i\in\mathcal{I}} X_i$ and a fixed class $\mathcal{A}$ along with functions ρ_A, one can plug the preceding estimate into (2.30), choosing $\beta_0 := \sum_{i\in\mathcal{I}} \mathbb{E}|X_i|^3$ and $\gamma_0 := \gamma^*(\mathcal{A} \mid \rho)$; (1.4) follows. This completes the proof. □

3. Derivation of the bound on the Gaussian perimeter of convex sets

In this section, we prove Theorem 1.2, and also state and prove Proposition 3.1, which is a generalization of Proposition 1.1. Throughout this section, fix $d \in \mathbb{N}$ and denote by $\mathcal{C}_d$ the class of all measurable convex sets in $\mathbb{R}^d$. From Section 1, recall the definitions of δ_A and A^t for a set $A \subseteq \mathbb{R}^d$. Recall also that $\mathcal{H}^r$ denotes the r-dimensional Hausdorff measure.

The first result of the section is closely related to Lemma 11 of Livshyts [17].

Proposition 3.1. *Let $\mathcal{A}$ be a class of certain convex sets in $\mathbb{R}^d$. Suppose that $A^t \in \mathcal{A} \cup \{\varnothing\}$ for all $A \in \mathcal{A}$ and all $t \in \mathbb{R}$. Take a continuous function $f : \mathbb{R}^d \to [0,\infty)$, which is integrable with respect to the Lebesgue measure. Then we have $\gamma_f(\mathcal{A}) = \gamma_f^*(\mathcal{A})$, where*

$$\gamma_f(\mathcal{A}) = \sup\left\{\int_{\partial A} f(x)\mathcal{H}^{d-1}(\mathrm{d}x);\, A \in \mathcal{A}\right\},$$

$$\gamma_f^*(\mathcal{A}) = \sup\left\{\frac{1}{\varepsilon}\int_{A^\varepsilon\setminus A} f(x)\,\mathrm{d}x,\, \frac{1}{\varepsilon}\int_{A\setminus A^{-\varepsilon}} f(x)\,\mathrm{d}x;\, \varepsilon > 0,\, A \in \mathcal{A}\right\}.$$

Before proving the preceding assertion, we need to introduce some notation and auxiliary results. For a map $g\colon A \to \mathbb{R}^n$, where $A \subseteq \mathbb{R}^d$ is a measurable set, and for a point $x \in A$ where g is differentiable, denote by $\mathbf{D}g(x)$ its derivative (i.e., Jacobian matrix) at x. For each $r = 0, 1, 2, \dots$, define $J_r g(x)$, the r-dimensional absolute Jacobian, as follows: if $\operatorname{rank}\mathbf{D}g(x) < r$, set $J_r g(x) := 0$. If $\operatorname{rank}\mathbf{D}g(x) > r$, set $J_r g(x) := \infty$. Finally, if $\operatorname{rank}\mathbf{D}g(x) = r$, define $J_r g(x)$ to be the product of r non-zero singular values in the singular-value decomposition of $\mathbf{D}g(x)$, that is, $\mathbf{D}g(x) = \mathbf{U\Sigma V}$, where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices and where $\mathbf{\Sigma}$ is a diagonal rectangular matrix with non-negative diagonal elements referred to as singular values. It is easy to see that the definition is independent of the decomposition. Notice that for $n = 1$, we have $J_1 g(x) = |\nabla g(x)|$.

The main tool used in the proof of Proposition 3.1 will be the following assertion, which can be regarded as a curvilinear variant of Fubini's theorem. As a special case, it also includes the change of variables formula in the multi-dimensional integral.

Proposition 3.2 (Federer [14], Corollary 3.2.32). *Let $A \subseteq \mathbb{R}^d$ be a measurable set, $f\colon A \to \mathbb{R}$ a measurable function and $g\colon A \to \mathbb{R}^n$ a locally Lipschitzian map. Take $0 \le r \le d$ and assume that $f J_r g$ is integrable with respect to the Lebesgue measure. Then $f|_{g^{-1}(\{y\})}$ is $\mathcal{H}^{d-r}$-integrable for almost all y with respect to $\mathcal{H}^r$, the function $y \mapsto \int_{g^{-1}(\{y\})} f\,\mathrm{d}\mathcal{H}^{d-r}$ is measurable and*

$$\int_{\{x\in A;\, \mathcal{H}^{d-r}(g^{-1}(\{g(x)\}))>0\}} f(x) J_r g(x)\,\mathrm{d}x = \int_{\mathbb{R}^n}\int_{g^{-1}(\{y\})} f\,\mathrm{d}\mathcal{H}^{d-r}\,\mathcal{H}^r(\mathrm{d}y).$$

Remark 3.1. The integrand in the left-hand side is defined for almost all $x \in A$, because g is almost everywhere differentiable by Rademacher's theorem.

Corollary 3.1 (Coarea Formula). *Let d, n, A, f and g be as in the preceding statement. Suppose that $d \geq n$. Then we have*

$$\int_A f(x) J_n g(x)\, \mathrm{d}x = \int_{\mathbb{R}^n} \int_{g^{-1}(\{y\})} f\, \mathrm{d}\mathcal{H}^{d-n}\, \mathrm{d}y.$$

Proof. Apply Proposition 3.2 with $d = n$ and observe that by the implicit function theorem, $J_n g(x) > 0$ implies $\mathcal{H}^{d-n}(g^{-1}(\{g(x)\})) > 0$. □

Now we turn to some simple properties of convex sets. First, one can easily check that if C is a non-empty convex set and $x \in \mathbb{R}^d$, there exists a unique point in $\overline{C}$ which is closest to x.

Definition 3.1. The *orthogonal projection* to a non-empty convex set C is a map $p_C^\perp : \mathbb{R}^d \to \overline{C}$, where $p_C^\perp(x)$ is defined to be the unique point in $\overline{C}$ which is closest to x.

Proposition 3.3. *Let C be a convex set, which is neither the empty set nor the whole $\mathbb{R}^d$.*

(1) *For each $x \in \mathbb{R}^d$ and each $\varepsilon > 0$, there exists $y \in \mathbb{R}^d$ with $0 < \delta_C(y) - \delta_C(x) = |y - x| < \varepsilon$.*
(2) *δ_C is almost everywhere differentiable.*
(3) *For each x where δ_C is differentiable, we have $|\nabla \delta_C(x)| = 1$.*
(4) *For each $t \in \mathbb{R}$, we have $\partial C^t = \{x; \delta_C(x) = t\}$.*

Proof. If $x \in \operatorname{Int} C$, there exists a point $z \in \mathbb{R}^d \setminus \operatorname{Int} C$ which is closest to x. For all $y = (1 - \tau)x + \tau z$, where $0 \leq \tau \leq 1$, we have $\operatorname{dist}(y, \mathbb{R}^d \setminus C) = \operatorname{dist}(x, \mathbb{R}^d \setminus C) - |x - y|$, that is, $-\delta_C(y) = -\delta_C(x) - |x - y|$. Next, if $x \in \mathbb{R}^d \setminus \overline{C}$, take $\tau \geq 0$ and let $y = (1 + \tau)x - \tau p_C^\perp(x)$. By convexity, we have $\langle w - p_C^\perp(x), x - p_C^\perp(x)\rangle \leq 0$ for all $w \in C$. As a result, $\operatorname{dist}(y, C) = \operatorname{dist}(x, C) + |x - y|$ for all $\tau \geq 0$. Finally, if $x \in \partial C$, it is well known that there exist a unit outer normal vector u (possibly more than one); then, for all $y = x + \tau u$, where $\tau \geq 0$, we again have $\operatorname{dist}(y, C) = \operatorname{dist}(x, C) + |x - y|$. This proves (1).

One can easily check that δ_C is non-expansive. By Rademacher's theorem (see also Remark 2.1), it is almost everywhere differentiable and $|\nabla \delta_C(x)| \leq 1$ for all x where it is differentiable. This proves (2). However, by (1), we have $|\nabla \delta_C(x)| \geq 1$. This proves (3).

From the continuity of δ_C, it follows that $\partial C^t \subseteq \{x; \delta_C(x) = t\}$. The opposite follows from (1). This proves (4). □

Proof of Proposition 3.1. Without loss of generality, we may assume that $\varnothing$ and $\mathbb{R}^d$ are not elements of $\mathcal{A}$. Take $A \in \mathcal{A}$. By the Coarea formula, we have

$$\int_{A^\varepsilon \setminus A} f(x) J_1 \delta_A(x)\, \mathrm{d}x = \int_0^\varepsilon \int_{\delta_A^{-1}(\{t\})} f(x) \mathcal{H}^{d-1}(\mathrm{d}x)\, \mathrm{d}t.$$

Applying Parts (3) and (4) of Proposition 3.3, this reduces to

$$\int_{A^\varepsilon\setminus A} f(x)\,\mathrm{d}x = \int_0^\varepsilon \int_{\partial A^t} f(x)\mathcal{H}^{d-1}(\mathrm{d}x)\,\mathrm{d}t \le \varepsilon\gamma_f(\mathcal{A}).$$

Similarly, we obtain

$$\int_{A\setminus A^{-\varepsilon}} f(x)\,\mathrm{d}x = \int_{-\varepsilon}^0 \int_{\partial A^t} f(x)\mathcal{H}^{d-1}(\mathrm{d}x)\,\mathrm{d}t \le \varepsilon\gamma_f(\mathcal{A})$$

(remember that $A^t \in \mathcal{A}\cup\{\varnothing\}$; for $A=\varnothing$, the inner integral vanishes). Dividing by ε, and taking the supremum over ε and A, we obtain $\gamma_f^*(\mathcal{A}) \le \gamma_f(\mathcal{A})$.

To prove the opposite inequality, observe first that, by Parts (2) and (3) of Proposition 3.3, $p_A^\perp$ is non-expansive. Next, observe that $p_A^\perp((1+\tau)x - \tau p_A^\perp(x)) = p_A^\perp(x)$ for all $x \in \mathbb{R}^d \setminus \overline{A}$ and all $\tau \ge 0$. Therefore, if $p_A^\perp$ is differentiable at $x \in \mathbb{R}^d \setminus \overline{A}$, we have $\operatorname{rank}\mathbf{D}p_A^\perp(x) \le d-1$ and, moreover, $J_{d-1}p_A^\perp(x) \le 1$. By Proposition 3.2, we have

$$\begin{aligned}\int_{A^\varepsilon\setminus A} f(x)\,\mathrm{d}x &\ge \int_{\{x\in A^\varepsilon\setminus\overline{A};\,\mathcal{H}^1((p_A^\perp)^{-1}(\{p_A^\perp(x)\}))>0\}} f(x)J_{d-1}p_A^\perp(x)\,\mathrm{d}x \\ &= \int_{\partial A}\int_{(p_A^\perp)^{-1}(\{y\}\cap(A^\varepsilon\setminus\overline{A}))} f\,\mathrm{d}\mathcal{H}^1\mathcal{H}^{d-1}(\mathrm{d}y).\end{aligned}$$

If u is a unit outer normal vector at $y \in \partial C$, then $p_A^\perp(y+\tau u) = y$ for all $\tau \ge 0$. Moreover, $y+\tau u \in (p_A^\perp)^{-1}(\{y\})\cap(A^\varepsilon\setminus\overline{A})$ for all $0<\tau\le\varepsilon$. Therefore, $\mathcal{H}^1((p_A^\perp)^{-1}(\{y\})\cap(A^\varepsilon\setminus\overline{A})) \ge \varepsilon$. As a result,

$$\int_{A^\varepsilon\setminus A} f(x)\,\mathrm{d}x \ge \varepsilon\int_{\partial A} f^{-\varepsilon}(y)\mathcal{H}^{d-1}(\mathrm{d}y),$$

where $f^{-\varepsilon}(x) := \inf_{|v|\le\varepsilon} f(x+v)$. Dividing by ε, we obtain

$$\int_{\partial A} f^{-\varepsilon}(y)\mathcal{H}^{d-1}(\mathrm{d}y) \le \gamma_f^*(\mathcal{A}).$$

Since f is continuous, we have $\lim_{\varepsilon\downarrow 0} f^{-\varepsilon}(x) = f(x)$ for all $x\in\mathbb{R}^d$. Applying the dominated convergence theorem and taking the supremum over all A, we obtain $\gamma_f(\mathcal{A}) \le \gamma_f^*(\mathcal{A})$. This completes the proof. □

The orthogonal projection will be one of two key maps used in the proof of Theorem 1.2. The other one will be the *radial projection*.

Definition 3.2. Let C be a convex set with $0 \in \operatorname{Int} C$. We define the *radial function* of C to be the map $\rho_C : \mathbb{R}^d\setminus\{0\} \to (0,\infty]$ defined by

$$\rho_C(x) := \sup\left\{r>0;\, r\frac{x}{|x|}\in C\right\} = \inf\left\{r>0;\, r\frac{x}{|x|}\notin C\right\}$$

and the *radial projection* of C to be the map $p_C^\rho\colon \{x \in \mathbb{R}^d \setminus \{0\}; \rho_C(x) < \infty\} \to \partial C$ defined by $p_C^\rho(x) := \rho_C(x)\dfrac{x}{|x|}$.

Lemma 3.1. *Let C be as before. Define the set $D := \{x \in \mathbb{R}^d \setminus \{0\}; \rho_C(x) < \infty\}$. Then:*

(1) *D is open and ρ_C and p_C^ρ are locally Lipschitzian on D.*
(2) *If ρ_C is differentiable at x, so is p_C^ρ, there is a unique outer unit normal vector at $p_C^\rho(x)$ and we have*

$$J_{d-1}p_C^\rho(x) = \left(\frac{\rho_C(x)}{|x|}\right)^{d-1}\frac{1}{\cos\theta},$$

where θ is the angle between x and the outer unit normal vector at $p_C^\rho(x)$.

Proof. Since $0 \in \operatorname{Int} C$, there exists $r_0 > 0$, such that $\{y \in \mathbb{R}^d; |y| < r_0\} \subseteq C$. Fix $x \in \mathbb{R}^d \setminus \{0\}$. Let $r_1 := \rho_C(x)$ and $v := x/|x|$. Take $w \perp v$ and $s, t \in \mathbb{R}$, and let $z := sv + tw$. By convexity, $z \in C$ if $0 \le s < r_1(1 - \frac{|t|}{r_0})$, and $z \notin C$ if $s > r_1(1 + \frac{|t|}{r_0})$. Consequently,

$$\frac{|z|}{\frac{s}{r_1} + \frac{|t|}{r_0}} \le \rho_C(z) \le \frac{|z|}{\frac{s}{r_1} - \frac{|t|}{r_0}},$$

provided that $s > 0$ and $|t| < sr_0/r_1$. Letting $s = |x| + \sigma$ and $t = \tau$, we obtain

$$\frac{\sqrt{(1 + \frac{\sigma}{|x|})^2 + (\frac{\tau}{|x|})^2}}{1 + \frac{\sigma}{|x|} + \frac{\rho_C(x)}{r_0}\frac{|\tau|}{|x|}} \le \frac{\rho_C(x + \sigma v + \tau w)}{\rho_C(x)} \le \frac{\sqrt{(1 + \frac{\sigma}{|x|})^2 + (\frac{\tau}{|x|})^2}}{1 + \frac{\sigma}{|x|} - \frac{\rho_C(x)}{r_0}\frac{|\tau|}{|x|}},$$

provided that $\sigma > -|x|$ and $|\tau| < (|x| + \sigma)r_0/\rho_C(x)$. From the preceding inequality, we deduce first that D is open, then that ρ_C is continuous on D, then that ρ_C is locally Lipschitzian on D and finally that the latter also holds for p_C^ρ. This proves (1).

Now suppose that ρ_C is differentiable at x. By the chain rule, so is p_C^ρ and straightforward computation yields

$$\mathbf{D}p_C^\rho(x)v = \langle\nabla\rho_C(x), v\rangle\frac{x}{|x|} + \rho_C(x)\left(\frac{v}{|x|} - \frac{\langle x, v\rangle x}{|x|^3}\right). \tag{3.1}$$

Observe that since $p_C^\rho(kx) = p_C^\rho(x)$ for all $k > 0$, we have, by the chain rule, $\mathbf{D}p_C^\rho(kx) = \frac{1}{k}\mathbf{D}p_C^\rho(x)$. Thus, letting $y := p_C^\rho(x) = \frac{\rho_C(x)}{|x|}x$, we have $\mathbf{D}p_C^\rho(x) = \frac{\rho_C(x)}{|x|}\mathbf{D}p_C^\rho(y)$. Taking y in place of x in (3.1) and noting that $\rho_C(y) = |y|$, we obtain

$$\mathbf{D}p_C^\rho(y)v = v - \langle y - |y|\nabla\rho_C(y), v\rangle\frac{y}{|y|^2}.$$

Differentiating $\rho_C(ky) = \rho_C(y)$ with respect to k, we obtain $\langle\nabla\rho_C(y), y\rangle = 0$. Making use of this identity, we find after some calculation that $\mathbf{D}p_C^\rho(y)$ is a projector.

If u is a unit outer normal vector at y, then u is perpendicular to the image of $\mathbf{D}p_C^\rho(y)$. However, since $\mathbf{D}p_C^\rho(y)$ is a projector, its image is the same as the set of its fixed points, which are precisely the vectors perpendicular to $y - |y|\nabla\rho_C(y)$. Therefore, u must be parallel to $y - |y|\nabla\rho_C(y)$. Since $\langle u, y\rangle > 0$ and since $\langle\nabla\rho_C(y), y\rangle = 0$, we have $u = \frac{y-|y|\nabla\rho_C(y)}{|y|\sqrt{1+|\nabla\rho_C(y)|^2}}$. Thus, there is indeed a unique unit outer normal vector. Taking the inner product with y, we find that $|\nabla\rho_C(y)| = \tan\theta$.

Without loss of generality, we may assume that $y/|y|$ is the first base vector and that $\nabla\rho_C(y)/|\nabla\rho_C(y)|$ is the second one, the latter provided that $\nabla\rho_C(y) \neq 0$. This way, we have

$$\mathbf{D}p_C^\rho(y) = \begin{bmatrix} 0 & \tan\theta & \\ 0 & 1 & \\ & & \mathbf{I}_{d-2} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & \\ -\sin\theta & \cos\theta & \\ & & \mathbf{I}_{d-2} \end{bmatrix} \begin{bmatrix} 0 & 0 & \\ 0 & 1/\cos\theta & \\ & & \mathbf{I}_{d-2} \end{bmatrix} \mathbf{I}_d.$$

The latter singular-value decomposition yields $J_{d-1}p_C^\rho(y) = 1/\cos\theta$. Recalling $\mathbf{D}p_C^\rho(x) = \frac{\rho_C(x)}{|x|}\mathbf{D}p_C^\rho(y)$, we obtain (2). □

Before finally turning to the proof of Theorem 1.2, we still need some inequalities regarding elementary and special functions. The first one regards the *Mills ratio*:

$$R(x) := e^{x^2/2}\int_x^\infty e^{-z^2/2}\,\mathrm{d}z = \int_0^\infty e^{-tx-t^2/2}\,\mathrm{d}t. \tag{3.2}$$

For $y > 0$, define

$$I(y) := \inf_{x\geq 0}\bigl(xy + R(x)\bigr) \tag{3.3}$$

and observe that $I(y) > 0$ and that I is strictly increasing.

Lemma 3.2. *For all* $0 < y < 1$, *the function* I *satisfies* $I(y) \geq 2\sqrt{y(1-y)}$.

Proof. By Formula 7.1.13 of Abramowitz and Stegun [1], we have $R(x) \geq \frac{2}{x+\sqrt{x^2+4}}$ for all $x \geq 0$. A straightforward calculation shows that the expression $\inf_{x\geq 0}(xy + \frac{2}{x+\sqrt{x^2+4}})$ equals $2\sqrt{y(1-y)}$ for $y \leq 1/2$ and 1 for $y \geq 1/2$. □

Lemma 3.3. *For all* $0 \leq x < \alpha$, *we have*

$$\left(1 - \frac{x}{\alpha}\right)^{-\alpha^2} e^{-\alpha x} \geq e^{x^2/2}, \tag{3.4}$$

$$\left(1 - \frac{x}{\alpha}\right)^{\alpha^2-1} e^{\alpha x} \geq e^{-x^2/2}\left(1 - \frac{x^3}{\alpha}\right). \tag{3.5}$$

Lemma 3.4. *Consider the function*

$$G(x,\alpha,\beta) := \left(1-\frac{x}{\alpha}\right)^{-\alpha^2} e^{-\alpha x}\left[\beta + \int_x^\alpha \left(1-\frac{y}{\alpha}\right)^{\alpha^2-1} e^{\alpha y}\,\mathrm{d}y\right].$$

For all $\alpha \ge 1$ and $\beta \ge 1/\sqrt{e}$, this function satisfies

$$\inf_{x<\alpha} G(x,\alpha,\beta) = \inf_{0\le x\le 1} G(x,\alpha,\beta).$$

The proofs of Lemmas 3.3 and 3.4 are deferred to the supplementary material [20].

Proof of Theorem 1.2. We basically follow Nazarov's [19] argument, tackling certain technical matters differently and expanding some arguments. First, observe that if a convex set C has no interior, then it is contained in the boundary of some half-space H, so that $\gamma(C) \le \gamma(H)$. Therefore, in the supremum in the definition of γ_d, it suffices to consider sets with non-empty interior. Next, if $0 \notin C$, we have $\gamma(C) \le \gamma(C - p_C^\perp(0))$ (for details, see Section 4 of Livshyts [17]). Therefore, it suffices only to consider sets C with the origin in the closure and with non-empty interior. Moreover, by continuity, it suffices to take sets containing the origin in the interior.

Let C be a convex set with $0 \in \operatorname{Int} C$. Take a random locally Lipschitzian map $G\colon \mathbb{R}^d \setminus \{0\} \to \partial C$ with $J_{d-1}G(x) \le J(x)$ for almost all $x \in A$, where $J : \mathbb{R}^d \setminus \{0\} \to (0,\infty)$ is another random function (random maps should be measurable as maps from the product of $\mathbb{R}^d \setminus \{0\}$ and the probability space with respect to the product of the Borel σ-algebra and the σ-algebra of the probability space). The random choices of G and J will depend on a parameter $p \in (0,1]$ (see below). By Proposition 3.2, we have

$$1 = \int_{\mathbb{R}^d} \phi_d(x)\,\mathrm{d}x \ge \mathbb{E}_p\left[\int_{\mathbb{R}^d\setminus\{0\}} \phi_d(x)\frac{J_{d-1}G(x)}{J(x)}\,\mathrm{d}x\right] \ge \int_{\partial C} \mathbb{E}_p\left[\int_{G^{-1}(\{y\})} \frac{\phi_d}{J}\,\mathrm{d}\mathcal{H}^1\right]\mathcal{H}^{d-1}(\mathrm{d}y).$$

Thus,

$$\gamma(C) \le \inf_{0<p\le 1} \frac{1}{\inf_{y\in\partial C} \xi_C(y,p)}, \tag{3.6}$$

where

$$\xi_C(y,p) := \frac{1}{\phi_d(y)}\mathbb{E}_p\left[\int_{G^{-1}(\{y\})} \frac{\phi_d}{J}\,\mathrm{d}\mathcal{H}^1\right].$$

Now define G as follows: for $x \in \overline{C}$, let $G(x) := p_C^\rho(x)$; for $x \in \mathbb{R}^d \setminus \overline{C}$, let $G(x) := p_C^\rho(x)$ with probability $1-p$ and $G(x) := p_C^\perp(x)$ with probability p. To define $J(x)$, recall Lemma 3.1 along with the fact that $p_C^\perp$ is non-expansive. Thus, we may take $J(x) := 1$ where $G = p_C^\perp$ and $J(x) := (\frac{\rho_C(x)}{|x|})^{d-1}\frac{1}{\cos\theta(p_C^\rho(x))}$ where $G = p_C^\rho$; here, $\theta(y)$ denotes the maximal angle between y and the outer normal of C at y. Notice that the maximum is attained because the set of all unit outer normal vectors is compact, and is strictly less than $\pi/2$ because $0 \in \operatorname{Int} C$; typically, the outer normal vector is unique by Lemma 3.1. As a result, we have $\xi_C(y,p) \geq \xi_{1,C}(y,p) + \xi_{2,C}(y,p)$, where

$$\xi_{1,C}(y,p) := \frac{\cos\theta(y)}{|y|^{d-1}\phi_d(y)}\bigg[\int_{(p_C^\rho)^{-1}(\{y\})\cap\overline{C}} |x|^{d-1}\phi_d(x)\mathcal{H}^1(\mathrm{d}x) + (1-p)\int_{(p_C^\rho)^{-1}(\{y\})\setminus\overline{C}} |x|^{d-1}\phi_d(x)\mathcal{H}^1(\mathrm{d}x)\bigg],$$

$$\xi_{2,C}(y,p) := \frac{p}{\phi_d(y)}\int_{(p_C^\perp)^{-1}(\{y\})} \phi_d(x)\mathcal{H}^1(\mathrm{d}x).$$

Observe that $\xi_{1,C}(y,p) = \cos\theta(y)\xi_1(|y|,d,p)$, where

$$\xi_1(r,d,p) := \frac{e^{r^2/2}}{r^{d-1}}\bigg[\int_0^r t^{d-1}e^{-t^2/2}\,\mathrm{d}t + (1-p)\int_r^\infty t^{d-1}e^{-t^2/2}\,\mathrm{d}t\bigg] = \frac{e^{r^2/2}}{r^{d-1}}\bigg[2^{d/2-1}(1-p)\Gamma\left(\frac{d}{2}\right) + p\int_0^r t^{d-1}e^{-t^2/2}\,\mathrm{d}t\bigg]. \tag{3.7}$$

As for $\xi_{2,C}(y,p)$, observe that $(p_C^\perp)^{-1}(\{y\}) \supseteq \{y+su; s>0\}$, where u is a unit outer normal vector at y. Take u with the maximal angle between u and y. As a result, we have

$$\xi_{2,C}(y,p) \geq \frac{p}{\phi_d(y)}\int_0^\infty \phi_d(y+tu)\,\mathrm{d}t = p\int_0^\infty e^{-t\langle y,u\rangle - t^2/2}\,\mathrm{d}t = pR\big(|y|\cos\theta(y)\big),$$

recalling the Mills ratio defined in (3.2). Combining all estimates after (3.6), plugging into the latter and taking the supremum over all convex sets with the origin in the interior, we find that

$$\gamma_d \leq \bar{\gamma}_d := \inf_{0<p\leq 1}\bar{\gamma}_{d,p}, \tag{3.8}$$

where

$$\bar{\gamma}_{d,p} := \frac{1}{\inf_{r,c>0}(c\xi_1(r,d,p) + pR(cr))}.$$

Substituting $cr = b$ and recalling that the function I defined in (3.3) is strictly increasing, we find the following alternative expression of $\bar{\gamma}_{d,p}$:

$$\bar{\gamma}_{d,p} = \frac{1}{\inf_{r,b>0}(\frac{b}{r}\xi_1(r,d,p) + pR(b))} = \frac{1}{pI(\inf_{r>0}\frac{1}{pr}\xi_1(r,d,p))}. \tag{3.9}$$

For each d, $\bar{\gamma}_d$ can be evaluated numerically. Some values are given in Table 1.

Remark 3.2. For $d=1$, we obtain the actual maximal Gaussian perimeter: we have $\gamma_1 = \bar{\gamma}_1 = \bar{\gamma}_{1,1}$. First, observe that $\inf_{r>0}\frac{1}{r}\xi_1(r,1,1) = \inf_{r>0}\frac{e^{r^2/2}}{r}\int_0^r e^{-t^2/2}\,\mathrm{d}t = 1$. Differentiating (3.2),

Table 1. Upper bounds on the Gaussian perimeter for some dimensions (with all values rounded upwards)

d	$\bar{\gamma}_d$	$\bar{\gamma}_d/d^{1/4}$	d	$\bar{\gamma}_d$	$\bar{\gamma}_d/d^{1/4}$
1	0.798	0.798	9	1.154	0.666
2	0.864	0.726	10	1.179	0.663
3	0.929	0.706	20	1.364	0.645
4	0.981	0.694	50	1.666	0.627
5	1.025	0.685	100	1.949	0.617
6	1.063	0.679	200	2.288	0.609
7	1.096	0.674	500	2.842	0.601
8	1.126	0.670	1000	3.357	0.597

we find that $R'(x) = -\int_0^\infty te^{-tx-t^2/2}\,\mathrm{d}t$ and $R''(x) = \int_0^\infty t^2e^{-tx-t^2/2}\,\mathrm{d}t$. Since $R'(0) = -1$ and $R''(x) > 0$ for all x, we have $R'(x) \geq -1$ for all $x \geq 0$. Therefore, for $y = 1$, the infimum in (3.3) is attained at $x = 0$, so that $\bar{\gamma}_{1,1} = 1/I(1) = 1/R(0) = \sqrt{2/\pi}$.

Now we continue with the estimation. From Stirling's formula with remainder (e.g., Formula 6.1.38 of Abramowitz and Stegun [1]), one can easily deduce that $\Gamma(x) \geq \sqrt{\frac{2\pi}{x}}(\frac{x}{e})^x$ for all $x > 0$. Plugging into (3.7), we obtain

$$\frac{1}{pr}\xi_1(r,d,p) \geq \frac{e^{r^2/2}}{r^d}\left(\frac{1-p}{p}\sqrt{\frac{\pi}{d}}\left(\frac{d}{e}\right)^{d/2} + \int_0^r t^{d-1}e^{-t^2/2}\,\mathrm{d}t\right).$$

Substituting $\alpha := \sqrt{d/2}$, $x = \alpha - r^2/(2\alpha)$, $y = \alpha - t^2/(2\alpha)$, we obtain after some calculation

$$\inf_{r>0}\frac{1}{pr}\xi_1(r,d,p) \geq \frac{1}{2\alpha}\inf_{x<\alpha}\left(1-\frac{x}{\alpha}\right)^{-\alpha^2}e^{-\alpha x}\left[\frac{1-p}{p}\sqrt{2\pi} + \int_x^\alpha\left(1-\frac{y}{\alpha}\right)^{\alpha^2-1}e^{\alpha y}\,\mathrm{d}y\right].$$

Now suppose that $\alpha \geq 1$ and $\frac{1-p}{p}\sqrt{2\pi} \geq \frac{1}{\sqrt{e}}$; this is ensured if $d \geq 2$ and $p < 0.8$. In this case, we can apply Lemma 3.4 to reduce the infimum over $x < \alpha$ to the infimum over $[0,1]$. By Lemma 3.3, we can further estimate

$$\inf_{r>0}\frac{1}{pr}\xi_1(r,d,p) \geq \frac{1}{2\alpha}\inf_{0\leq x\leq 1}e^{x^2/2}\left[\frac{1-p}{p}\sqrt{2\pi} + \int_x^\alpha e^{-y^2/2}\left(1-\frac{y^3}{\alpha}\right)\mathrm{d}y\right].$$

Since $\alpha \geq 1$, $y \geq \alpha$ implies $y^3 \geq \alpha$, so that the upper limit α can be replaced with the infinity:

$$\begin{aligned}\inf_{r>0}\frac{1}{pr}\xi_1(r,d,p) &\geq \frac{1}{2\alpha}\inf_{0\leq x\leq 1}e^{x^2/2}\left[\frac{1-p}{p}\sqrt{2\pi} + \int_x^\infty e^{-y^2/2}\left(1-\frac{y^3}{\alpha}\right)\mathrm{d}y\right] \\ &= \frac{1}{2\alpha}\inf_{0\leq x\leq 1}\left[\frac{1-p}{p}\sqrt{2\pi}e^{x^2/2} + R(x) - \frac{x^2+2}{\alpha}\right]\end{aligned}$$

$$\geq \frac{1}{2\alpha} K(p) - \frac{3}{2\alpha^2}$$
$$= \frac{1}{\sqrt{2d}} K(p) - \frac{3}{d},$$

where $K(p) := \inf_{0\leq x\leq 1}[\frac{1-p}{p}\sqrt{2\pi}e^{x^2/2} + R(x)]$. Plugging into (3.9) and applying Lemma 3.2, we find that

$$\bar{\gamma}_{d,p} \leq \frac{1}{2p\sqrt{(\frac{1}{\sqrt{2d}}K(p) - \frac{3}{d})(1 - \frac{1}{\sqrt{2d}}K(p) + \frac{3}{d})}},$$

provided that $d \geq 2$, $p \leq 0.8$ and $d > \max\{\frac{(K(p))^2}{2}, \frac{18}{(K(p))^2}\}$. As $d \to \infty$, the preceding upper bound asymptotically equals $\frac{d^{1/4}}{2^{3/4}p\sqrt{K(p)}}$. Now choose p so that this asymptotic bound is optimal, that is, so that $p^2K(p)$ is maximal. Numerical calculation shows that this occurs approximately at $p = p^* := 0.72$ (which is less than 0.8). Moreover, one can numerically check that $K(p^*) > K^* := 1.98$. This indicates that the coefficient at $d^{1/4}$ in the bound on γ_d can be set to $\frac{1}{2^{3/4}p^*\sqrt{K^*}} < 0.59$.

Choosing $p = p^*$, we re-estimate $\bar{\gamma}_d$, using Lemma 3.2 once again:

$$\bar{\gamma}_d \leq \bar{\gamma}_{d,p^*} \leq \frac{1}{p^* I(\frac{K^*}{\sqrt{2d}} - \frac{3}{d})}$$
$$\leq \frac{1}{2p^*\sqrt{(\frac{K^*}{\sqrt{2d}} - \frac{3}{d})(1 - \frac{K^*}{\sqrt{2d}} + \frac{3}{d})}} =$$
$$\leq \frac{0.59d^{1/4}}{\sqrt{1 - (\frac{3\sqrt{2}}{K^*} + \frac{K^*}{\sqrt{2}})d^{-1/2}}}$$
$$\leq \frac{0.59d^{1/4}}{\sqrt{1 - 3.55d^{-1/2}}},$$

provided that $d \geq 3.55^2$, that is, $d \geq 13$. Taking $d = 932$, observe that $(1 - 3.55d^{-1/2})^{-1/2} \leq 1 + (\frac{1}{0.59}\sqrt{\frac{2}{\pi}} - 1)d^{-1/4}$; this inequality also holds in the limit as $d \to \infty$. Since the function $x \mapsto (1 - 3.55x^2)^{-1/2}$ is convex, the latter inequality must hold for all $d \geq 932$. This completes the proof for the latter case. For $d < 932$, the desired result can be verified numerically, evaluating (3.8) directly. □

Acknowledgements

The author is grateful to Mihael Perman for a fruitful discussion that led to the appearance of this paper, and for useful comments.

Supplementary Material

Proofs of certain technical issues (DOI: 10.3150/18-BEJ1072SUPP; .pdf). The supplementary file contains a proof of continuous differentiability of f_A^ε in Lemma 2.1, and proofs of Lemmas 3.3 and 3.4.

References

[1] Abramowitz, M. and Stegun, I.A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. *National Bureau of Standards Applied Mathematics Series* **55**. Washington, DC: U.S. Government Printing Office. MR0167642

[2] Ball, K. (1993). The reverse isoperimetric problem for Gaussian measure. *Discrete Comput. Geom.* **10** 411–420. MR1243336

[3] Banach, S. (1938). Über homogene Polynome in L^2. *Studia Math.* **7** 36–44.

[4] Barbour, A.D. and Čekanavičius, V. (2002). Total variation asymptotics for sums of independent integer random variables. *Ann. Probab.* **30** 509–545. MR1905850

[5] Bentkus, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. MR1965117

[6] Bentkus, V. (2005). A Lyapunov type bound in $\mathbf{R}^d$. *Theory Probab. Appl.* **49** 311–323. MR2144310

[7] Bentkus, V.Yu. (1986). Dependence of the Berry–Esseen estimate on the dimension. *Lith. Math. J.* **26** 110–114. MR0862741

[8] Bhattacharya, R. and Holmes, S. (2010). An exposition in Götze's estimation of the rate of convergence in the multivariate central limit theorem. Available at arXiv:1003.4251v1.

[9] Bochnak, J. and Siciak, J. (1971). Polynomials and multilinear mappings in topological vector spaces. *Studia Math.* **39** 59–76. MR0313810

[10] Chen, L.H.Y. and Fang, X. (2015). Multivariate normal approximation by Stein's method: The concentration inequality approach. Available at arXiv:1111.4073v2.

[11] 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. MR1841329

[12] Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Supplement to "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors". doi:10.1214/13-AOS1161SUPP. MR3161448

[13] Esseen, C.-G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law. *Acta Math.* **77** 1–125. MR0014626

[14] Federer, H. (1969). *Geometric Measure Theory*. *Die Grundlehren der Mathematischen Wissenschaften*, *Band* 153. New York: Springer. MR0257325

[15] Gaunt, R.E., Pickett, A.M. and Reinert, G. (2017). Chi-square approximation by Stein's method with application to Pearson's statistic. *Ann. Appl. Probab.* **27** 720–756. MR3655852

[16] Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739. MR1106283

[17] Livshyts, G. (2014). Maximal surface area of a convex set in $\mathbb{R}^n$ with respect to log concave rotation invariant measures. In *Geometric Aspects of Functional Analysis*. *Lecture Notes in Math.* **2116** 355–383. Cham: Springer. MR3364697

[18] Nagaev, S.V. (1976). An estimate of the remainder term in the multidimensional central limit theorem. In *Proceedings of the Third Japan–USSR Symposium on Probability Theory* (*Tashkent*, 1975). *Lecture Notes in Math.* **550** 419–438. Berlin: Springer. MR0443043

[19] Nazarov, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis*. *Lecture Notes in Math*. **1807** 169–187. Berlin: Springer. MR2083397

[20] Raič, M. (2018). Supplement to "A multivariate Berry–Esseen theorem with explicit constants." DOI:10.3150/18-BEJ1072SUPP.

[21] Röllin, A. (2013). Stein's method in high dimensions with applications. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **49** 529–549. MR3088380

[22] Sazonov, V.V. (1972). On a bound for the rate of convergence in the multidimensional central limit theorem. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California*, *Berkeley*, *Calif*., 1970/1971), *Vol*. *II*: *Probability Theory* 563–581. Berkeley, CA: Univ. California Press. MR0400351

[23] Sazonov, V.V. (1981). *Normal Approximation – Some Recent Advances*. *Lecture Notes in Math*. **879**. Berlin: Springer. MR0643968

[24] Shevtsova, I.G. (2010). Refinement of estimates for the rate of convergence in Lyapunov's theorem. *Dokl*. *Math*. **82** 862–864. MR2790498

[25] Slepian, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell Syst*. *Tech*. *J*. **41** 463–501. MR0133183

[26] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California*, *Berkeley*, *Calif*., 1970/1971), *Vol*. *II*: *Probability Theory* 583–602. Berkeley, CA: Univ. California Press. MR0402873

[27] Tihomirov, A.N. (1980). Convergence rate in the central limit theorem for weakly dependent random variables. *Teor. Veroyatn*. *Primen*. **25** 800–818. MR0595140

Received February 2018 and revised September 2018